VMASC

Virginia Modeling, Analysis and Simulation Center

Modeling, Simulation & Visualization
Student Capstone Conference

Final Proceedings

April 14, 2011
Old Dominion University

2011 marks the fifth year of the VMASC Capstone Conference for Modeling, Simulation and Gaming. This year our conference attracted a number of fine Student written papers and presentations, resulting in 29 that were presented on April 14, 2011 at the conference.

The tracks that the papers were divided up into included the following:

- Gaming
- M&S in Engineering
- Training & Education
- General Science
- Immersion/Virtual Reality
- Transportation
- Homeland Security/Military
- Medical & Healthcare

The following proceedings document is organized into chapters for each of the tracks, in the above order. At the beginning of each section is a front piece giving the names of the papers, the authors, and also indicating which of the papers in each track were recipients of the Gene Newman award for 2011.

Gaming

Track Lead: Dr. Yiannis Papelis, VMASC

Judge: Dr. Ginger Watson, ODU

Digi Island: A Smartphone Application for Learning Digital Circuit Optimization

Author(s): Joe Miller, Yuzhong Shen, and Michael Harper

2nd Place Gene Newman Award Winner, Gaming Track

Defining Learning Space in a Serious Game in Terms of Operative and Resultant Actions

Author(s): Michael Martin, Yuzhong Shen

Abstract Only

1st Place Gene Newman Award Winner, Gaming Track

1st Place (tied) for Best Paper of Conference

Digi Island: A Smartphone Application for Learning Digital Circuit Optimization

J. Miller, M. Harper, Y. Shen

***Abstract*—Karnaugh maps, also known as K-maps, are a tool used to teach digital circuit optimization. The Digi Island game is a serious game designed for aiding students to learn K-map optimization. Digi Island converts conventional classroom learning into a fun process involving the development of real estate such as resorts and amusement parks on a beautiful island in the Pacific Ocean. Digi Island allows the user to learn and progress at a personal pace rather than the pace dictated inside the classroom. The game has been modified from its original format and is now deployable on smart phone platforms in order to take advantage of the availability and popularity of smart phones such as Google's Android, Apple's iPhone and Microsoft's Windows Phone. This paper discusses the design, development, and current results of the Digi Island game as well as the ramifications of educational gaming on portable devices.**

I. INTRODUCTION

Electronic games are a pervasive aspect of American culture and entertainment: as many as 65 percent of American households play games [1]. With a revenue of $18.6 billion in 2010 [2], the game industry has evolved into an important sector that is larger than the film industry. The passion for games can be exploited for more vital purposes such as education, training, and marketing via "serious games." Game-based learning uses serious games with defined learning outcomes and objectives. The values of game-based learning have been recognized by organizations such as National Science Foundation and National Research Council [3-5]. NSF considers games as an important form of cyber-learning platform and technology [6]. The latest ground-breaking game technologies, such as Nintendo Wii and Microsoft Kinect, have significant impact on gamers and the game industry transforming gameplay into a more positive and healthy experience. The Kinect alone sold over 8 million units within 60 days of its launch in November, 2010 [18].

Digital circuits are embedded in almost all electronic equipment and devices in use today, such as computers, MP3 players, and digital cameras. Digital circuit optimization, or simplification, is a process to reduce the complexity of the digital circuits so that electronic devices will have a smaller size and therefore less weight, as well as less power consumption resulting in prolonged battery life. Various techniques have been developed in the last several decades for digital circuit optimization. Among them, the Karnaugh map is the standard method to teach digital circuit optimization in introductory digital circuit courses because its graphical representations facilitate logic simplification, providing an intuitive and systematic way for circuit optimization. However, many students have difficulties learning circuit optimization using Karnaugh maps merely because it is the first time for them to be exposed to Karnaugh maps and class lectures do not provide enough coverage and exercises. A serious game that exploits students' interest and curiosity with games would be helpful for learning circuit optimization using Karnaugh maps. As part of a Senior Design Project at the Department of Electrical and Computer Engineering of Old Dominion University, the authors developed a serious game, Digi Island, to aid teaching and learning digital circuit optimization using Karnaugh Maps.

The smart phone is rapidly becoming one of the most popular electronic devices in today's mobile phone market. Consumer reports indicate that the sales of smart phones, such as Google's Android and Apple's iPhone, contributed approximately 19% of 2010's total mobile phone sales, which is a 73% increase from smartphone sales in 2009 [19]. The meteoric rise in demand for these smartphones is due to the ability to download and run applications which perform a myriad of services and tasks. Smartphone applications, which can be bought and downloaded from carrier specific stores such as the Android Market, Blackberry App World and Microsoft Marketplace for Mobile, allows the smartphone to act as a GPS, a portable library, a portable gaming system or even a time clock for billable working hours. It is the flexibility and popularity of these smartphone applications that prompted the authors of the Digi Island learning game to port Digi Island into a format suitable for deployment on a smartphone. This paper discusses the design and development of Digi Island as well as the results of the port to the Windows Phone platform.

II. Digital Circuit Optimization Using K-maps

There are two different types of electronic circuits: analog circuits and digital circuits. Analog circuits represent and process information in continuous or analog form while digital circuit information is represented and processed in discrete binary form. Most components of modern electronic devices are digital circuits and the transformation from analog to digital is still underway. Use the media for storing music as an example. The traditional audio cassette tapes store music as analog signals, while the MP3 music players that became extremely popular in the last decade store music as digital signals. Compared with analog systems, digital systems have many advantages in terms of flexibility, programmability, computational capability, numerical accuracy, information storage and retrieval, error detection and correction, and miniaturization [7].

The same logic function can be implemented by different digital circuits with varied complexities. Thus, it is necessary to find the optimal digital circuit with minimum complexity for the desired function. Such process is called digital circuit optimization or simplification, which is important to reduce the size and weight of electronic devices and prolong their battery life. For example, consider the evolution of cell phones since their inception in terms of size, weight, and battery life. Circuit optimization is an important theoretical concept covered in introductory digital circuit courses. Various techniques have been developed in the last several decades for digital circuit optimization, including Boolean algebraic manipulation and minimization, Karnaugh maps, Quine-McCluskey, Petrick's algorithm, Espresso, and others [7]. The Karnaugh map is the standard method to teach digital circuit optimization in introductory digital circuit courses because it is a graphical representation that facilitates logic simplification providing a standard and systematic way for circuit optimization.

Karnaugh maps, also known as K-maps, are graphical representations of logic circuits which can also be represented by Boolean algebraic expressions. The sides of a K-map present circuits inputs while each cell of a K-map represents the corresponding circuit output with values of 1 or O. Figure 1 shows a K-map representing a circuit with 4 inputs. K-map optimization is essentially the process of finding a minimum number of maximal aggregations of K-map cells with values of 1 according to a set of rules. Circuit simplification using K-maps requires an understanding of several key concepts including implicant, prime implicant, and essential prime implicant [7-8]. To find the optimized expression of a K-map, all prime implicants must be identified first. The optimized expression is the logic sum of all essential prime implicants and other prime implicants consisting of minterms not included in the essential prime implicants. The remaining nonessential prime implicants can be determined using a selection rule that minimizes the overlap among prime implicants [8]. K-maps are introduced in introductory digital logic circuit courses such as ECE 241 Digital Logic Circuit at Old Dominion University. Without understanding and using K-maps proficiently, students are likely to fail in this introductory course and more advanced digital circuit courses.

AB \ CD	00	01	11	10
00	1	0	1	0
01	1	0	1	0
11	1	1	1	0
10	1	0	1	0

Fig. 1. A 4 variable K-map.

III. Game Design

The goal of the game Digi Island is to provide both a formal introduction to K-maps and an engaging game setting that encapsulates the K-maps. To this end, Digi Island is designed as a construction based strategy game [9-10]. The game has three modes: Tutorial Mode, Practice Mode, and Play Mode. The Tutorial Mode provides several tutorials about the K-map through exemplary circuits with 2, 3, and 4 inputs. This mode identifies implicants, prime implicants, and essential implicants in these circuits, illustrates the procedure of selecting essential prime implicants and nonessential implicants using the selection rule, and finally generates the Boolean expression for the optimized logic circuit. The Practice Mode displays arbitrarily generated K-maps with 2, 3, and 4 input variables and asks the player to identify the essential prime implicants. User interfaces should be provided to start and end identification of these terms. The player also needs to generate the optimized Boolean expression. User interfaces should be provided to the player to enter input variables, their complements, and the logical OR operation. Both the Tutorial Mode and Practice Mode provide instructions and practices using standard representations of K-maps that are used in regular classrooms.

The Play Mode is the fun part of the Digi Island game. There are no more K-maps in the Play Mode and instead the player sees a beautiful island (called Digi Island) in the Pacific Ocean. In the game, the player is an adventurer to the Digi Island and will transform it into a tourist attraction by developing real estates such as amusement parks and hotels. However, not all places can be exploited by the player such as rocks and reserves for wild lives. The player is given a map of Digi Island that labels all usable and unusable spaces. Large buildings bring about better financial outcomes as they provide more efficient utilization of the space. Thus, the goal of the player is to construct a minimum number of buildings as large as possible covering all the usable spaces, while satisfying the regulations of the Pac Republic, which exercises sovereignty over Digi Island. Some sample regulations are listed below.

- All the spaces occupied by each building must be adjacent to each other.
- The number of usable spaces (blocks) in each building must be a power of 2.
- Sharing spaces between adjacent buildings is allowed, but should be minimized.

Initially, the player has a certain amount of cash that can be used to exploit the island and construct buildings on the usable spaces. Depending on the importance of the buildings, they have different values and are represented differently, e.g. single houses or skyscrapers. Larger buildings are more valuable and generate more profit for the player and thus more points.

In addition to the game design discussed above, the game has a number of other requirements. The game must be deployed and playable on personal computers and the Microsoft Windows Phone The touch screen of the Windows Phone needs to be utilized as an input device. Sound effects are also required to provide user feedback. Voice instructions should be provided as well.

IV. Game Development

The development of the game Digi Island contains two major components: front end and back end. The front end mainly contains a graphical user interface which displays menus, tutorials, and K-maps. The front end has different user input modes for K-map manipulation and provides user feedback for their actions. The back end contains the major logic for digital circuit optimization, that is, for an input digital circuit, the back-end generates the optimized Boolean expression, compares that with the player's solution (answer), and provides feedback to the player. In the following, the back end is discussed first.

A. Logic Circuit Optimization

Since the goal of the game is to help the player learn logic circuit optimization using K-maps, the game must know the final optimized Boolean expression for an input logic circuit. Although K-maps are effective visual tools used by humans for manual circuit optimization, they are not suitable to be implemented on computers for automatic circuit optimization. Several other logic circuit optimization algorithms have been developed. However, here we mainly discuss the two most widely used methods: Espresso algorithm and Quine-McCluskey algorithm.

The Espresso algorithm is the de facto industry standard for circuit optimization and it was initially created by Brayton et al. [11] and later revised by Rudell of University of California at Berkeley [12]. The Espresso algorithm is a heuristic but effective algorithm in terms of memory usage and computational complexity. Although it does not guarantee to produce an optimized circuit, in practice it always leads to a solution that is either optimal or very close to optimal. The source code of the Espresso algorithm in C programming language is available for downloading from the University of California at Berkeley [13]. There are several options to utilize the Espresso algorithm in the game: 1) porting the Espresso C source code to C# in the game, 2) compiling the source code into libraries and calling the libraries in the game, and 3) calling the executable of Espresso directly. Considering that the Espresso algorithm has a high degree of complexity, it was not feasible to port the Espresso source code from C to C# in this project. Options 2 and 3 worked for the game running on the personal computers (PCs), but not for the game on the Windows Phone since there is no C compiler that can generate object code for the Windows Phone platform. Considering these obstacles, the Espresso algorithm was not deemed suitable for this project.

The Quine-McCluskey algorithm is another widely used method for circuit optimization [14-15]. The Quine-McCluskey algorithm is designed to work similarly to the human brain's pattern recognition. It is a systematic method that guarantees to produce the optimized Boolean expression and its tabular form making it suitable for computer implementation. The Quine-McCluskey algorithm first finds all implicants with n variables, then combines some of them to implicants with n - 1 variables, and continues this combination process until all prime implicants are found. The algorithm then identifies all essential prime implicants using a prime implicant chart. However, the circuit is not fully optimized or minimized yet as the remaining prime implicants may still have overlap. A covering procedure is then utilized to select a minimum number of remaining non-essential prime implicants in the prime implicant chart so that the circuit function is fully covered. Unlike the Espresso algorithm, no authoritative source code exists for the Quine-McCluskey algorithm. Since it is a systematic method that is straightforward to implement, the team decided to develop the C# code for the Quine-McCluskey algorithm from scratch.

The Microsoft .NET Framework and C# programming language were utilized to develop the code. C# is an object-oriented programming language drafted by Microsoft and has been approved as a standard by ISO. The final version of the developed algorithm based on the Quine-McCluskey method takes an input of an n variable K-map and iteratively discovers all possible solutions, eliminating overlaps and unnecessary implicants through each iteration. In order to facilitate the best possible gaming experience, the algorithm returns not one solution but a list of all possible solutions which allows for the greatest flexibility in the playstyle of the game.

B. Game Play

Microsoft XNA Game Studio is a game development toolkit for Windows, Xbox 360, Zune HD players, and Windows phones. The XNA Game Studio consists of two parts: XNA Framework and a set of tools and templates for game development. The XNA Framework is an extensive set of libraries for game development based on the Microsoft .NET Framework. It encapsulates low-level technical details so that game developers can focus more on content and high-level development [16]. XNA provides templates for common tasks, such as development of games, game libraries, and game components. It also provides utilities for cross-platform development, publishing, and deployment. Developers can make use of both the XNA Framework and the .NET Framework in the game with the former for game-specific tasks such as graphics rendering and managing inputs and latter for more general programming tasks.

XNA Game Studio is a powerful tool for rapid development of cross-platforms and was selected as the tool to implement the graphical user interface and game play. XNA Game Studio 3.1 was used in the original development of Digi Island, but was later upgraded to the recently released XNA 4.0 which enabled support for the newly released Windows Phone. The new update allowed for a wide variety of touch screen interactions including gestures and multi-touch capabilities. Object oriented programming (OOP) was utilized and a number of classes were developed to represent different game scenes, graphical user interface, user inputs, K-map, and logic circuit optimizations. XNA Game Studio provides a fundamental class Game that handles game logic update and drawing. A class diagram of the game is shown in Figure 2.

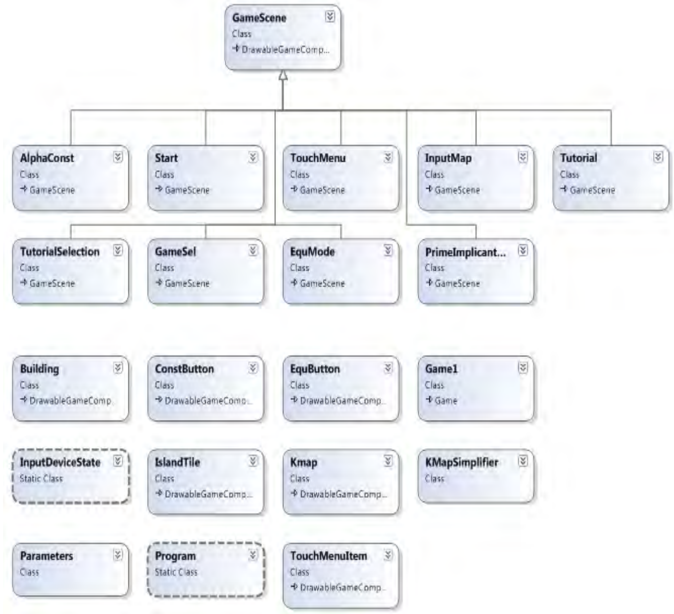
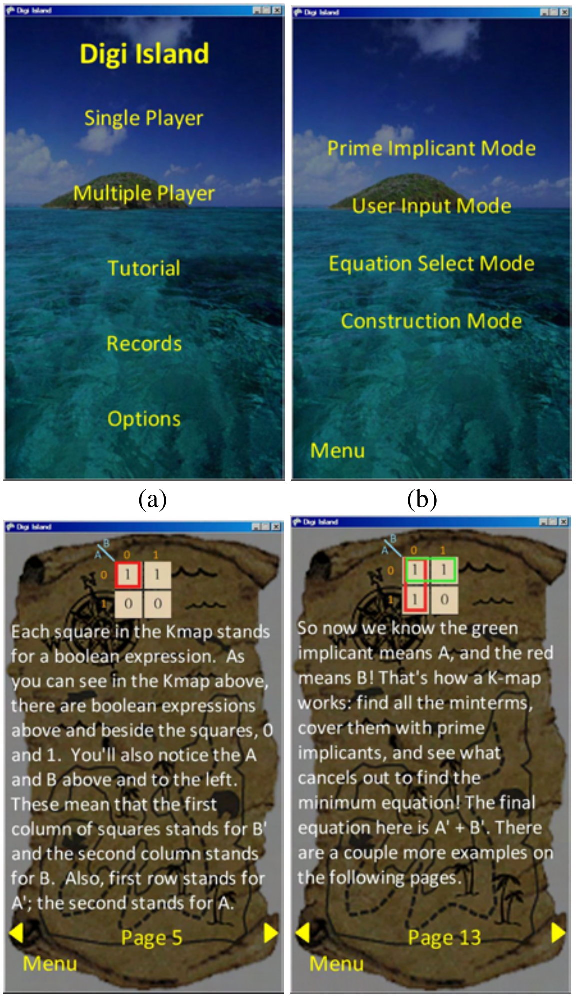


Fig. 2. Class diagram of Digi Island. Main classes connected to GameScene; support classes shown below.

V. RESULTS

As of February 2010, the Digi Island game is almost fully operational with a few minor graphics and visualization updates yet to be implemented. The game has a fully functional tutorial which demonstrates the form and function of Karnaugh maps to students and aids them in learning how to interpret a K-map and solve for the most optimal digital logic solution. The PC version contains voice overs for the tutorial recorded by the authors for the edification and clarification of the player. The voice overs were omitted in the Windows Phone version, however, due to space requirements for smart phone applications.

The main menu is the player's access point to the rest of the game. The user may choose to immediately access the single player mode or read the tutorial first; the multiple player and record modes are not functional at this time and will be implemented in the future. The Tutorial Mode is designed to familiarize the player with the concepts behind K-maps and how to solve them quickly and efficiently. The player starts the tutorial on page one and may flip back and forth through the different pages of the Tutorial Mode by clicking on the large yellow arrows on the bottom left and right of the screen. The tutorial begins with very simple, straightforward introductions to a K-map, but the topics and teachings become more advanced as the tutorial progresses. Once finished, the player may exit to the main menu by clicking the "Menu" button in the bottom left corner of the screen. The submenu displays when the user selects the "Single Player" option on the main menu and acts as a gateway between the four playable modes. The player may return to the main menu in the same manner as the tutorial: by clicking the menu option in the bottom left corner of the screen. See Fig. 3 for screen captures displaying the main menu, Play Mode selection menu and Tutorial Mode.

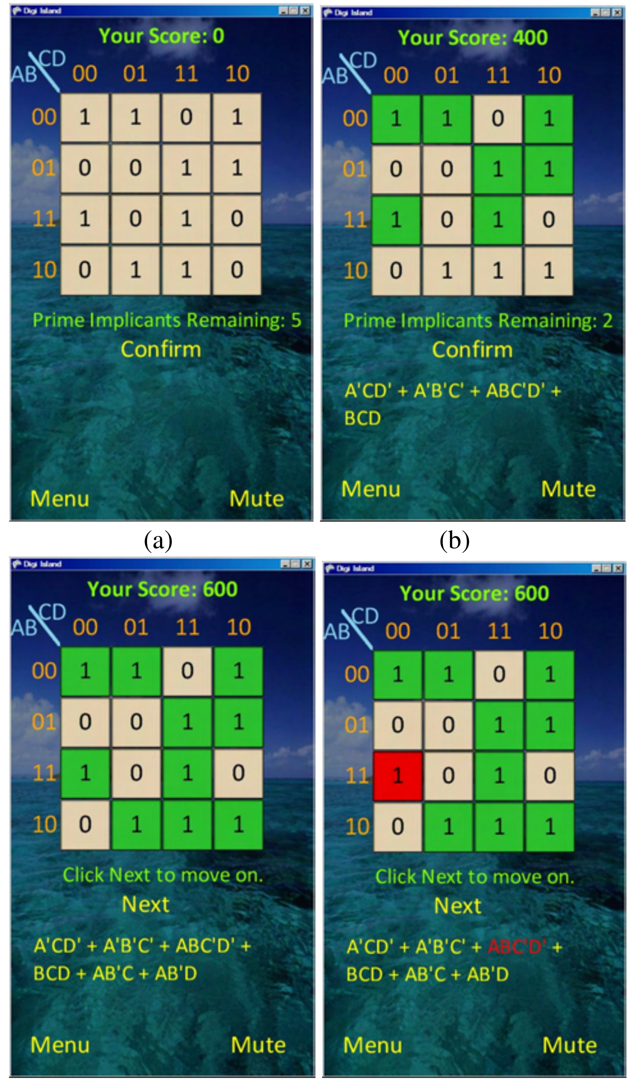


(a) (b)

(c) (d)

Fig. 3. Screen captures of Digi Island menus and tutorial. (a)Main menu. (b)Single Player submenu. (c)Tutorial Mode. (d) Example of advanced topic later in the tutorial.

The Practice Mode is also fully functional and contains three separate types of practice options, up from the original two. The first mode is the Prime Implicant mode which generates a random K-map and asks the user to select the prime implicants. Each time the player selects the correct implicant, their score is increased by one hundred points and the logical value of the selected implicant is displayed for the user to associate with the selected implicant. Once all implicants are selected, the user can see the entire optimized equation and may scroll over each term to see how it corresponds to the map. Further K-maps are generated randomly in this manner until the player exits the Prime Implicant mode. An example of the process of solving a K-map in this mode is shown in Fig. 4.



(a) (b)

(c) (d)

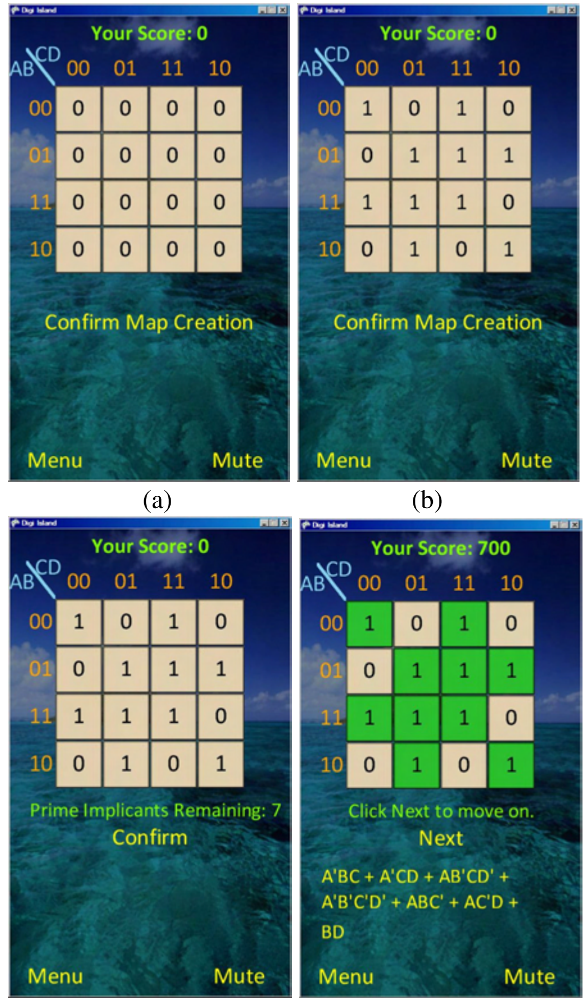
Fig. 4. Examples of Prime Implicant Mode. (a)Randomly generated unsolved K-map. (b)Partially solved K-map with 2 remaining unsolved implicants. (c)Fully solved K-map. (d)Mouseover Boolean logic text highlights corresponding implicant.

The Equation Select mode displayed in Figure 5 also generates a random K-map in a manner similar to the Prime Implicant mode. However, instead of requiring the user to highlight implicants on the map itself they are required to input the logical value of the implicant instead. For example in a four variable K-map the user can select a value using some combination of "A", "B", "C", "D" and their respective NOT symbols which generates the logical value for the implicant they wish to select.

(a) (b)

Fig. 5. Screen captures Equation Select Mode. (a)Randomly generated K-map ready to be solved by the player. (b)Solved K-map that highlights implicants corresponding to the Boolean logic value mouseover text.

The final Practice Mode option is the User Select Mode, which differs from the Prime Implicant mode because he/she must instead input a map of his/her own choosing and solve it, rather than solving a computer generated map. This allows the player the freedom of testing their understanding about a particular K-map, which is useful as some possible permutations of K-maps have multiple solutions and can be difficult to interpret. For example, in Figure 6 below, the player begins with a completely blank K-map or a K-map filled entirely with zeroes. The player wants to practice a very difficult K-map and creates the map by clicking on the blank map to fill in ones in appropriate squares. The player then clicks the button "Confirm Map Creation" on the K-map he/she wants to solve. The game then takes the map the player designed and generates all possible minimal solutions then returns the number of prime implicants the player needs to locate. The player may then solve the map in a fashion similar to that used in the Prime Implicant mode.



(c) (d)

Fig. 6. Screen captures of User Input Mode. (a) A blank K-map. (b) User generated K-map. (c) User generated K-map converted to solvable K-map. (d) Solved user generated K-map.

The Play mode as currently implemented replaces the K-maps of the Practice mode with Digi Island, a beautiful and untamed land that the player must develop with real estate in order to attain the maximum profit. The user may select a building size from one of eight buttons on the playing field and once generated may drag the building to a site on the island. If the selected area on the island is suitable for building, then the building is constructed and the player receives income specific to the constructed building. If the location is unsuitable or overlaps with another building, the player receives a penalty for not optimizing the available space.

The Digi Island game is now fully deployable and playable on the Windows Phone platform. The phone has a touchscreen for input so the user may select implicants and/or buildings with a touch rather than a mouse click. Implementation on the phone also allows for a few gesture innovations such as resetting the mode currently being played via touching the screen with two fingers. See Fig. 7

below for an example of the Prime Implicant Mode as played on the Windows Phone.

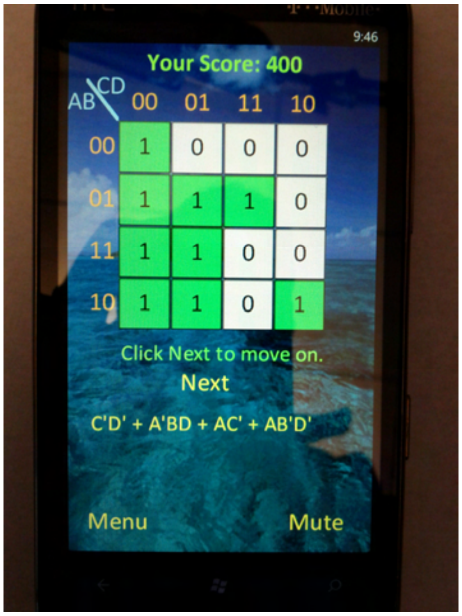


Fig. 7: Digi Island deployed on the Windows Phone 7.

VI. Analysis and Future Developments

The Digi Island game already demonstrated many advantages over traditional teaching methods. For example, the typical K-maps in the textbooks contain many overlapping rectangles representing different primary implicants, making them visually confusing and difficult to understand. In the Digi Island game, each primary implicant can be selected and highlighted individually through user interactions, leading to a much clearer representation and better understandings. Advanced rendering techniques, such as transparency control using alpha maps, can be used to align two K-maps to facilitate groups of minterms, which is not possible by using just plain textbooks or paper and pencil methods.

One important principle of learning is to connect new concepts and understanding to pre-existing knowledge [17]. The Digi Island game elegantly converts boring 1s and 0s in digital circuits into usable and unusable spaces on a beautiful island and transforms K-map optimization into real estate development, an activity with which many students are familiar and also interested in. The rules for K-map optimization exhibit themselves as construction regulations for real estate development. Players will be more engaged when they deal with real assets such as skyscrapers and amusement parks rather than blocks of abstract 1 s and 0s.

The conversion of the Digi Island project into a format deployable on the Windows Phone resulted in another advantage over traditional learning methods: portability. Once submitted to Microsoft's App Hub, Windows Phone users will be able to download and play the Digi Island game in a matter of moments, allowing them to play the game by merely pulling out their smartphone and starting the application. This way, the player may execute the game in nearly any location, without the need for a computer. The learning process will no longer be limited to the lecture hall or confined to the PC; the user may play the game in any place, at any time.

An area of future development is to include "don't cares" in the K-maps, which are the outputs of certain inputs that do not matter. The "don't cares" can be treated as either 1 or 0. Usually including the "don't cares" in the K-map optimization can further simplify the circuit's logical solution.

The game can be further expanded to have multi player mode to form more competitive and engaging game plays. Players can face off against each other to see who can solve a K-map the fastest. With networking, a game server can also be setup to store player configuration and performance. The Windows Phone in particular supports the multiplayer upgrade, as each phone supports multiplayer connections via the Windows Live. One may connect to the Windows Live service, look for other people also playing the Digi Island game and connect with them for a multiplayer experience, in a manner similar to that of XBOX Live multiplayer.

VII. Conclusion

K-maps are an important tool for teaching digital circuit optimization and simplification which is critical to reduce physical size and power consumption of electronic devices and prolong their battery life. This paper discussed the design and development of Digi Island, a serious game based on K-map optimization. The game was developed as a product of a senior design project at Old Dominion University and continued as a research project in the Modeling and Simulation graduate department of Old Dominion University. Digi Island is a fun and engaging game that offers many advantages over traditional teaching methods of K-maps and it will be completed and released in the near future.

REFERENCES

[1] Entertainment Software Association, "Industry Facts," http://www.theesa.com/facts/index.asp, 2010.
[2] NPD Group, "2010 TOTAL CONSUMER SPEND ON ALL GAMES CONTENT IN THE U.S. ESTIMATED BETWEEN $15.4 TO $15.6 BILLION," http://www.npd.com/press/releases/press_110113.htm/, 2010.
[3] The National Academies, "Improving Undergraduate Instruction in Science, Technology, Engineering and Mathematics," 2003.
[4] Project Kaleidoscope, “Recommendations for Action in Support of Undergraduate Science, Technology, Engineering, and Mathematics," 2002.
[5] American Association for the Advancement of Science, "Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and Mathematics Education," 2005.
[6] NSF Task Force on Cyberlearning, "Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge," National Science Foundation, 2008.
[7] V. P. Nelson, H. T. Nagle, J. D. Irwin, and B. D. Carroll, *Digital Logic Circuit Analysis & Design*: Prentice Hall, 1995.
[8] M. M. Mano and C. R. Kime, *Logic and Computer Design Fundamentals*. Upper Saddle River, NJ: Pearson Prentice Hall, 2008.
[9] T. Fullerton, *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*: Morgan Kaufmann, 2008.
[10] J. Schell, *The Art of Game Design: A Book of Lenses*: Morgan Kaufman, 2008.
[11] R. K. Brayton, A. L. Sangiovanni-Vincentelli, C. T. McMullen, and G. D. Hachtel, *Logic Minimization Algorithms for VLSI Synthesis*. Norwell, MA: Kluwer Academic Publishers, 1984.
[12] R. L. Rudell, "Multiple-Valued Logic Minimization for PLA Synthesis," *Memorandum No. UCBIERL M86-65*, 1986.
[13] University of California at Berkeley, "Espresso," http://embedded.eecs.berkeley.edu/pubs/downloads/espresso/index.htm ,2010.
[14] W. V. Quine, "The Problem of Simplifying Truth Functions," *American Math Monthly*, vol. 59, pp. 521-531, 1956.
[15] E. J. McCluskey, "Minimization of Boolean Functions," *Bell System Technical Journal*, vol. November, pp. 1417-1444, 1956.
[16] Microsoft*, XNA Game Studio Documentation*, 2009.
[17] Board on Behavioral, Cognitive, and Sensory Sciences and Education, *How Students Learn: History, Mathematics, and Science in the Classroom*. Washington, DC: The National Academies Press, 2005.
[18] Josh Lowenson, “Timeline: A look back at Kinect's history,” http://news.cnet.com/8301-10805_3-20035039-75.html, 2011.
[19] Gartner, Inc., “Gartner Says Worldwide Mobile Device Sales to End Users Reached 1.6 Billion Units in 2010; Smartphone Sales Grew 72 Percent in 2010,” http://www.gartner.com/it/page.jsp?id=1543014, 2011.

Defining Learning Space in a Serious Game in Terms of Operative and Resultant Actions

Author(s): Michael Martin, Yuzhong Shen

Keywords: Serious Games, Cognitive Load Theory, Operative Actions, Resultant Actions, Learning Space

Abstract: This paper explores the distinction between operative and resultant actions in games, and proposes that the learning space created by a serious game is a function of these actions. Further, it suggests a possible relationship between these actions and the forms of cognitive load imposed upon the game player. Association of specific types of cognitive load with respective forms of actions in game mechanics also presents some heuristics for integrating learning content into serious games. Research indicates that different balances of these types of actions are more suitable for novice or experienced learners. By examining these relationships, we can develop a few basic principles of game design which have an increased potential to promote positive learning outcomes.

M&S in Engineering

Track Lead: Dr. Rafael Diaz, VMASC

Judge: Dr. Jose Padilla, VMASC

Using Decision-making Techniques in Support of Simulation Training Transfer
Author(s): Jane Bachman and Patrick Hester
Abstract Only
Invited Paper

Simulation–Based Education for Teens Concerning Their Health Now and In Years to Come
Author(s): Charles Reese
3rd Place Gene Newman Award Winner, Engineering Track

Use of Architecture Modeling for Insertion of New Technology
Author(s): Kevin Michael and Andreas Tolk
2nd Place Gene Newman Award Winner, Engineering Track

Comparison of Conceptual Modeling Frameworks to Facilitate Standardization
Author(s): Rabia Haq

Agile Project Management Methodology for Transition to Next Generation Architecture
Author: Rupesh Narkar
1st Place Gene Newman Award Winner, Engineering Track

Using Decision-making Techniques in Support of Simulation Training Transfer

Author(s): Jane Bachman and Patrick Hester

Keywords: Analytical Modeling, Human Computer Interface, Human Performance, Military Training, Multi-criteria Decision-making models, Simulation Training Transfer

Abstract: In the United States, the military services are encouraged to promote and execute war fighting excellence. The execution of military training is a very important aspect of promoting and executing war fighting excellence. During the military training planning process, requirement versus capability preparations often involve additional training approaches that include new methods or techniques that are not currently conducted in military training simulations. How are decisions made when determining the selection of military training simulations with respect to military training requirements? Is there a general methodological approach that decision-makers are utilizing? This research will first perform a literature review to identify relevant previous literature in this area. Previous literature will be examined to determine if multi-criteria decision-making techniques were utilized, and to what extent. Utilizing the results of the literature review, the focus of this paper is to identify the benefits of having a general methodological approach and to identify, what, if any, previous approaches can be utilized in the development of a comprehensive methodology.

Simulation–Based Education for Teens Concerning Their Health Now and In Years to Come

Charles Reese
Old Dominion University
Norfolk, VA 23508 USA

***Abstract*- One of the fastest growing problems in the United States today is the issue of childhood obesity. It causes issues that are far greater than most realize. Most adults that are overweight began that process of gaining weight many years ago when they were first making choices concerning their levels diet and exercise. Now as an adult they realize their mistakes but are forced now to pay the price for the uninformed decisions they made years earlier. It is much more difficult to reverse these effects as an adult as it would have been to stop these events from occurring.**

This paper discusses a simulation to inform high school health class students that the choices you make as a child/teen or early adult can impact your life for years to come. The simulation introduces the student to variables such as their weight, calories consumed, BMI, and BMR. The student can then experiment with the simulation to determine their current level of health and the state of their health in years to come. The objective is to create a mental model that the learner can refer to when they make the decisions not exercise or to overeat. That mental model will serve as a guide to help them make the correct choices since they saw the negative effects bad choices made at this age will have on their life in the future.

I. INTRODUCTION

Many Americans die each day from complications related to obesity; despite billions of dollars spent by the government for health costs (hospital visits and medicine) the number continues to rise. The best solution would be to address this problem at its infancy before it becomes out of hand. One solution is education before or during the critical years when teens become adults and begin to make the choices concerning their level of exercise and what and how much they eat. The goal is to allow them to interact with a simulation that explains this underlying process in a manner that most students this age are comfortable with.

While traditional education is still applicable such as a health class, the introduction of a computer simulation increases the likelihood of the student being motivated to learn the material. The easiest way to get someone to learn is to make it fun and not to let them know they are learning. This paper describes the simulation design process, from the inception of the idea to creating a working simulation that can be used for the purposes described above.

II. LITERATURE REVIEW

The general instructional approaches this simulation will follow are cognitive strategies to help the learner construct a mental model of how diet and exercise affect weight and health. The main instructional approach within the simulation will be generative, and more specifically through elaboration. To help the learner build a mental model and link it to existing schemata, the learner will be asked questions both within the simulation and on an associated worksheet. These questions will require the learner to predict what will happen within the simulation and in the future. One example would be, “If you reach this body weight, what impact would it have on your ability to do your job? What impact will it have on your social life?” These questions will also allow the affective aspects of the objectives to be addressed as well.

The scenario is based on recommendations from Campbell and Monson (1994) to provide information to the learner on a just-in-time basis. Rather than beginning with expository content, the learner has the option to begin immediately with the simulation, the personal exercise, or an orienting section describing basic operation of the simulation based upon their level of knowledge as a novice or expert. Choi (1997) suggests that the simulation provide an opportunity for coaching on how to use the simulation, along with a situation to engage the learner’s interest. The concepts are introduced along the way.

This model is verified through comparison of outputs from a scientific model developed by Kosinski (2005). The Kosinski (2005) model is much more in depth and approaches the same process considering the internal chemical and physiological processes associated with the body’s processes as factors of types of activity and calorie intake. One limitation of the Kosinski (2005) model was in its limitations of specific physical activities to be factored into the process. An additional limiting factor in the Kosinski (2005) model was that it presented two separate models based upon gender. The current model developed as a part of this process uses the equations listed later in this paper in calculations for BMI and BMR. It is believed that the model fidelity may actually permit the model to be used at dual levels: 1) understanding the predictive results to change behavior, and 2) allow more advanced learners to investigate the model itself and provide a greater transfer of learning by understanding the relationships of the variables involved.

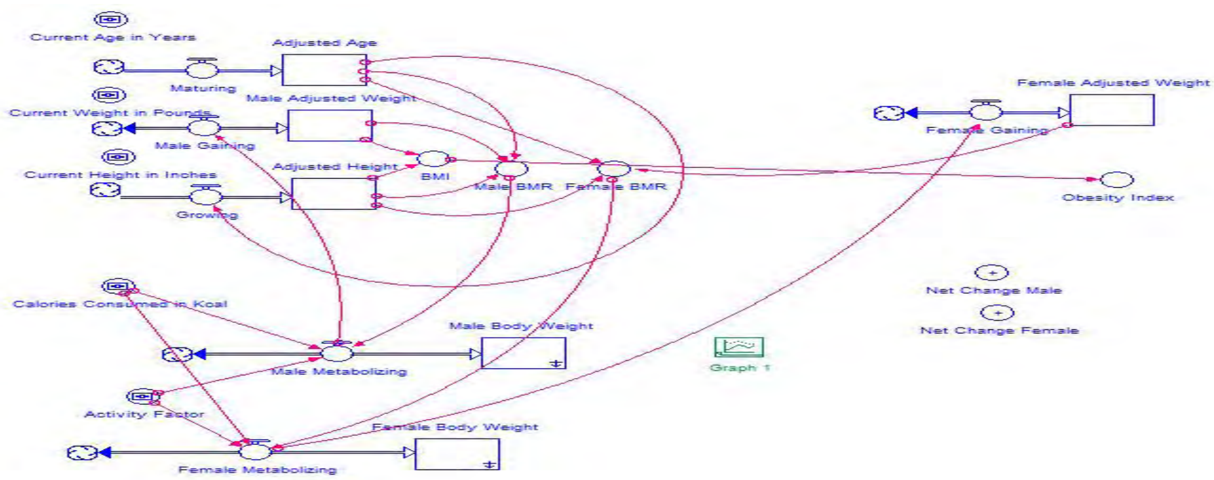


Figure 1. STELLA Model

III. PROJECT WORK

The purpose of the simulation is to illustrate the impact and relationship diet and exercise has on diabetes and heart disease in adolescents in order to affect their dietary and exercise choices. The instructional objectives are to ensure that at the end of this simulation, the learner will be able to:

1) Calculate their ideal body weight using height/weight charts;
2) Identify the concepts of Body Mass Index (BMI) and Basal Metabolic Rate (BMR);
3) State the effect exercise and nutrition have on BMI and BMR;
4) State the link between obesity and heart disease/diabetes;
5) Compute the caloric intake and level of exercise required for them to maintain or achieve ideal body weight using the simulation; and
6) Construct a 7-day nutrition and exercise plan to reach their ideal body weight in a healthy manner using the guidelines from the FDA, an exercise data table, and a spreadsheet.

The Learner

This simulation is aimed at a high-school audience; learners are in grades 9-12 and are typically 14-18 years of age. They will use this simulation as part of a health and physical education class, but they possess limited knowledge about nutrition. Specifically, these learners: 1) are comfortable using window computer environments including web browsers; 2) may have low motivation to complete assignments in a pass-fail health/physical education class; and 3) are able to speak, read, and write English at a minimum of an eighth grade level. Because this instruction will take place in a standard, required course, learners represent a wide range of physical and mental abilities. Some learners will have physical or learning disabilities. Some accommodation outside the simulation will be necessary for more severe cases.

As learners begin using the instructional simulation, they must already be able to: 1) turn on and log on to a personal computer on a school network; 2) control programs using a mouse; 3) enter numbers using the keyboard; 4) enter data into a spreadsheet; 5) use standard graphical user interface elements including radio buttons, check boxes, scroll bars, sliders, and buttons; and 6) use a web browser to navigate a web site.

Learners are taking this course to fulfill a requirement. Motivation will vary widely from learner to learner. Some students may be motivated to learn about nutrition and maintain a healthy body weight because of personal or family history. However, many learners' motivation will be low because in many schools the health and physical education component is pass or fail. The simulation and the surrounding instruction will have to convince the learners of its usefulness since learners will enter the instruction with differing perceptions of its use. Accountability will happen as part of the course when the learners have to turn in several worksheets for course credit.

This simulation is directed at schools with either computer labs or one-to-one programs in which students have access to a computer. Every learner should have access to his or her own computer for the entire time period of the simulation. The learner may need access to a printer for some of the resource materials unless they are printed in advance and distributed with the worksheets and instructions.

Learners are expected to apply the knowledge gained from this simulation in their lives. Unfortunately, this is a very large context over which the designers have no control. For example, even in the school cafeterias good nutrition is not as important as the bottom line. Learners will need to make nutritional choices in an environment that offers many bad options and few appealing good options. The instruction will have to take this into account and equip the learner to make good choices in spite of a difficult environment.

The Simulation

Learners sit down individually in front of the simulation as part of a high school health class. They enter some basic information including name and gender, and are then introduced to a case study matching their gender. For example, a story of a teenager close to their age with a brief biography and a photograph is displayed. The next screen takes them to the main simulation screen, but all the controls except for the Run button are locked. The case study shows an individual of average weight, with their caloric intake and exercise level fixed. When the user runs the simulation, the case study's weight slowly increases over time, as displayed by the growing silhouette and on the chart. The user is then randomly presented with a weight-related health issue such as diabetes or heart disease that the case study is said to experience.

The learner is then introduced to the concept of BMI; a definition is presented and the location of BMI on the main simulation screen is pointed out by a callout. The learner is then informed that one of the contributing factors to BMI is caloric intake. The control for caloric intake is then unlocked in the simulation and the learner is encouraged to run the simulation again after changing the value. When the learner clicks the run button, a dialog will ask the learner to predict what will happen over time; this is a generative strategy to reinforce the concept of BMI. If caloric intake is adjusted down and the simulation is run again, the case study's weight will still increase but less drastically.

Next the learner is introduced to the concept of Basal Metabolic Rate (BMR), which influences how the body stores or expends excess calories. The control for exercise level is then unlocked, and the learner is encouraged to run the simulation again modifying both caloric intake and exercise activity. Again when the user runs the simulation, they will be prompted to type in a prediction and explain why they made that prediction. With both controls unlocked, now it is possible for learners to keep the case study's BMI within a desired range. The learner receives feedback based on how the case study does. If the case study's BMI changes too much or if the input values were not realistic, they are encouraged to try again.

Learners are then prompted to enter their height, weight, and BMI. These attributes are located on a worksheet from a prior day in health class where this was measured by the teacher. A height and weight chart is displayed on the screen along with instructions on how to read it. The learner is then asked to locate the ideal weight for their height and build on the chart by clicking on it. If the learner clicks incorrectly they are given additional help, but after a third incorrect try the correct weight is highlighted on the chart. Feedback is then given based on the comparison between their BMI and weight to the ideal values displayed on the height/weight chart and BMI range chart.

Next the learner is returned to the main simulation screen, but this time the simulation is based on their actual height and weight. The caloric intake input is set high by default, to represent an average American diet, and the exercise slider also starts at the low end. Learners may run the simulation with these values to see what happens first, or they may begin by changing the values. Under the exercise level control, an exercise data table is linked so that learners may look up the value corresponding to their actual or desired activity level.

While using the simulation with their real data, learners follow a worksheet. This worksheet asks them to record each set of inputs they enter and then record the results. At this point, learners will be able to run the simulation in five year increments and adjust input values at each interval. When learners have reached a reasonable caloric intake and activity level and have reached or maintained a healthy BMI, they record the results in the next section of the worksheet. Using the exercise data table in the simulation, they create an exercise plan to reach that level of activity and record it on the worksheet. As a final activity, they are taken to a website listing the calorie contents of common foods. Then using a spreadsheet, they create a 7-day nutrition plan which fits the caloric intake they identified in the simulation.

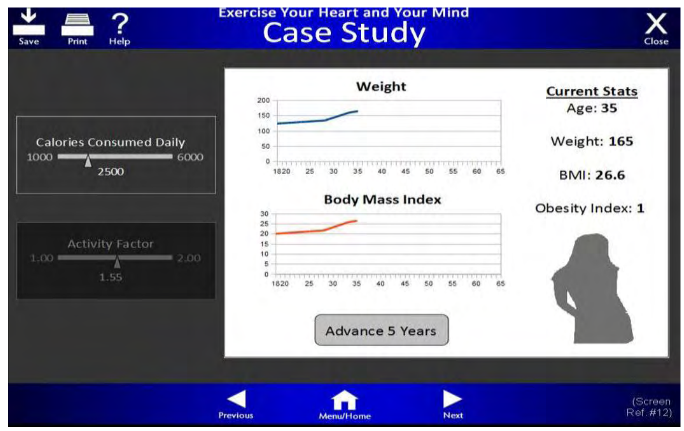


Figure 2. User Interface Storyboard

PowerPoint is used as a storyboarding tool to illustrate the user interface and interact with the model underlying the simulation. An example screen is shown in Figure 2.

STELLA

The model was developed using STELLA v9.1.4 as shown in Figure 1. It graphically depicts changes in learner weight as a function of physical activity level and calorie intake over time. The predictive value of the model is found in the resultant BMI, which may be used as a measure for potential alerts specific to Type II diabetes and heart disease. Variables specific to the user's age, weight, height, and gender are used to calculate BMI. The BMI calculations are based upon widely accepted formulas as well as inputs into calculating the user unique BMR to determine calorie expenditures. The underlying model and formulas used in the calculations from the equations screen are presented below. The interface screens displayed below are presented here only for understanding as they relate to the model.

The User Input Variables:
The following are variables the user supplies to the simulation. Figure 3 shows the user interface for manipulating these variables.

- Current Weight: Weight of learner in pounds.
- Current Height: Height of learner in inches.
- Current Age: Age of learner in years.
- Calories Consumed: The learner's actual, estimated, or planned caloric consumption derived from worksheets and additional instructional content linked to the model.
- Activity Level: The learner's actual, estimated, or planned activity level derived from worksheets and additional instructional content linked to the model. The activity level is based on the Harris Benedict indicators of activity as it impacts metabolic rate due to level and intensity and is used in determining total number of calories expended on a daily basis.

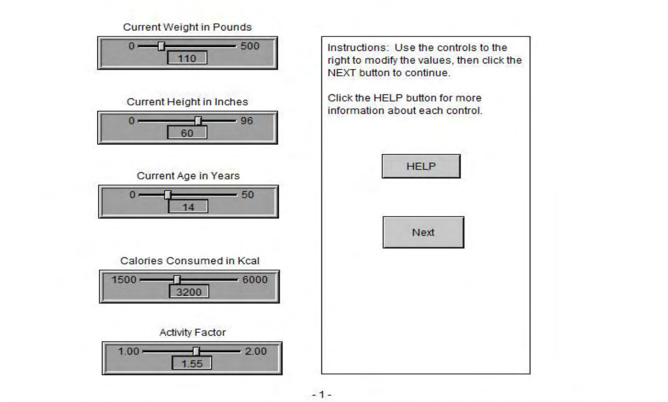


Figure 3. STELLA Screen for User Input Variables

The Hardcoded Variables:
The following variable is hardcoded into the simulation.

- Calories per Pound Conversion Rate: Used to determine the number of calories associated with a pound of body fat, and is set to 3,500 based upon web-based searches.

The Calculated Variables:
The simulation calculates the following variables based on the equations supplied.

- BMI: Based upon the learner's height and weight as it changes through the execution of the model. It is used as a factor in determining the obesity index as an indicator for diabetes and heart disease related complications due to nutrition and activity levels. The formula is defined as:
$$BMI = (705 \times Weight) \div (Height \times Height)$$
- Obesity Index: An integer value derived from the BMI and is used as an indicator for potential diabetes and heart disease risks over time. The indices are:
  0 = underweight (BMI < 18.5),
  1 = normal (BMI < 25),
  2 = Overweight (BMI < 30),
  3 = Obese (BMI < 40),
  4 = Morbidly Obese (BMI > 40).
- Male BMR: Calculated by converting English measurements to metric equivalents and used to derive the number calories expended by males attributed to natural physiological metabolic activity. The formula is defined as:
$$Male\ BMR = 66 + \big(13.75 \times (Weight \div 2.2)\big) + \big(5 \times (Height \times 2.54)\big) - (6.8 \times Age\ in\ years)$$
- Female BMR: Calculated by converting English measurements to metric equivalents and used to derive the number calories expended by females attributed to natural physiological metabolic activity. The formula is defined as:
$$Female\ BMR = 655 + \big(9.56 \times (Weight \div 2.2)\big) + \big(1.85 \times (Height \times 2.54)\big) - (4.68 \times Age\ in\ years)$$

The Stored Variables:
The following variables are maintained internal to the simulation.

- Adjusted Age: Tracks the age progression from Current Age as the model executes and is used in adjusting for height changes due to maturation.
- Adjusted Height: Tracks the height progression as the model executes and is used in adjusting for BMI and BMR calculations.
- Male Body Weight: Tracks the adjusted weight for males as a factor of calories consumed, activity level, BMR, and weight measures over time.
- Female Body Weight: Tracks the adjusted weight for females as a factor of calories consumed, activity level, BMR, and weight measures over time.
- Male Adjusted Weight: Tracks the adjusted weight for males as a factor of calories consumed, activity level, BMR, and Male Body Weight measures over time.
- Female Adjusted Weight: Tracks the adjusted weight for females as a factor of calories consumed, activity level, BMR, and Female Body Weight measures over time.

The Output Variables:
The following variables are computed and displayed to the user. They are output as shown in Figure 4.

- Male Weight: Tracks the adjusted weight for males as a factor of calories consumed, activity level, BMR, and Male Body Weight measures over time.
- Female Weight: Tracks the adjusted weight for females as a factor of calories consumed, activity level, BMR, and Female Body Weight measures over time.
- Male Body Weight Gain/Loss: Tracks the adjusted weight gain/loss for males as a factor of calories consumed, activity level, BMR, and weight measures over time.
- Female Body Weight Gain/Loss: Tracks the adjusted weight gain/loss for females as a factor of calories consumed, activity level, BMR, and weight measures over time.

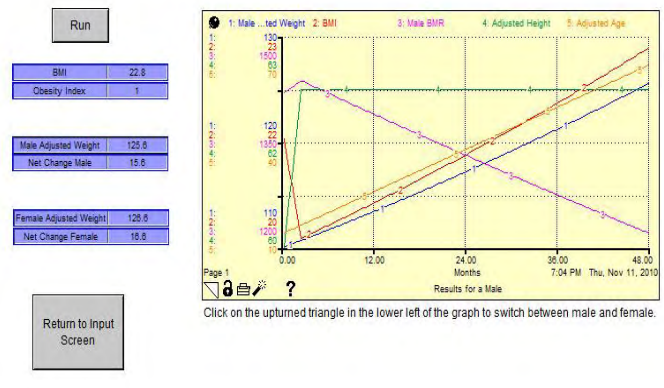


Figure 4. Output Variables

Instructor Support

The instructor packet which is delivered as part of the instructional simulation contains the guide and worksheet as one document. They have been separated in this design document for descriptive purposes for use by the final design team. The Activity Worksheet described in the instructor guide is contained below. The instructor guide is provided to the teacher of the high school health class; even though they should be able to instruct the students on how use this simulation with little trouble.

The teacher should assist the students in measuring their height, weight, and measuring their BMI accurately. The height can be measured with a tape measure, the weight can be measured with a simple scale, and the BMI can be measured using a skin fold thickness caliper. This caliper can be of any type such as "Figure Finders" 14W5058 from Ward's catalog at a price of $14.50. The teacher should ask the students for their permission and assistance in performing the BMI testing. It should be done by a female for female students and by a male for male students.

The activity worksheets contained in the instructor guide packet and shown in Figure 5 are used to record the student's height, weight, and BMI readings as described above. These items should be recorded under Part 1 of the worksheet. Direct students to run the simulation to determine their caloric intake and activity levels required to meet their desired ideal weight. Have students record the results under Part 2 of the worksheet. Additional copies of the worksheet may be provided as needed.

Questions concerning the instruction, model, interface, or any other aspects of the simulation may be directed to: Charles Reese at the Support Information and Additional Resources help desk available at crees008@odu. For additional instructional resources, you may refer to: Kosinski, B. (2005). Exploring weight regulation using *Stella* modeling software.

Name: ________________________

Exercise Your Heart and Your Mind: Student Worksheet

Part 1. Your teacher will help you measure your height, weight, and BMI. Record the measurements you get here.

Height: (inches) _________ Weight: (pounds) ________
BMI: _________

Part 2. Bring this worksheet when your class goes to the computer lab to run the simulation. When you get to the **"Your Turn"** part, use the tables below to record the values you try and the results. If one run through the simulation doesn't get you good results, just move on to the next copy of the table and try again.

Attempt 1

Age	Input (before you click Advance 5 Years)		Results (after advancing 5 years)	
	Calories Consumed	Activity Level	Weight	BMI
18				
25				
30				
35				
40				
45				
50				
55				
60				

Attempt 2

Age	Input (before you click Advance 5 Years)		Results (after advancing 5 years)	
	Calories Consumed	Activity Level	Weight	BMI
18				
25				
30				
35				
40				
45				
50				
55				
60				

Figure 5. Simulation Worksheet

IV. FUTURE WORK

Currently learners are instructed to record caloric intake and activity data on worksheets external to the simulation. Future design considerations include the recommendation to build these worksheets into the simulation to avoid a Cognitive Load Theory (CLT) split-attention effect between the worksheet and simulation. Therefore, a more suitable type of media may need to be used for the interface.

The model is an underlying representation of the phenomenon specific to be used within an instructional context. Currently two variables need definitive and validated sources from scholarly resources: Calories per Pound and the adjustment factor for changes in height. The latter variable may require the addition of another computational procedure based upon models for height prediction as a factor or parental height. Additionally, as many other factors are associated with accurately predicting weight gain and loss, the precision of the data does not include other complicating lifestyle or health and wellness factors. The model makes some basic assumptions about the learner, to include a standard degree of emotional, physical, and mental health and ability. These factors should be at least addressed as disclaimers within the instructional content.

The final design team shall be responsible for a complete formative evaluation, to include SME Review, learner talk-aloud protocol testing, and small group review. The results of these evaluations shall be referred back to the preliminary design team for assessment. The preliminary design team shall conduct as small scale field test of the product in support of a final production decision. To test the simulation in the short term, the teacher could monitor the students' behavior as it applies to their diet and exercise and measure changes in weight and BMI over the course of the health class. For long term testing, the simulation would have to be used in the manner listed above for short term testing but then check back with the students at regular intervals no later than each 10 year high school reunion. During this 10 year evaluation, the teacher can measure the students again and ask what impact the students feel the simulation had on their current overall level of health.

V. SIMPLIFIED VERSION

An Excel version of the simulation was also created to make it easier to use by others and to be able to explain the process better. The first spreadsheet, shown in Figure 6, lists all the variables and the outputs in chart form as the years go by. The second spreadsheet list all the formulas used in the first spreadsheet to explain the calculations to others as shown in Figure 7. Finally, there is a simplified model shown in Figure 8 explaining the impacts one variable has from the input through the output.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Initial Age	35				Time (years)	Current Age	Current Height	Current Weight	Current BMR	Current BMI	Net Change	Obesity Index
2	Initial Height	70				0	35	70	215	2060.75	30.93	0	3
3	Initial Weight	215				1	36	70	214.98	2053.84	30.93	-0.017357143	3
4	Maturing	1				2	37	70	214.97	2046.93	30.93	-0.017357143	3
5	Activity Factor	1				3	38	70	214.95	2040.02	30.93	-0.017357143	3
6	Calories Consumed	2000				4	39	70	214.93	2033.12	30.92	-0.017357143	3
7	Change in Years	1				5	40	70	214.91	2026.21	30.92	-0.017357143	3
8						6	41	70	214.90	2019.30	30.92	-0.017357143	3
9	You can change any of the above conditions					7	42	70	214.88	2012.39	30.92	-0.017357143	3
10						8	43	70	214.86	2005.48	30.91	-0.017357143	3
11	Activity Factor of 0 equals no exercise					9	44	70	214.84	1998.57	30.91	-0.017357143	3
12	Activity Factor of 1 equals walking level of exercise					10	45	70	214.83	1991.67	30.91	-0.017357143	3
13	Activity Factor of 2 equals running level of exercise					11	46	70	214.81	1984.76	30.91	-0.017357143	3
14	Activity Factor of 3 equals professional athlete level of exercise					12	47	70	214.79	1977.85	30.90	-0.017357143	3
15						13	48	70	214.77	1970.94	30.90	-0.017357143	3
16						14	49	70	214.76	1964.03	30.90	-0.017357143	3
17	Growing	0				15	50	70	214.74	1957.12	30.90	-0.017357143	3
18	Male Gaining	-0.017357143				16	51	70	214.72	1950.21	30.89	-0.017357143	3
19	Male Metabolizing	-0.017357143				17	52	70	214.70	1943.31	30.89	-0.017357143	3

Figure 6. Excel Spreadsheet #1

	A	B	C	D	E	F
1	**Initial Age**	35		**Time (years)**	**Current Age**	**Current Height**
2	**Initial Height**	70		0	=B1	=B2
3	**Initial Weight**	215		=D2+$B$7	=E2+$B$4*$B$7	=F2+$B$17*$B$7
4	**Maturing**	1		=D3+$B$7	=E3+$B$4*$B$7	=F3+$B$17*$B$7
5	**Activity Factor**	1				
6	**Calories Consumed**	2000				
7	**Change in Years**	1				
8				**Current Weight**	**Current BMR**	
9	**You can change any of the above conditions**			=B3	=66+(13.75*(D9/2.2))+(5*(F2*2.54))-(6.8*E2)	
10				=D9+$B$18*$B$7	=66+(13.75*(D10/2.2))+(5*(F3*2.54))-(6.8*E3)	
11	Activity Factor of 0 equals no exercise			=D10+$B$18*$B$7	=66+(13.75*(D11/2.2))+(5*(F4*2.54))-(6.8*E4)	
12	Activity Factor of 1 equals walking level of exercise					
13	Activity Factor of 2 equals running level of exercise					
14	Activity Factor of 3 equals professional athlete level of exercise					
15				**Current BMI**	**Net Change**	**Obesity Index**
16				=705*D9/(F2*F2)	=D9-D9	=IF(D16<18.5,0,IF(D16<24.9,1,IF(D16<29.9,2,IF(D16<39.9,3,4))))
17	**Growing**	=IF(E3<=16,1,0)		=705*D10/(F3*F3)	=D10-D9	=IF(D17<18.5,0,IF(D17<24.9,1,IF(D17<29.9,2,IF(D17<39.9,3,4))))
18	**Male Gaining**	=B19		=705*D11/(F4*F4)	=D11-D10	=IF(D18<18.5,0,IF(D18<24.9,1,IF(D18<29.9,2,IF(D18<39.9,3,4))))
19	**Male Metabolizing**	=(B6-(E9*B5))/3500				

Figure 7. Excel Spreadsheet #2

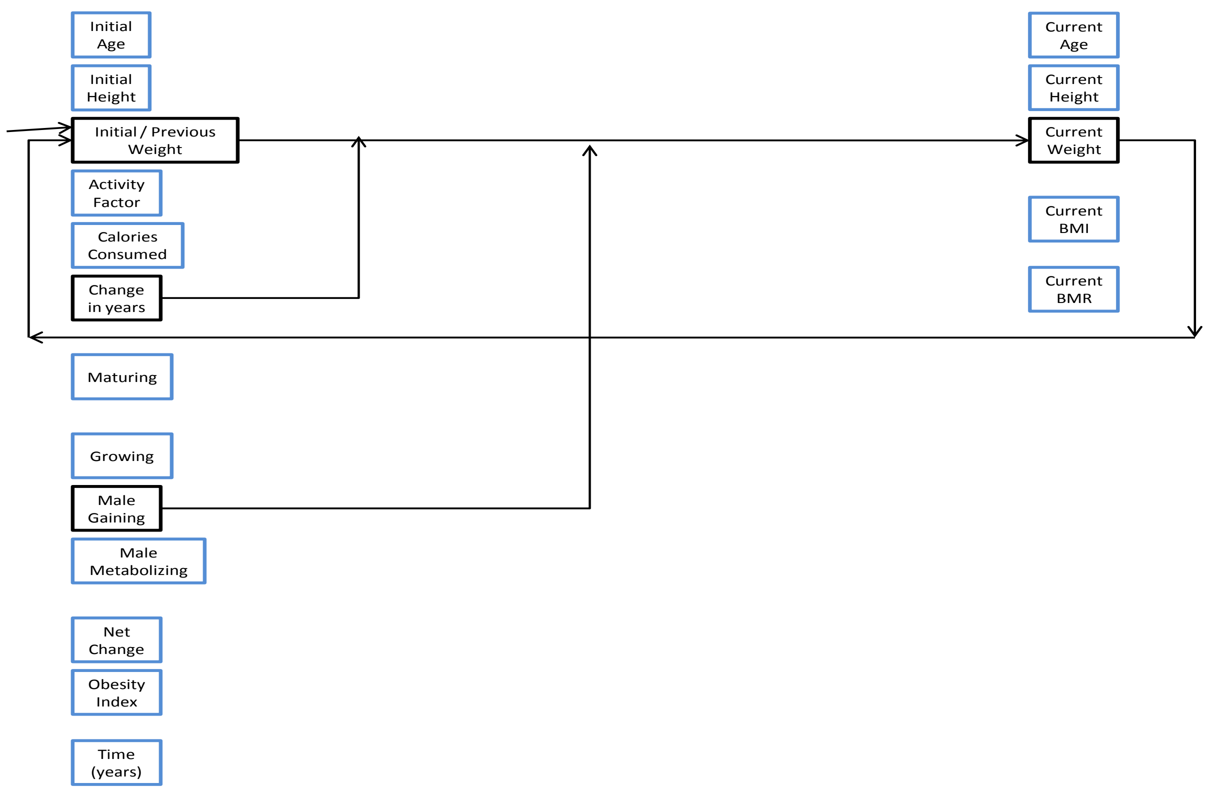


Figure 8. Simplified Excel Model

VI. CONCLUSIONS

This paper addresses some of the ways to help fight the obesity problem in the United States. It only deals with one main cause and only one way to fix that issue. To deal more effectively with this problem there needs to be more research and education into way obesity is increasing and ways of fighting. This simulation only focuses on teens making poor choices with diet and exercise as the cause and the potential solution to this problem. However, even this simple approach has many flaws such as with the equations, many factors that are directly related to weight gain and level of health are not addressed.

This simulation is a success because if used in its intended manner, it will inform high school health class students that there is an issue. The way to fix this problem is through knowledge and the simulation provides a place to see the actions or inactions first hand without experiencing them in real life. Also it shows the impacts over time that are not so apparent at their current age.

Although this paper and associated simulation tackled only a small part of the problem, it certainly brought the issue to light in a manner that is more understandable for teens in a format that they are more likely to learn from. I hope that this idea is refined to the point that it can be used in a health class setting. Hopefully one day, strives in education such as this may make a difference in reducing the problem of obesity.

REFERENCES

[1] About.Com. (2010). Using a BMI Chart for Obesity Management. Retrieved from http://diabetes.about.com/od/nutrition/a/BMIchartobesity.htm

[2] BMI Calculator. (2010). BMR Calculator. Retrieved from http://www.bmi-calculator.net/bmr-calculator/bmr-formula.php

[3] BMI Calculator. (2010). Harris Benedict Equation. Retrieved from http://www.bmi-calculator.net/bmr-calculator/harris-benedict-equation/

[4] Campbell, R., & Monson, D. (1994). Building a goal-based scenario learning environment. *Educational Technology*, 34(9), 9-14.

[5] Choi, W. (1997). Designing effective scenarios for computer-based instructional simulation: Classification of essential features. *Educational Technology*, 37(5): 13-21.

[6] STELLA (Version 9.1.4) [Computer Software]. Lebanon, NH: isee systems, inc.

[7] Kosinski, B. (2005). Exploring weight regulation using *Stella* modeling software. Pages 419-423, *in* Tested Studies for Laboratory Teaching, Volume 26 (M.A. O'Donnell, Editor). Proceedings of the 26th Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 452 pages. Retrieved from http://www.ableweb.org/volumes/vol-26/mini.Kosinski.pdf

[8] Calories. (2010). Determining how many calories equal 1 pound. Retrieved from http://www.caloriesperhour.com/tutorial_pound.php

[9] Calories. (2010). Determining how many calories equal 1 pound. Retrieved from http://www.annecollins.com/weight-loss/calories-per-pound.htm

Use of Architecture Modeling for Insertion of New Technology

Kevin Michael, Old Dominion University
Dr. Andreas Tolk, Old Dominion University

***Abstract* – This paper examines the issues involved in the insertion of new technology for long life-cycle, complex, mission critical legacy systems. Many of these systems are part of a system of systems or family of systems, which further increases the complexity. As technology advances rapidly (Moore's Law) legacy systems must be able to incorporate technological advances in order to improve their efficiency and effectiveness while avoiding obsolescence. The unavailability of hardware components and software updates (obsolescence) requires expensive reverse engineering, or expensive replacement of entire systems. Many approaches to this problem have been developed—some successful and some not so successful.**

The purpose of this paper is to address modeling approaches in order to address this rapidly growing problem. Specifically, this paper will focus on those systems which have a life-cycle measured not in months or years, but rather in decades, and will focus further on those systems that have mission critical implications, although the approaches examined may also apply to non-mission critical systems. There are many examples of such systems, both commercial and Government.

***Index Terms* – Architecture modeling, Technology Insertion, Technology Infusion, Mission Critical Systems**

I. INTRODUCTION

Inserting new technology in legacy systems is not only desirable, but in many cases is critical. One example is updating the technology in U.S. combat and weapon systems in order to maintain superiority, or at least parity, with potential adversaries. There are several pertinent questions to be answered. What are the most effective Engineering Management and Systems Engineering approaches for inserting new technology into legacy systems? What are unique applications of these approaches that can be applied to systems under different circumstances and with different requirements? What new guidelines can be established for technology insertion? What are the technology insertion implications for systems having a proprietary versus open systems architecture?

II. SYSTEM OF SYSTEMS AND FAMILY OF SYSTEMS

Keating et al, [1] define a system of systems (SoS) as "The design, deployment, operation, and transformation of metasystems that must function as an integrated complex system to produce desirable results. These metasystems are themselves comprised of multiple autonomous embedded complex systems that can be diverse in technology, context, operation, geography, and conceptual frame." Baldwin [2] describes SoS engineering within the Department of Defense (DoD) as "… planning, analyzing, organizing, and integrating the capabilities of a mix of existing and new systems into an SoS capability greater than the sum of the capabilities of the constituent parts."

While a system of systems represents a group of tightly coupled systems, with well-defined interfaces, all working together to accomplish a single purpose, a family of systems (FoS) represents a more loosely coupled group of systems. The systems in a FoS rely on one another to accomplish their individual purpose; however they may not have well defined interfaces or share a common purpose.

An example of a SoS is an integrated weapon system. The SoS may be made up of the weapon itself, the weapon control system that performs mission planning and interfaces with the launcher, and the Command & Control (C2) system that receives and distributes higher level plans and manages communications. An example of a family of systems is an entire weapon system, coupled with a combat control system (e.g., Aegis), the associated sensor systems (e.g., RADAR), the associated launch systems, etc. In a FoS, the individual systems may be used for purposes other than supporting the specific system under consideration. For example, the Space Shuttle itself is a SoS, however in order to accomplish its mission it relies on other system that serve other purposes, such as satellites and ground stations.

III. BACKGROUND

While the author is most familiar with United States DoD combat systems and weapon systems, other types of systems will be examined as well. The approach and success of technology insertion into these systems will be compared and evaluated. Some have enjoyed a moderate or high level of success while some have suffered a more dubious record of "success."

Several studies and examples of technology insertion are available for this research paper and will be examined. The Management of Accelerated Technology Insertion (MATI) consortium, founded in 1996, provides a forum for commercial, university and Government organizations to collaborate on strategic, rapid, timely and integrated technology innovation. Several specific efforts at technology insertion are described in the following paragraphs.

The Acoustic Rapid COTS Insertion (ARCI) Sonar system was a successful effort by the U.S. Navy to provide "significant, rapid improvements in acoustic sensors and processing by using commercial technology" as well as "substantially reduced costs with significantly improved processing capability" [3]. The Navy has attempted the insertion of new technology in numerous other programs, including surface ships, missiles, computer networks and financial systems. The results have generally been successful, although cost and schedule overruns have been encountered.

Another Navy initiative, the Navy Marine Corps Intranet (NMCI), has over the past decade replaced most non-tactical computer networks, servers, workstation, laptops and productivity software such as Microsoft Office. Technology refresh was one of the major driving factors behind NMCI; however the results have been mixed. While some Navy activities initially received better technology and service than before NMCI, the general result across the Navy has been a slower technology refresh, with the "current" NMCI workstations running Windows 2000, Windows XP, Microsoft Office 2003, Microsoft Office 2007 and Internet Explorer v6 or v7 – all one or more versions behind the current software releases.

The Federal Aviation Administration (FAA) for decades now has had difficulties modernizing the nation's air traffic control system for several reasons. As the GAO stated in one report [4] "The Federal Aviation Administration's (FAA) multibillion-dollar effort to modernize the nation's air traffic control (ATC) system has resulted in cost, schedule, and performance shortfalls for over two decades and has been on GAO's list of high-risk federal programs since 1995". Problems modernizing the nation's air traffic control system continue to this day.

These are but a few examples of the difficulty of successfully performing new technology insertion for long life-cycle, complex, mission critical legacy systems. Successfully transferring new technology requires emphasis at the top level of the organization, while Engineering Managers and Systems Engineers must address a large number of implications which will be examined in detail in this research paper. According to Sandborn and Singh [5], "Methodologies are needed that address how to optimally design a system in order to minimize the cost of concurrently managing both inevitable obsolescence problems and technology insertion."

IV. APPROACH

While managing the process and engineering of inserting new technology, it is obviously important not to break the existing system or otherwise degrade its functionality, capability or performance. The process of inserting new technology in a long life-cycle, complex, mission critical legacy system must address numerous concerns, including:

- Identification of applicable technologies that can improve system functionality, capability and performance, or reduce cost.

- Management of requirements to ensure functionality, capability and performance are maintained or enhanced.

- Effects on logistics, such as procuring and maintaining inventory of new technology (e.g., hardware) for future system support.

- Provision of training for users and support personnel.

- Testing and analysis to ensure reliability, maintainability and availability are not adversely affected.

- Cost analysis to ensure the new technology either provides better performance, capability or functionality at the same or lower price, or any cost increase is justified by improved performance, capability or functionality.

- Security posture to ensure the system is not compromised by introduction of the new technology.
    - Frequently security vulnerabilities in new technology begin arising shortly after introduction of the new technology, especially with COTS software.
    - The introduction of new technology into a legacy system can provide an attack vector for an adversary to sabotage a system, such as a combat or weapon system. Increasing reliance on foreign produced hardware containing firmware and microcode represents a potential risk. (This is a topic worthy of a separate research paper.)
- Interoperability, which must be maintained since the systems that are the topic of this research paper are generally part of a system of systems or family of systems.
- User acceptance, which is essential; inserting new technology that the system users cannot or will not accept, will be of no value. This will be maximized in cases where the new technology is transparent to end users, or where the new technology reduces the complexity or improves the user interface.
- Architecture modeling, which is a key aspect of new technology insertion as this will sometimes require changes to the system architecture. Introducing or maintaining an open architecture is essential for the effective insertion of new technology over the life cycle of a system.

V. OPEN SYSTEM ARCHITECTURE

Open system architecture is important to achieving rapid technology insertion, whether for legacy systems or new systems. According to Davis [6] “Defense dollars are becoming scarcer and technology is advancing at an ever-increasing rate. This combination is putting increasing pressure on the available acquisition budget. To combat this, the Department of Defense (DoD) has moved toward commercial-off-the-shelf (COTS) equipment and open system architecture (OSA) to reduce acquisition cost, shorten development time and allow rapid technology insertion into existing avionics systems.”

The Air Force has hands-on experience with using open systems architecture to combat obsolescence. Hardenburg [7] states “The replacement of legacy ATE with commercial ATE, however, is only the start of the problem. Once an ATE is upgraded, the problem remains with constantly emerging technology that promotes rapid obsolescence.”

Several approaches to technology insertion are already in practice. Cooperative Research and Development Agreements (CRADAs) are becoming more common between Government and industry. Gallagher [8] states “CRADAs are highly cost-effective from the government’s point of view. Participating companies fund the research efforts, but a CRADA gives technology developers an inside view of the agency’s needs, and as a result they are able to tailor their products to meet them, and by extension the needs of other potential government customers.”

VI. METRICS

While the large sample size of these systems will preclude providing an exhaustive list, the approaches derived can be categorized, summarized and evaluated. For each individual approach identified, a representative sample of projects can be selected and metrics generated. The representative samples selected can be selected such that each approach will have a mix of projects of similar cost, life cycle and complexity. Each approach can be scored on how well they accomplished the following criteria, as described in more detail above. A proposed scoring system to be used is as follows:

5: Outstanding

4: Very Good

3: Good

2: Acceptable

1: Marginal

0: No Benefit

-1: Negative Effect

The scoring method for each criterion can be documented before applied to the individual technology insertion approaches. For example, Effects on Logistics may be rated -1 (Negative Effect) if the approach means obsolete parts will have to remain in inventory because the new technology cannot be configured as a drop in replacement and at the same time adding the new technology to the logistics requirements of the sample system will introduce extensive new logistics requirements. Conversely, Effects on Logistics may be rated 5 (Outstanding) if the approach means the new technology can be quickly integrated into the supply chain with minimal installation/configuration effort to the system. The following is an example of how various technology insertion approaches may be evaluated based on specific metrics.

Evaluation Criterion	Spiral Devel.	Rules Engines[1]	SBIR[2]	CRADA
Cost Savings				
Schedule Relative to Development				
Identified Technologies				
Requirements Management				
Effects on Logistics				
Requirements for Training				
Reliability, Maintainability and Availability				
Affect on Security Posture				
Level of Interoperability				
User Acceptance				
Use of Open Architecture				
Composite Score				

[1] Rules Engines are an alternative to hard coding logic into computer programs. They allow a domain expert to develop complex rules and logic outside of the software, which are then compiled and executed at run time by a rules engine. Rules Engines provide a consistent and visible basis for decisions made during software execution. While rules engines have been used for business applications, they are being evaluated for combat systems.

[2] SBIR is the Small Business Innovative Research Program, which focuses on development of new technology as well as improvements to existing technology. SBIR awards impose strict cost and schedule limits for Phases I and II, with Phase III providing for a transition to commercialization of the technology.

The evaluations can include a quantitative score indicating the overall level of success for each evaluation criterion for each approach. Statistical analysis will be performed on individual approaches, and the author will attempt to identify and estimate the combined success of complimentary combinations. This may be accomplished by comparing the individual evaluation criterion scores for each approach and selecting the “best of breed”, then analyzing for any inconsistencies or overlaps. In this way, best practice for recommended approaches can be developed.

Metrics, such as those described above, can be developed that will provide a basis of comparison between the different technology insertion approaches. Important categories of metrics for each approach include:

- Cost: percentage relative to the overall development cost of the system
- Schedule: absolute time as well as time relative to development time of the system
- Capability Improvements: assessment of system performance improvements
- Regulatory/Statutory Requirements: system dependent; examples include interoperability with the Global Information Grid (GIG) and compliance with Information Assurance mandates such as the Defense Information Assurance Certification and Accreditation Process (DIACAP).

Selected criteria may be examined in detail across the different approaches for similar technology. For example, it may be possible to compare the historical cost of improving storage capacity in a combat control system by a factor of ten and performance (throughput) by a factor of three using the various approaches identified. This may provide an historical analysis showing the method that would provide the most cost savings.

The quantitative analysis may be accompanied a qualitative analysis consisting of selected case studies to illustrate how the quantitative scores were derived and to provide further insight into the particular technology insertion approaches that work particularly well, or in some cases not so well.

New technologies to be considered will be dependent upon the type of system under consideration, and will include both current and emergent technologies. Examples of technology insertion may include blade servers, storage area networks, network attached storage, solid state storage devices, virtualization, service oriented technology, knowledge management, high density multi-core multi-processor systems, parallel processing and quantum computing. While these examples may not seem to be the latest start-of-the-art, this research paper will be focused on legacy projects decades in duration, many of which have already been in operation for years.

Currently, it is very difficult to perform rapid technology insertion for long life-cycle, complex, mission critical legacy systems. However this is critical to provide required performance at reasonable cost, and in many cases Government and industry are falling further and further behind. This can be especially true for combat and weapon systems, as well as other types of systems where obsolescence or poor performance relative to adversaries can lead to catastrophic loss of life, equipment or capability. Providing an improved Engineering Management and Systems Engineering solution set to this problem has the potential to save lives, reduce cost, and improve mission capability

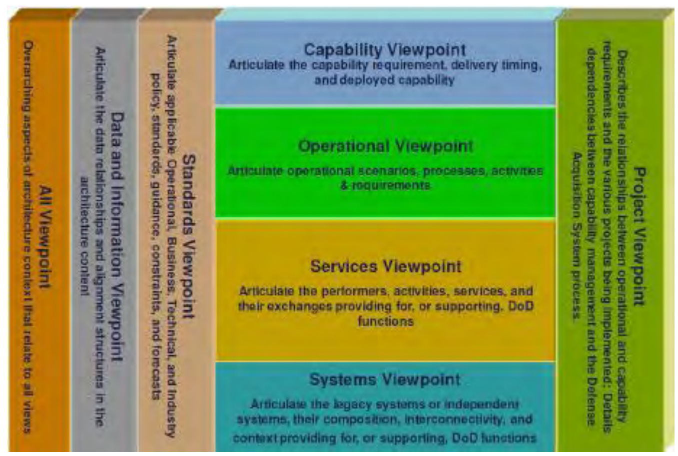
VII. MODELING AND ARCHITECTURE

Architecture represents the earliest set of design decisions for a system. According to Bass, Clements and Kazman [9] "These early decisions are the most difficult to get correct and the hardest to change later in the development process, and they have the most far-reaching effects."

There are several methodologies for modeling system design and system upgrades. While developing a system architecture is accepted practice in the design of new systems, it is not always used for system upgrades – whether upgrading hardware or updating software. In cases where an existing system architecture does not exist or is outdated, creating or updating the architecture can be beneficial for inserting new technology.

When developing the system architecture, there are several frameworks available, each of which provide specific views or models. The accepted framework for DoD systems is the DoD Architecture Framework (DoDAF). The DoDAF is closely aligned with the Ministry of Defence Architecture Framework (MODAF) used by the United Kingdom. The DoDAF has evolved since its introduction. The DoDAF v1.0 included All Views, Operational Views, System Views and Technical Standards Views. The DoDAF v1.5 added Services Views to describe the use of software (e.g., web) services in the architecture. The DoDAF v2.0 has added Capabilities Viewpoints (DoDAF v2.0 renamed views to viewpoints), Data and Information Viewpoints and Project Viewpoints.

The DoDAF v2.0 architecture viewpoints are shown below, and include the products specified in previous versions of the DoDAF.



One of the main differences between the DoDAF v2.0 and previous versions is in the underlying structure of the architecture models, or views. While the DoDAF does not prescribe a specific modeling methodology or language, such as Universal Modeling Language (UML) or Systems Modeling Language (SysML), previous versions of the DoDAF were better addressed by structured analysis such as IDEF0, whereas the DoDAF v2.0 is better addressed by object oriented design. Structured analysis uses functional decomposition, while object oriented design uses entity definition. These are very different design approaches, and therefore require a different modeling approach and modeling tools.

It is advantageous where possible to use an object oriented methodology with the DoDAF v2.0. The predominant object oriented methodology in use for DoDAF v2.0 compliance is UML, with SysML rapidly gaining in popularity. UML provides the ability to perform analysis, design, and implementation of software based systems. SysML provides support for the specification, analysis, design, verification, and validation of a broad range of complex systems including hardware, software, information, processes, personnel, and facilities. SysML provides systems engineering extensions to UML is intended to address a broad range of systems beyond software.

When developing or updating the system architecture for new technology insertion, several of these models, views and viewpoints are helpful. It should be noted that the DoDAF does not require or describe several key architectural products including operational scenarios, mission threads and use cases. These additional architecture products are very useful in developing the other products. The operational scenarios describe the specific missions in which a system will be employed, including force structure, doctrine and tactics, chain of command, and operational details. The mission threads provide a time sequenced series of steps to accomplish the mission, and include additional information such as operational nodes/actors and assumptions. The use cases, unlike the scenarios and threads, are functional. They provide the steps required to perform system functions, used throughout the scenarios and threads, to accomplish discrete functions such as planning a mission or launching a weapon.

Some of the DoDAF v2.0 viewpoints are critical when implementing new technology insertion. Operational Viewpoints are critical to ensure that the new system architecture will support the operational environment. This includes the overall operational scenario, the command structure (both chain of command and information passed between command elements) and the operational activities employed by the system users. System Viewpoints are critical to show the interfaces between system elements and components – mainly hardware. Services Viewpoints are critical to show the interaction between software components. Many of the remaining viewpoints are essential as well, and provide supporting information for the system implementation.

VIII. CONCLUSION

Modeling of the system is necessary when inserting new technology. This helps to verify that it will meet the operational needs of users, that the new hardware and software will work as intended through the systems and services viewpoints, and that the system overall will function as intended.

Modeling a system that has been in existence for many years or decades can present additional challenges. Normally the architecture of a system must be developed, documented and may be evaluated prior to system design and development. In some cases the system architecture was never developed, never properly documented, or the system has evolved beyond the documented architecture. In these cases, reconstruction of the current architecture will be necessary prior to inserting new technology into the system. This involves documenting the "as-built" architecture of the existing system. The reconstructed architecture can then be used for re-documenting the architecture, analyzing the architecture and re-engineering the system. In this way the reconstructed architecture can serve as the basis for system analysis and insertion of new technology.

REFERENCES

[1] Keating, C., Padilla, and Adams, K. (2008). System of Systems Engineering Requirements: Challenges and Guidelines. Engineering Management Journal, Vol. 20, No. 4

[2] Baldwin, J.K. (2008). Systems Engineering Guide for Systems of Systems. Office of the Deputy Under Secretary of Defense for Acquisition and Technology. United States Department of Defense. Ver. 1.0 http://www.acq.osd.mil/sse/docs/SE-Guide-for-SoS.pdf

[3] Lockheed Martin Corporation (2003). Acoustic Rapid COTS Insertion (ARCI). Lockheed Martin Maritime Systems & Sensors: Manassas, VA.

[4] United States Government Accountability Office (November 2004). Report to the Chairman and Ranking Minority Member, Committee on Government Reform, House of Representatives; GAO-05-23: Air Traffic Control.

[5] Sandborn, Peter and Singh, Pameet (2005). Forecasting Technology Insertion Concurrent with Design Refresh Planning for COTS-Based Electronic Systems and CALCE. Dept. of Mechanical Engineering, University of Maryland. Proc. Reliability and Maintainability Symposium. Arlington, VA.

[6] Davis, D.E. (1999). Support of COTS and open system architecture modules [in military avionics systems]. IEEE Systems Readiness Technology Conference, Page(s): 719 – 722: San Antonio, TX.

[7] Gallagher, Sean (2009). U.S. mapping agency retools cooperative research with industry. C4ISR Journal. VA: Gannett Corporation. June 03, 2009.

[8] Hardenburg, G.L. (1999). Successful migration of a legacy system to an open architecture. IEEE Systems Readiness Technology Conference, Page(s): 405 – 410: San Antonio, TX.

[9] Bass, L., Clements, P. and Kazman, R. (2003). Software Architecture in Practice. Carnegie Mellon Software Engineering Institute. Pearson Education, Inc: Boston, MA
[10] Buede, D. (2009). The Engineering Design of Systems. John Wiley and Sons: Hoboken, NJ
[11] Clements, P., Kazman R. and Klein, M. (2002). Evaluating Software Architectures. Carnegie Mellon Software Engineering Institute. Addison-Wesley: Westford, MA
[12] Clements, P., et al. (2003). Documenting Software Architectures, Views and Beyond. Carnegie Mellon Software Engineering Institute. Pearson Education, Inc: Boston, MA
[13] Levis, A., Wagenhals, L. and Zaidi, A. (2010). DoDAF Architecture Framework Implementation. AFCEA Coursebook 503-BO. Armed Forces Communications and Electronics Association: Fairfax, VA
[14] Katina, P. (2010). Modeling & Simulation: A System of Systems Engineering View. Old Dominion University: Norfolk, VA
[15] Object Modeling Group. Systems Modeling Language (SysML), Version 1.2. June 2010.
[16] Object Modeling Group. Unified Modeling Language (UML), Version 2.3. May 2010.
[17] United States Department of Defense. DoD Architecture Framework, Version 1.5. April 23, 2007.
[18] United States Department of Defense. DoD Architecture Framework, Version 2.0. May 28, 2009.

COMPARISON OF CONCEPTUAL MODELING FRAMEWORKS TO FACILITATE STANDARDIZATION

Rabia Haq, Old Dominion University

ABSTRACT - No set of standard tools and languages yet exist to facilitate the conceptual modeling instantiation, communication and validation and verification process, rendering the conceptual modeling design process subjective to the understanding of the developers. This paper reviews three discipline-independent conceptual model development frameworks to establish a list of criteria that is the recommended steps towards formulation of a concept model. Identification of essential stages of conceptual model development frameworks is addressed to formulate a synthesis of conceptual model development requirements and a cohesive technique of design that can be applied across various disciplines.

INDEX TERMS - conceptual modeling framework, overview, standardization.

I. INTRODUCTION

The development of a conceptual model is widely being accepted by the research community as an imperative step in simulation development to facilitate simulation fidelity and reduce development costs. Due to the lack of a standardized conceptual modeling framework, researchers in various disciplines have been motivated to formulate a conceptual model in accordance with, and pertaining specifically to, the requirements of the simulation project in a particular discipline and its problem scope under consideration.

Conceptual models are widely utilized in translating user requirements from domain experts to simulation developers. They facilitate the requirements analysis process to mitigate simulation error and reduce cost, whereby increasing simulation validity and feasibility for intended use. It is also used as part of the validation and verification process of simulation model development to ensure the resulting simulation is acceptably accurate in the specified problem domain.

No set of standard tools and languages yet exist to facilitate the conceptual modeling instantiation, communication and validation and verification process, rendering the conceptual modeling design process subjective to the understanding of the developers [1]. Therefore, simulation modeling is considered to be both art and science [7].

This paper reviews three discipline-independent conceptual model development frameworks to establish a list of criteria that is the recommended steps towards formulation of a concept model. The reviewed frameworks are then compared to establish commonalities and differences in these approaches that can be extended towards future research for determination of a single concept modeling framework. Identification of essential stages of conceptual model development frameworks is addressed to formulate a synthesis of conceptual model development requirements and a cohesive technique of design that can be applied across various disciplines.

Section II conducts a review of three discipline-independent conceptual modeling frameworks. Section III enumerates the various steps and criteria that can be assimilated to formulate a conceptual model, and uses this criteria to compare and contrasts the implementation of these steps in the framework under consideration. Section IV provides insight into future work and Section V concludes the paper.

II. CONCEPTUAL MODELING FRAMEWORKS RESEARCH REVIEW

II.1 Simulation Conceptual Model Construct [2]

[2] introduces a conceptual modeling framework developed in the military domain with emphasis on the development of conceptual models for the purpose of simulation validation and verification. The author defines the conceptual model as "collection of information that describes a simulation developer's concept about the simulation and its pieces" [2]. It is a tool utilized for transformation of modeling requirements into a detailed design framework through which simulation implementation details can be extracted. It is assumed that the developer creates the conceptual model to facilitate simulation development and aid in mitigation of the development cost and an increase in simulation fidelity.

Key elements of all development paradigms have been identified as shown in Figure 1, extracted from [2]. This paradigm identifies *Reality* from which the *Problem Statement* via application of constraints on reality is formed. The *Conceptual Model (CM)* can be extracted from *Reality* through a simulation context with intended use and requirements. *Fidelity referent data* can also be extracted from *Reality*, which does not have an intended use in relation to the simulation context and is utilized for the verification of design and implementation to the *Specifications Design Implementation*. The tools that are required for development of simulation and algorithms are extracted per the requirements of the *Problem Statement*, and are utilized for formation of the *Simulation Requirements*. These requirements are further used for verification and validation of the CM.

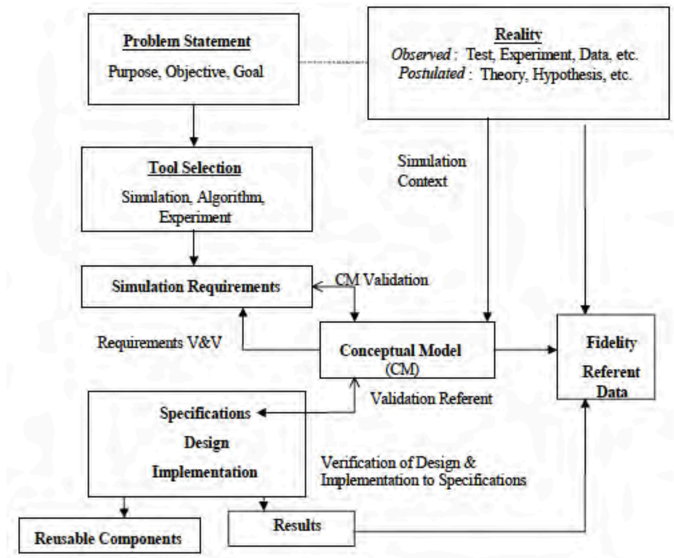


Figure 1: Simulation Development Elements from [2]

A four-stage approach towards conceptual modeling development is adopted in this conceptual modeling framework:

1. collecting authoritative information regarding the problem domain
2. identification of entities and processes that need to be addressed (model decomposition)
3. identification of simulation elements (model abstraction)
4. identification of simulation functions and relationships between elements.

The author proposes this to be an iterative process to achieve model completeness with respect to the intended use and requirements of the simulation.

i. Collecting Authoritative Information regarding the Problem Domain

This is the first of the four-stage development process of a simulation conceptual model. It is asserted to be an iterative process of requirements gathering whereby the simulation problem domain is communicated from the subject matter experts (SMEs) to the simulation developer. It is vital that the personnel involved in the process have a common level of communication to mitigate information and requirements ambiguity that may later result in a development error.

ii. Identification of Entities and Processes (Model Decomposition)

This stage of the model development is the process of identification of simulation elements via decomposition of the conceptual model and requirements analysis. The author provides a six-item checklist for conceptual model decomposition [2] that can be followed to ensure that the element extraction process is complete and concise.

iii. Identification of Simulation Elements (Model Abstraction)

This third step in modeling development determines the scope of the simulation. The process of model abstraction is detrimental towards achievement of simulation element accuracy and precision. Various knowledge engineering abstraction techniques and information modeling languages, such as Unified Modeling Language (UML) are adopted to represent the correct level of simulation detail and precision. Various criteria of simulation quality, such as correctness, completeness, consistency and coherence have been used to evaluate model abstraction.

iv. Identification of Simulation Functions and Element Relationships

This stage of conceptual modeling development ensures that the simulation to be developed is limited to the bounds and constraints imposed by the problem domain. The author also proposes documentation of the various stages of the conceptual modeling development process to facilitate validation and simulation reuse.

II.2 Simulation Model Paradigms [6]

[6] establishes two simulation model paradigms for the purpose of simulation model verification and validation to the model development process. Therefore, the author has a different perspective of conceptual model development with the intention of development of a model used for verification and validation of the resulting simulation model. The development of a conceptual modeling framework with this purpose of use as a V & V tool has resulted in the development of a more detailed paradigm . This paradigm describes in detail the modeling process components and their interaction and transformation between each other that aid in the verification and validation of not only the simulation model (what has been built) but also the conceptual model (what is to be built). This facilitates the determination of the level of model validity for its determined purpose. [6] provides four approaches for determining simulation model validity and provides a simplified and a detailed method of verification and validation to the simulation model development process. Figure 2 describes the detailed paradigm consisting of the Real World and the Simulation World.

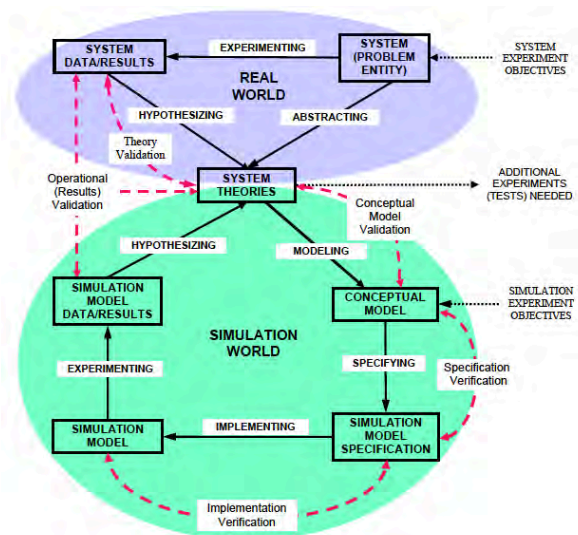


Figure 2: Real World and Simulation World Relationships and Interactions with Verification and validation from [6]

Figure 2 shows the Real World consisting of a Problem Entity which can be abstracted to form System Theories. A set of System data/results can be obtained from the Problem Entity via experimentation if possible. These System Theories form the synthesis point for the Real and the Simulation World. The System Theories can be modeled for the objectives of the simulation study to form a Conceptual Model, which has been described as "the mathematical/logical/verbal representation (mimic) of the system developed for the objectives of a particular study" [6]. The Simulation Model Specification is a detailed document describing the software design detail and specification that can be utilized as a specifications documentation for the development of the Simulation Model. The Simulation Model is an executable representation of the model that can be executed to obtain a set of Simulation Model data/results through experimenting. This detail decomposition of the system theories and models allows a more detailed validation and verification process at various stages of development. Similar to other frameworks, this paradigm describes an iterative development process.

The Conceptual model validation between the System Theories and the Conceptual Model is the determination of the underlying theories and assumptions required for formulation of a model conformed to the established simulation intended use. Specification verification between the Conceptual Model and the Simulation Model Specification document is conducted to ensure that all assumptions and theories identified in the conceptual model have been correctly specified in the resulting document. Implementation Verification is the process of ensuring that the simulation development has followed the specification document. Operational (Results) Validation determines that the simulation output results conform to sufficient accuracy with respect to the simulation intended use and model's intended functional requirements.

[6] proposes various face validation and statistical validation techniques to test the underlying simulation assumptions, whereby describing verification and verification a primary purpose of a simulation model development.

II.3 Conceptual Model Design [3]

[3] explores a conceptual modelling framework with the purpose of defining the problem scope and level of detail required for simulation model development. The vast impact of conceptual model development on the problem domain and requirements and the efficiency of the development process is recognized. It is established as an aiding tool towards a more complete and valid simulation development within the allotted project time and budget.

A conceptual mode has been described as "a non-software specific description of the simulation model that is to be developed, describing the objectives, inputs, outputs, content, assumptions and simplifications of the model."

[3] describes four conceptual modeling requirement qualities of validity, credibility, utility and feasibility that can be employed to determine whether the conceptual model under construction fulfills the quality requirements in relation to simulation intended use. A framework for developing a conceptual model has been introduced that can be used a guiding tool for

1. understanding the problem situation
2. determining objectives and requirements
3. identification of inputs (experimental factors)
4. identification of outputs (responses)
5. the extraction of problem scope and level of detail of the project.

Figure 3 taken from [5] describes the revised conceptual modeling framework of [3].

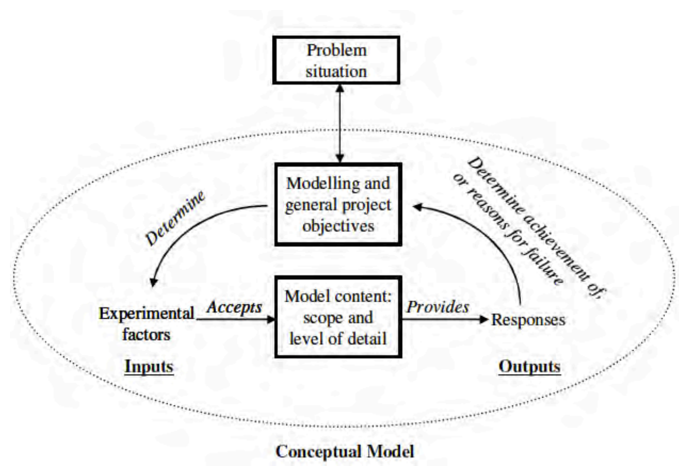


Figure 3: A framework for Conceptual Modelling from [5]

Figure 3 describes a problem situation, similar to the problem domain in other frameworks, which is outside the scope of the conceptual model. The first stage of conceptual modeling development is the understanding of the problem situation. A set of modeling and general project objectives can be extracted from the problem situation. These objectives can be used to determine the conceptual model by identification of the inputs and then the outputs of the conceptual model, leading to the determination of the model content in relation to scope and level of detail. Any underlying assumptions and abstractions are identified throughout the development process. The extracted set of outputs can be further utilized to determine whether the modeling objectives originally set have been achieved or not. This framework provides another cyclical process that is improved over various iterations.

i. Understanding the Problem Situation

[3], [5] addresses the involvement of the SMEs and domain experts for establishment of the problem situation. Development of a good understanding of the cause and effects of the system is vital for extraction of a model scope and level of detail. Formal and informal problem structures, such as soft systems methodology, causal loop diagrams and cognitive mapping have been recommended for better understanding the problem scope [3].

ii. Determining Objectives and Requirements

The objectives and requirements elicitation should be approached with three elements under consideration: what the clients want to achieve, the level of performance required and elicitation of constraints.

iii. Conceptual Model Design: Inputs and Outputs

The initial step towards conceptual model design is the elicitation of the inputs and outputs of the conceptual model. The range of acceptable values for the inputs and outputs should also be noted to aid in establishment of a level of confidence of the resulting simulation model.

iv. Conceptual Model Design: Model Content

This step requires the modeler to consider the appropriate modeling approach, and whether simulation modeling is the correct approach for solving the problem domain. The key interconnections and functions transforming the inputs to the outputs in the real world should be recognized within the defined scope of the model. An appropriate level of detail of the model content should also be established that accurately represents the defined components and its interconnections within the model scope.

III. COMPARISON OF CONCEPTUAL MODELING PARADIGMS AND DEVELOPMENT PROCESS

III.1 Comparison of Modeling Paradigm Approaches

[2] explores a military domain-centric conceptual modeling paradigm approach which is aligned with the paradigm of High Level Architecture (HLA) widely used for distributed simulations. His conceptual modeling development approach is aligned with the need of a conceptual model for verification and validation of the resulting simulation with the extracted simulation requirements. This would not only increase simulation validity within the intended use domain, but would also increase simulation model confidence and decrease the overall development cost by mitigating the risk of simulation development error. This approach provides a clean step-by-step development technique for a conceptual model that would serve as an efficient translation tool between the problem domain requirements and the simulation.

[6] describes a detailed Simulation Model paradigm with Real and Simulation World interactions that has been utilized for verification and validation throughout the simulation development lifecycle. This level of development and validation process detail would ensure that the resulting simulation developed addresses the underlying assumptions of the model and provides acceptable model accuracy and confidence with respect to intended use.

[3] research succinctly describes the lack of research that addresses the development of valid simulation conceptual models, and establishes a need for formalization of a conceptual modeling framework. The definition provided for a conceptual model contains the key elements of the conceptual model, thus providing a more comprehensive conceptual model intended applica-

tion and use. A framework is originally established in [3]. This is currently being extended to assimilate various cross-disciplinary framework techniques, such as [4] and [5] to formulate the basis of a framework not limited by a research area or discipline. This could serve as a catalyst for establishment of a standard conceptual modeling framework through further research.

[4], [5] assimilate various implementations and approaches to formulate a more generic concept model. [4] establishes the need for recognizing concept model requirements – what is required of a conceptual model – and tabulates these requirements against various techniques to compare various conceptual model development approaches in discipline-centric research areas of engineering management and supply-chain management. It is noteworthy that [4] addresses the implementation of conceptual modeling quality requirements in various research areas.

This paper is comparing and contrasting the techniques by which various conceptual modeling frameworks approach the steps towards conceptual modeling design and development.

III.2 Comparison of Conceptual Modeling Development Approaches

i. Comparison Criteria

Conceptual modeling is a fundamental research topic in the simulation field. The importance of the development of a conceptual model within the simulation model lifecycle is being widely established, accepted and practiced in various research disciplines. Various conceptual modeling frameworks and techniques have been developed in specific areas of interest, such as healthcare, operations research, mechanical engineering, discrete event simulation and software engineering to address the specific requirements of the problem specific to the area of interest.

Various criteria that can be used for formulation of the conceptual model have been extracted and recognized as the synthesis of the various stages recommended for formulation of a conceptual model.

1. Conceptual model definition
2. Definition of model constraints and scope via abstraction. The conceptual model development entails an abstraction of reality, with applied constraints and boundaries. These may be quantifiable in terms of inputs and outputs or qualitative in terms of observation of certain behavior.
3. Knowledge gathering for collection of observed data
4. Conceptual model elements description
5. Conceptual model functions and objectives description
6. Conceptual model validation and verification
7. Implementation simplicity

Three discipline-independent conceptual modeling framework components and techniques are compared in this paper to formulate a commonality between the various framework development approaches that may facilitate further research towards a synthesis of a single discipline-independent conceptual model framework. This comparison can be further extended to extract necessary steps required for formulation of a conceptual modeling framework standard. Development of a common technique for conceptual model development and associated representative languages would result in improved model interoperability and simulation reuse.

ii. Conceptual Modeling Development Comparison

Table 1 compares the conceptual model development frameworks of [2], [6], [3] and [5]. It can be inferred from the table that frameworks that provide more complexity satisfy a wider criteria for the steps recommended for development of a conceptual model, but are more difficult to implement due to design complexity.

	[2]	[6]	[3],[5]
Conceptual Model Definition	Provided. Centric towards simulation developers view	Provided. Centric towards simulation model V & V process	Provided. Identifies concept model elements
Model Constraints Definition	Determination of Problem Statement	Development of System Theories	Modelling and general project objectives
Knowledge Gathering	Not explicit. Part of Fidelity Referent Data and Simulation Context	System Data/Results and Simulation Model data/Results	Not explicit. The Responses are used for comparison with the project objectives
Conceptual Model elements description	Simulation Context	Conceptual Model & Simulation Model Specification	Model inputs and outputs
Conceptual Model functions description	Simulation specifications	Conceptual Model & Simulation Model Specification	Model Content: scope and level of detail
Conceptual Model V & V	Requirements V&V, Validation referent, Verification of Design & Implementation to Specifications	Conceptual Model Validation, Specification Verification	Conducted via output results comparison and documentation
Implementation Simplicity	Moderate	No	Yes

Table 1: Comparison of conceptual model development frameworks

IV. FUTURE WORK

Future work by Robinson that addresses further steps in conceptual modeling can be included to expand this list of conceptual modeling development criteria. [1] refers to the establishment of a standard for evaluation of quality of conceptual models. The various practical issues addressed in this existing research can also be applied towards understanding the reasons for the lack of a standard conceptual modeling framework.

V. CONCLUSION

This paper aimed at review of three discipline-independent conceptual modeling framework development methods, establishment of a criteria common to these frameworks that may be utilized as a comparison criteria to determine commonalities and differences between these various approaches to aid in further research in a discipline-independent conceptual modeling framework development.

The criteria list introduced in this paper is by no means exhaustive, and has ideally remain simple to remain on the path of reaching a common objectives resolution, as a more detailed list of development criteria would require a more detailed framework, narrowing the interoperability and applicability of a framework across various research disciplines.

ACKNOWLEDGMENTS

Dr. Eric Weisel's outlook and understanding on the subject of conceptual modeling was a driving force in the synthesis of this paper.

REFERENCES

[1] Moody, D. L. 2005. Theoretical and practical issues in evaluating the quality f conceptual models: current state and future directions. In *Data and Knowledge Engineering*, 55:243-276.

[2] Pace, D. K. 2000. Simulation Conceptual Model Issues: Development Methods (Part 1), Interaction with Simulation Requirements (Part 2), and Simulation Development Costs and V&V Costs (Part 3). In *Proceedings of the 2000 Summer Computer Simulation Conference*, ed. W. F. Waite, 488-499. Vancouver, British Columbia, Canada.

[3] Robinson, S. 2004. Designing the Conceptual Model in Simulation Studies. In *Proceedings of the 2004 Operational Research Society Simulation Workshop*, ed. Brailsford, S.C., Oakshott, L., Robinson, S. and Taylor, S.J.E., 259-266. Operational Research Society, Birmingham, UK.

[4] Robinson, S. 2008a. Conceptual modelling for simulation Part I: definition and requirements. In *The Journal of the Operational Research Society* 59(3): 278-290.

[5] Robinson, S. 2008b. Conceptual modelling for simulation Part I: a framework for conceptual modelling. In *The Journal of the*

Operational Research Society 59(3): 291-304.

[6] Sargent, R. G. 2001. Some Approaches and Paradigms for Verifying and Validating Simulation Models. In *Proceedings of the 2001 Winter Simulation Conference*, 2:106-114.

[7] Shannon, R. E. 1975. *Systems Simulation: The art and Science*. Prentice-Hall, Englewood Cliffs, NJ.

AUTHOR BIOGRAPHIES

RABIA HAQ is a PhD student in the department of Modeling, Simulation and Visualization at Old Dominion University. She received her B.S. and M.S. degrees in Computer Science from Old Dominion University.

Agile Project Management Methodology for Transition to Next Generation Architecture

Rupesh P. Narkar
Engineering Management and Systems Engineering
Old Dominion University
Norfolk, VA 23529, U.S.A
rnark001@odu.edu

Current JLVC simulations employ resource heavy HLA and RTI to support the additions of federates, accounting to significant costs incurred while moving /upgrading each simulation into new versions of RTI or HLA. With this way ahead, upgrading existing federates and addition of new federates will result into high costs. Transition to an innovative next generation M&S architecture will solve this issue by leveraging benefits from next generation technologies. This transition process from the current architecture to NGA is a complex project which may reveal several management issues as a result of changing priorities, delays in decision, changes in requirements, and other uncertainties. The paper identifies and provides an initial guideline on applying an agile project management technique which would embrace change and help to better manage uncertainty for smooth transition from the current architecture to NGA.

1. Introduction

In order to effectively support the joint war fighting in future, the Next Generation Architecture Team (NGAT) (2010) recommends Joint Forces Command (JFCOM) to evolve its Joint Live Virtual Constructive (JLVC) environments which are plagued by tightly-coupled simulations that demand increasing resources for integrating new and upgrading existing federates.
As per the study and recommendations in the report, following are key points (NGAT, 2010),

1. The growth and demand of new federates are taxing, spurring the need for an orchestrated, well-researched way ahead. To optimize future resources, JFCOM must establish an M&S solution – an innovative way ahead – for a Next Generation M&S architecture to support independent services collaboration within a common framework.
2. A NextGen M&S architecture standard will resolve problems associated with tightly coupled simulations within a federation, including inflexibility with change and composability, achieving a more efficient "loosely-coupled" platform resulting in significant cost savings.
3. Designed in three phases, the strategy gently transitions the existing architectures from being problematic to being proficient. Phase 1 decouples Run Time Infrastructure (RTI) functionality into common services, phase 2 delivers decomposed models resulting in composable operational M&S services and phases 3 uses next generation technology employing intelligent agents, metadata and model bases to allow on the fly M&S solution development.

The planned phase wise transition to NGA is similar to complex project involving stakeholders and uncertainties such as changes in requirements, priority, standards, policies etc. A project management technique should be identified for the smooth migration from the current architecture through these three phases to next generation architecture. The identified project management methodology should be one that supports phase wise or iterative development same as in the IT project management and which is not rigid or dependent on the sequential processes such as the traditional water fall model.

With the ever growing demand to manage growing customer expectations and to combat increasing competitive pressure to sustain, the organizations today want to automate, modernize, implement and evolve to new technologies which would strengthen their operations, reduce operating cost, integrate their business with other services, and enable coordination, sharing and exchange of services among partners and groups. Such project based organizations are in a constant quest to identify better ways to manage the inevitable change brought by implementing complex development projects. Large scale technology projects with exceeding budgets and scope running over several years are common these days. Complexity addition to such large projects is a result of several unpredictable factors such as long execution time, involvement of multiple stakeholders and their priorities, changes in the economic policies, government spending and reforms, funding priorities by sponsors, etc. Project Management as defined by PMBOK (2008) is the application of knowledge, skills and techniques to project activities to meet the project objectives within the constraint of scope, time, schedule, priorities, cost, quality and risk. Most of the tools and methods recommended by the literature have a focus on implementing such tools on "predictable environments" (Landaeta & Tolk, 2010).

In order to manage complex projects (which are governed by unpredictability), PMBOK (2008) recommends division of project lifecycle into project phases, and project phases into project processes that take place sequentially within every project phase, as a structured approach to implement project management controls to a project lifecycle. Struggling with managing the unpredictability in the IT projects, a group of IT practitioners used management practices such as quality control incremental iterative development, complex adaptive system, self managed cross functional teams, and fractal teams among others, to define agile project management frameworks and agile principles that can guide managers to successfully manage complex IT projects (Landaeta & Tolk, 2010).

The Agile project management framework is not rigid and accounts for the occurrence of unpredictable events with a guideline to manage them. One of these frameworks is Scrum, which has been applied successfully to manage IT projects in Fortune 100 organizations in the last years (Landaeta & Tolk, 2010). The paper shows that migration to NGA is a complex process and may reveal complexities that require a project management technique such as agile for managing them. It also proposes the applicability of Scrum as an agile method in this domain.

2. NGAT and Modeling and Simulation

In order to use JLVC simulations, JFCOM is currently employing resource heavy High Level Architecture (HLA) and RTI to support the additions of federates, accounting to significant costs incurred to move /upgrade each simulation into new versions of RTI or HLA. With this reliance on HLA and RTI, adding and upgrading the existing federates in future will result into huge spending. As per the NGAT report (2010), the M&S community needs to shift to an innovative next generation M&S architecture paradigm leveraging benefits from SOA's, Cloud computing, intelligent agents, composable services and other forward next generation technologies. To optimize future resources, JFCOM must establish an M&S solution – an innovative way ahead – for a Next Generation M&S architecture to support independent services collaboration within a common framework (NGAT, 2010).

A collection of innovative technologies will reduce the resources demand currently required to support the JLVC's "tightly-coupled" simulations. A team called Next Generation Architecture Team is composed to research and provide recommendations for an innovative Next Generation Modeling & Simulation (M&S) Architecture. The brief description and phases involved are (NGAT, 2010):

The NGAT has assembled the knowledge and developed the innovation to describe a new generation of M&S architecture to support the Department of Defense (DoD) and the war fighter in the defense. Designed in three phases, the strategy gently transitions the existing architectures from being problematic to being proficient.

Phase 1:

Phase 1 is the foundation of the NextGen architecture and establishes the initial integration of web services and commences the decomposition of services, building the framework for the future phases. A primary tenet of Phase 1 is the gradual JLVC transition from the existing HLA 1.3 implementation to a hybrid SOA implementation. At the close of Phase 1, the infrastructure will provide services robust enough to provide the foundation for a grid and computing environment while fully utilizing the SOA infrastructure and web services.

Phase 2:

The objective of Phase 2 is to provide composable M&S services. Phase 2 will further develop the NextGen M&S architecture by expanding the cloud computing environment and grid computing environments which will be characterized by intelligent agents and a model-based data engineering framework. Upon completion of Phase 2, the infrastructure will include a secure cloud computing environment, a grid computing plan and resource usage, intelligent services and agents, a MBDE framework and Model Query Language (MQL). These infrastructure elements will result in a cost-efficient, standardized, fully functioning, reusable NextGen M&S architecture.

Phase 3:

Phase 3 realizes the full application of the NextGen M&S architecture through development and implementation of autonomous intelligent services. These services monitor and learn environment information and become intelligent enough to intuitively determine, select, compose and run simulations from a model base of existing simulations. Phase 3 will evolve the NGAT architecture to autonomous intelligent services and model bases. At this stage, highly intelligent services will be in development to compose the simulations from existing simulation models.

3. Agile Project Management

Businesses have to respond quickly to changing needs, or could waste scarce resources doing non productive things. Most of the business, in anticipation of such changes, develop or alter business strategies from time to time to realign products and services and to meet market expectations. This is more evident in case of complex projects, where the priorities, requirements and decisions regarding the projects are constantly changing with time. As a result, such projects need a project management philosophy that embraces change, yet strictly following the management guidelines. This gave birth to agile project management development. Traditional project management aims to avoid change and rework once the project deliverables are finalized. Any change or rework is viewed as expensive aspect wasting time, money and resource. In most of the IT development, software development or system upgrade projects change in requirements or priorities are inevitable. The agile project management advocates the practice of managing the change rather than avoiding the change itself (Kotikalpudi, 2010). This aspect of agile project management is making it more popular, day by day, among the communities of IT and software development organizations.
The origins of agile have been laid decades ago. The United States Department of Defense (DoD) and NASA have used iterative and incremental development methods since 1950's (Sliger & Broderick, 2008). A white paper on the new product development published by Takeuchi and Nonaka in 1986 suggested that the rules of game in product development and its management are changing. Many companies have discovered that "it takes speed and flexibility" (Sliger & Broderick, 2008) to excel in today's competitive market. Anderson & Schragenheim (2003) define agility as 'an ability to cope with change'. Agile development methodology aims to provide many opportunities to assess the direction of a project throughout the development lifecycle. This agile motto is achieved through breaking down of work into deliverables, known as iterations. At the end of the iteration, the team working on it must present a dispatchable increment of work/task. During beginning and end of every iteration, the stakeholder is involved to gain feedback, preferences or approval on the plan or delivery of that iteration. In case the tasks need some changes, they are re-worked upon in the following iterations. By focusing only on creating the features that are of highest value and priority for the customer, agile project management maintains scope for changes in priority and decision without leading to loss of time and resources.

The waterfall development models are based on traditional rigid concepts "assuming that requirements are fixed and well defined at the beginning of the project" (Landaeta & Tolk, 2010). Based on this assumption, the traditional project management approach decomposes the activities to lowest possible level and then work up with a bottom to top approach to estimate the cost, time and resource associated with the project. The fundamental mistake is that we are using traditional project management tools to try to predict the future of a complex adaptive system, therefore if there is any change upon the agreed scope in the contract, the team is perceived to have made mistake in the project initialization phase (Landaeta & Tolk, 2010). In such cases the teams only have one chance to get each aspect of a project right. Agile has a focus on the iterative work cycles with a belief that project requirements are bound to change and it will be beneficial to deliver only the needed high value tasks without planning to deliver everything at the outset. In an agile paradigm, every aspect of development namely requirements, design, coding, testing etc. can be revisited and changed as per the changing requirements throughout the lifecycle, hence it is iterative and incremental approach suitable for the development processes today. Agile is about continuous improvement and a focus on minimizing cost and avoidance of waste. Agile methodologies generally promote stakeholder involvement, feedback, objective metrics and effective controls right from the project beginning. Agile project management with its principles as listed in the agile manifesto can better cope up with uncertainty. The principles behind the agile manifesto are (Sliger & Broderick, 2008):

1. “Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
2. Welcome the changing requirements, even late in the development. Agile processes harness the change for the customer’s competitive advantage.
3. Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
4. The most important and effective method of conveying information to and within a development team is face to face conservation.
5. Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
6. Simplicity – the art of maximizing the amount of work not done – is essential.
7. At regular intervals, the team proposes on how to become more effective, then tunes and adjusts its behavior accordingly.

In agile, the value of features to the customer drives the order of the work and the values outlined in the manifesto drive the way the work itself is defined. Some of the well known agile software development methods are:

1. Extreme programming
2. Feature driven development
3. Scrum

Scrum and Extreme Programming have gained popularity on a global level and emphasize on the same principles that describe ‘agile’ and share the values described in the agile manifesto. While XP is more of an agile engineering methodology that specifies software engineering practices such as test driven development, refactoring, pair programming, and simplicity in design, etc, Scrum is more of an agile management methodology that focuses on developing and managing software projects and is applicable outside of the pure IT domain as well (Kotikalpudi, 2010).

4. Scrum Overview

The term Scrum was first coined by Takeuchi and Nonaka (1986) in a study that was published in the Harvard Business Review (Kotikalpudi, 2010). In this study Takeuchi and Nonaka discussed the management of chaos by rugby approach. Rugby has self organizing teams, and the players work together to gain the control of the ball and move it up to the field. Takeuchi and Nonaka found that in a sequential system, crucial problems tend to occur at the point where one group passes the project tasks to the next. The rugby approach handles this problem by maintaining continuity across phases. This laid the foundation principal of the Scrum approach.

Scrum is an iterative, incremental software development process (Schwaber, 2004). Scrum is based the on the philosophy that the project requirements will change and that it will be a waste of project resources if we try to predict the future and set up a project plan (Landaeta & Tolk, 2010). Based on the principles of the agile project management technique, Scrum is designed to decompose various tasks of the project into iterations called sprints. Scrum seeks to manage chaos by implementing short time line (i.e. iterations) plans at which the cost and time are fixed by the customer and negotiated with the project team to the find out what attributes of the product or services can get done with these two constraints (Landaeta & Tolk, 2010). Several short sprints help in a manner that, owing to the limited productivity, in case the requirements or priorities change during the execution of the project, the team has not spent resources and time in producing something that has no value, is changed or no more needed.

The Scrum team is composed of three principal roles: Product owner, Scrum Master, and the team. The product owner defines the feature of the product and is responsible to prioritize them. The Scrum Master is the owner of the process; he or she interacts with the customer and the management and elects and teaches the product owner how to do his/her job. The teams do the actual work and are self organizing and cross functional.

5. Scrum and its Related Work in M&S

The field of IT and complex software development process is governed by issues of changing requirements and priorities. Scrum as explained, is an agile process that is used to manage and control “chaos” in complex software development projects by providing energy, focus, clarity and transparency to project teams by using iterative, incremental practices (Landaeta & Tolk, 2010). Consequently, it would be feasible to propose and implement Scrum for successfully managing the M&S projects. Scrum would enable an effective project management guideline for the reasons listed below (Kotikalpudi, 2010)

1. Scrum is scalable, from single process to entire project and because scrum is also adaptive, it can be customized to some extent to fit the unique situations and conditions of the project environment. In this case, each phase can be planned into several iterations and principles of scrum can be applied effectively.
2. Scrum is light project management frame work; it does not prescribe any specific technical practices which makes it feasible to integrate with any project
3. Although Scrum advocates face to face communications as the best form of communication; Scrum also adapts to work well with distributed teams as indicated by the numerous empirical studies. Scrum is designed to overcome the "human and organizational issues" by implementing the cross functional, self managed teams and by organizing the "Daily Scrum" meetings to ensure effective communication and collaboration.

Though Scrum seems to be a perfect fit for M&S projects there is very little literature supporting Scrum as a project management methodology for M&S projects. Kotikalpudi (2010) has demonstrated that the simulation development part of the M&S project is very similar to IT development project except for the "modeling" part. This work identified and established Scrum as a promising solution for enabling excellent agile project management in M&S federation development projects. A systems engineering process proposed by Kewely & Tolk (2009) for development of federated simulations in M&S, proposed an agile project management methodology like Scrum as a good fit. Further to this work, Landaeta and Tolk (2010) in their paper "Project Management Challenges for Agile Federation Development: A Paradigm shift" identified challenges for development of M&S federations and proposed Scrum as an agile process methodology to manage such complex and dynamic project. Further studies need to be conducted to check the applicability of Scrum on M&S projects to confirm the proposals.

6. NGA and Project Management

The project management challenges with any project are usually associated to scheduling, execution time, cost/budget, managing resources and any other unplanned events that may surface out during execution. Same principles apply to the migration from the current architecture to next generation architecture "which does not share the traditional philosophy of fixed requirements" (Landaeta & Tolk, 2010). One of the first management issues will be preparing to move to HLA 1516-2010 (NGAT, 2010), and it would be valid to assume that the transition process may reveal several other management issues which may be composed of changing priorities, delays in decision, rework and changes in requirements, that need to be addressed with a strong project management guideline such as agile, that would embrace and account for change. Such a complex project involving stake holders and their anticipated changing priorities can be similar to a complex software development or an IT project. The current sequential waterfall model or V model is based on the traditional concepts which plans for the entire project's deliverables during the initial stages assuming that requirements and priorities will not change during the execution of the project, This is the reason for unsuitability of such methods in managing complex projects since most of them constantly have to face changes in the priorities, scheduled, or agreed scope resulting in unfruitful results for the work done so far.

7. Scrum Application for NGA

The transition to NGA can be categorized into three phases where each phase is composed of several main activities and sub activities ranging from establishment of management and governance, standards, to needed infrastructure. A sample of activities for the needed infrastructure in each phase is provided below and explained by Fig.1. (NGAT, 2010):

Phase 1 decouples RTI functionality into common services. The infrastructure in phase 1 will be established in multi step process which includes steps as follows: 1.1. Establishing SOA Infrastructure, 1.2 Abstraction of the transport layer, 1.3 Establishing Web Services, 1.4 Interfaces/Service Level Agreements (SLAs), 1.5 Phased approach for federates integration, 1.6 Lay foundations for Phase 2 by ensuring that services know Data location, Data meaning and context, Data format required for data to be useful. At the close of Phase 1, the infrastructure will provide services robust enough to provide the foundation for a grid and computing environment while fully utilizing the SOA infrastructure and web services. Phase 2 delivers decomposed models resulting in composable operational M&S services.

The Phase 2 infrastructure will be implemented in a multi-step method as follows: 2.1. Develop secure cloud computing environment, 2.2. Determine use of grid computing resources, 2.3. Develop intelligent services/agents, and 2.4. Lay foundations for Model Based Data Engineering (MBDE). Upon completion of Phase 2, the infrastructure will include a secure cloud computing environment, a grid computing plan and resource usage, intelligent services and agents, a MBDE framework and Model Query Language (MQL).

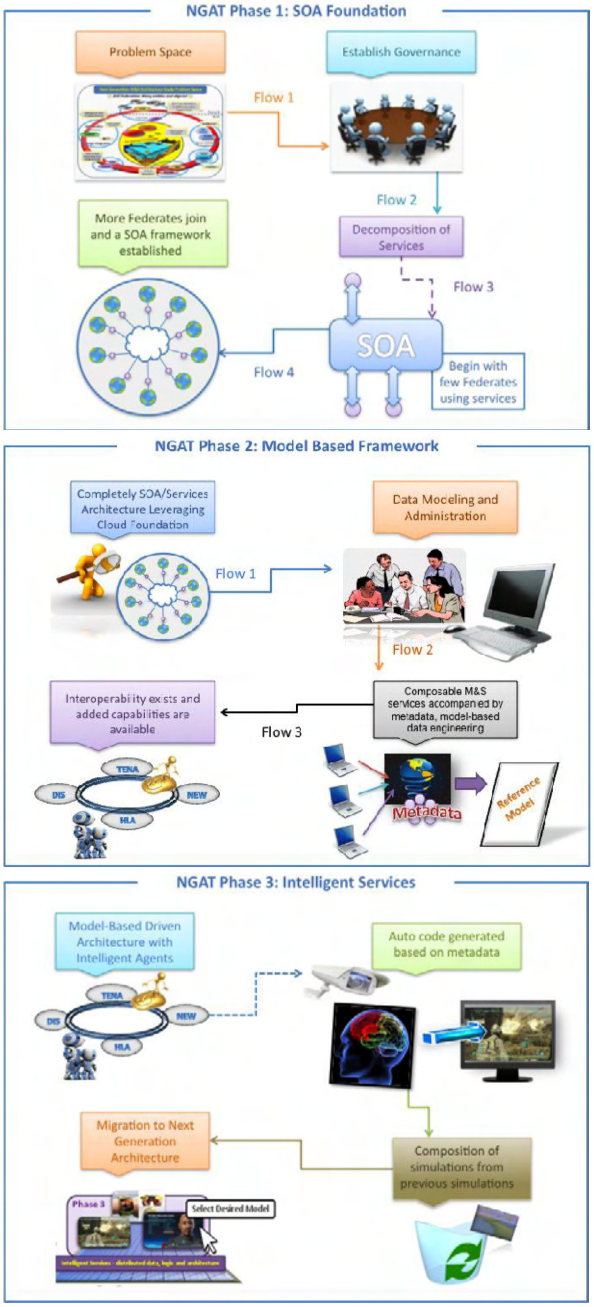


Fig.1: Phases in NGA transition (Taken from NGAT, 2010)

Phase 3 uses next generation technology employing intelligent agents, metadata and model bases to allow on the fly M&S solution development. The recommended steps to complete the desired infrastructure are as follows: 3.1. Build services that are now able to learn from their environment, 3.2. Take what the services are learning and store it as scenarios in a repository creating a Model, 3.3.Continued development of model bases, 3.4. Develop new Model Query Language (MQL) using Structured Query Language (SQL) as a guide/standard, develop a language to extract, add, delete and update simulations in the model bases. By the conclusion of Phase 3, the infrastructure will have functioning fully autonomous intelligent services, model bases and MQL as part of a cost-efficient, standardized, fully functioning, reusable NextGen M&S Architecture.

Looking from the Scrum approach, all these activities can be organized into a product backlog and prioritized. These activities can be further broken down into releases and organized into a release backlog and sub activities of each item in the release backlog can be developed and implemented in sprints/iterations (Landaeta & Tolk, 2010). If the stake holders decide to have SOA Infrastructure and Web Services ready in the first iteration, "they can be planned during the release planning meeting and the subsequent activities associated within each of the release backlog can be broken down into sprints and implemented" (Landaeta & Tolk, 2010). Applying Scrum to each phase will act as a project management guideline which will result in better and faster migration to NGA. Stake holders at different locations can be involved in the release planning meetings which will enable them to control, participate and plan the phase activities in each releases based on their priority and necessity. They can also engage in interactions and help to make collective decisions. Scrum is an effective tool, but should be implemented the right way, there must be experienced Scrum Master, product Owners, and qualified members of the project teams involved so that efficiently facilitate and control Scrum process (Landaeta & Tolk, 2010). An eagle's eye view of Scrum applicability to the prominent activities in phase1 is demonstrated through the example below:

The prominent activities along with their sub activities in phase 1 are:

1. Establishment of Governance and Standards:
    - Leverage existing standards
    - Identify and appoint board members

- Ensure members stay current on the standards and identify areas of best practices
- Continue standards for discovery and structural metadata

2. Decomposition of the current services and development of SOA infrastructure :
   - Developing SOA infrastructure
   - Abstraction of transport layer
   - Establishing web services
   - Managing interferences/Service level Agreements (SLA's)
   - Designing phased approach for federates integration
   - Lay foundations for phase 2 by ensuring that services know data location, data meaning, and data format required for data to be useful.
3. Migrate select federates using services:
   - Identify and migrate the current services
   - Assist the migration and joining of new federates

Fig.2.displays a complete Scrum project management approach, as explained by Schwaber, R, (2004), in the book "Agile project Management with Scrum". The Scrum project starts with planning a vision / road map of the system to be developed, which in this case would be migration to NGA. The vision will become clearer as the project proceeds. The product owner creates the product backlog which is a list of functional and non functional requirements that, when turned into functionality, will deliver this vision. The product backlog for this migration would contain activities planned in the phases. The product backlog is prioritized so that the items most likely to generate value or top priority are divided into proposed releases during the release planning meeting. In our case, a release would constitute implementations of phases. All the work needed to be delivered at the end of the release is done in sprints. Each sprint is an iteration of preplanned number of days. Each sprint is initiated with a sprint planning meeting, where the product owner and team get together to plan about the features to be completed in the sprint, which in our case would be the activities to be completed in phase 1. Daily stand-up meetings help to track the progress of the sprints.

If the stake holders decide to have SOA Infrastructure and web services to be ready in the first place, "they can be planned during the release planning meeting and the subsequent activities associated with each of the release backlog can be broken down into sprints and implemented" (Landaeta & Tolk, 2010). Applying Scrum to each phase will offer such flexibility and will act as a project management guideline which will ensure smooth, fast, better managed transition to NGA. The stake holders, even if placed at the different locations can be involved in the release planning meetings enabling them to participate, plan and control the phase activities in each release based on their priority and necessity.

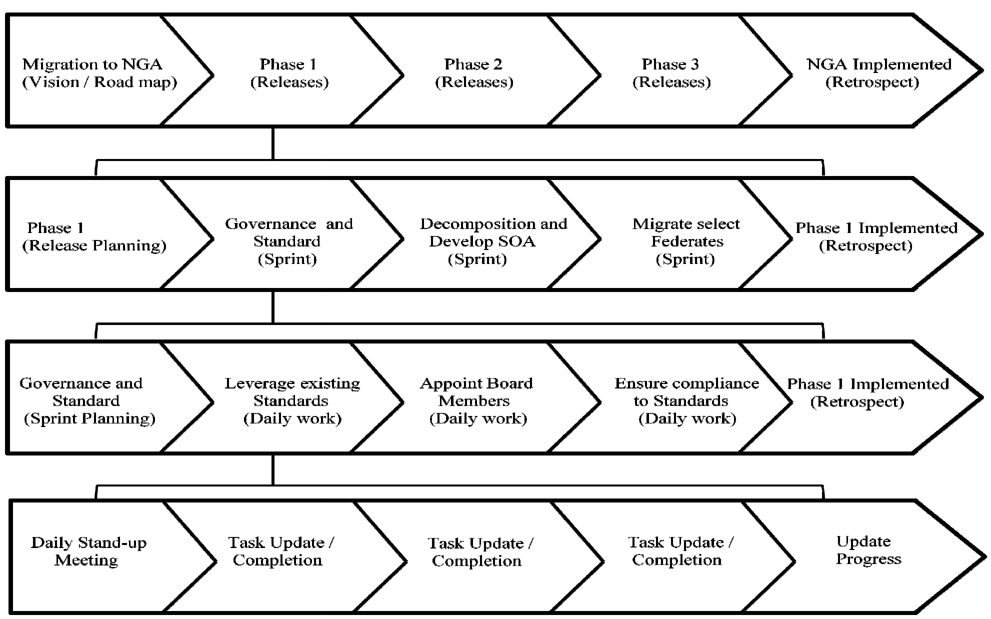


Fig.2: Scrum Stages for NGA (Adapted from Sliger & Broderick, 2008)

8. Summary

The paper presents an initial guideline and benefit of applying Scrum to manage the transition from current architecture to NGA. In order to resolve problems associated with the "tightly-coupled" resource heavy HLA and RTI, NGAT report (2010) recommends transition from current architecture to efficient and "loosely-coupled" NGA, which will resulting in significant cost saving. The three phase transition is an M&S project similar to managing a complex software development project owing to the "chaos" that stake holder's changing priorities, requirements and other uncertainties that may be revealed ahead. A project management methodology such as agile which can accommodate and manage change and uncertainty is needed to guide this transition. Stated from the definitions, agility is 'an ability to cope with change'. Unlike traditional project management, agile project management believes in better managing the change and uncertainty rather than avoiding them. Currently, agile project management is widely implemented in IT industry to manage development projects and its success is proven by its popularity and acceptance among the IT community

In their work (Kewley & Tolk 2009, Kotikalpudi 2010, and Landaeta & Tolk 2010) have identified, proposed and recommended agile project management methodology, specifically Scrum for managing M&S projects. Scrum is more of an agile management methodology that focuses on managing change in complex projects and appears to be a suitable choice to guide this transition. Scrum proposes solutions to efficiently cope with requirement changes, keeping the overall development process stable without supporting the fulfillment of no longer valid requirements (waste of resources) and, in parallel, supporting the fulfillment of new requirements (providing updated functional capability) (Landaeta & Tolk, 2010). The paper proposes and recommends Scrum as an agile project management methodology looking at its suitability and applicability to manage the transition to NGA

9. Future Work

The work presented in this paper is conceptual at this stage, based on the literature review and sample implementation, but will be evaluated in detail by the author in his Master Thesis on this topic.

Acknowledgement

The author is grateful to Dr. Andreas Tolk for suggesting and granting opportunity to conduct research on the topic of the paper for Master Thesis and also to Dr. Jose Padilla for his continuous guidance, motivation and support which helped a great deal to endeavor the task of writing this paper.

10. References:

Anderson, D., & Schragenheim, E. (2003). *Agile Management for Software Engineering: Applying the Theory of Constraints for Business Results.* New Jersey: Prentice Hall.

Camci, A., & Kontour, T. (2006). Technology Complexity in Projects: Does Classical Project Management Work? *PICMET 2006 Proceedings,*, (pp. 1-5). Istanbul, Turkey.

Kewley, R., & Tolk, A. (2009). A Systems Engineering Process for development of federated simulations. *Spring Simulation Multiconference.* Society for Computer Simulation International.

Kotikalpudi, P. (2010). *How to Improve Project Management For Agile Federation Development. Masterof Science Thesis in Engineering Management and Systems Engineering.Old Dominion Univeristy.* Norfolk, Virginia.

Landaeta,R., & Tolk, A. (2010). Project Management Challanges for Agile Federation Development: A Paradigm Shift. *Fall Simulation Interoperability Workshop* , 1-6.

NGAT. (2010). *Toward a Next Generation M&S Architecture.* Norfolk, VA. (NGAT -Next Generation Architecture Team)

PMBOK. (2008). *A Guide to Porject Management Body of Knowledge (4th Edition).* Newtown Square, PA: Project Management Institute.

Schwaber, K. (2004). *Agile Project Management with Scrum.* Redmond , WA: Microsoft Press.

Schwable,K. (2006). *Information Technology Project Management.* Boston, MA:Course Technology.

Sliger, M., & Broderick, S. (2008). *The Software Project Managers Bridge to Agility.* Boston, MA: Pearson Education.

Sutherland, J. (2010, March 12). *The Scrum Papers : Nuts, Bolts, and Origins of an Agile Process.* Retrieved March 4, 2011, from Scrum training: http://scrumtraininginstitute.com/home/stream_download/scrumpapers

10. Author's Biography

Rupesh P. Narkar is a Master of Science Student in the Engineering Management and Systems Engineering Department at Old Dominion University. He has a B.E degree in Mechanical Engineering from the University of Pune (2006). His research interests include project management and applications of agile project management to modeling and simulation projects.

Training and Education

Track Lead: Mr. Sol Sherfey, VMASC

Judges: Dr. Larry Bartosh, JFCOM, Dr. Eric Weisel, Weisel Scientific & Technology Corporation

Development of Web-Based Engineering, Educational and Assessment Modules for Learning Numerical Methods

Author(s): Subhash Chandra Bose S V Kadiam, Ahmed Ali Mohammed and Duc Thai Nguyen

3rd Place Gene Newman Award Winner, Training and Education Track

Low versus High Intensity Emotion in Animated Pedagogical Agents: A Qualitative Study

Author(s): Enilda Romero

Abstract Only

1st Place Gene Newman Award Winner, Training and Education Track

Pushing the Limits of the ARCS Model in the Motivational Design of Instructional Simulation

Author(s): Don Robison

2nd Place Gene Newman Award Winner, Training and Education Track

Pushing the Limits of the ARCS Model in the Motivational Design of Instructional Simulation

Don G. Robison, M.S.

***Abstract*-Learner motivation is fundamental to effective learning. Keller's ARCS Model of learner motivation is a macro-level model that synthesizes motivation research and provides heuristics for instructional designers to design instruction that is motivating. This paper applies the ARCS Model to instructional simulation motivational design.**

The ARCS Model is described in terms of its four components: attention, relevance, confidence, and satisfaction. The motivational design process associated with each ARCS component is presented and processes for developing motivational objectives and strategies are described. Additionally, a process for audience motivation analysis using the ARCS Model is provided. The motivational design opportunities and challenges of instructional simulations are analyzed within the framework of the ARCS Model, and heuristics for the design of motivationally engaging instructional simulations are offered. The implications of four significant motivational influences not addressed by the ARCS Model—social factors, visual design, structured story, and flow theory—are discussed and design heuristics based on these concepts are provided.

***Index terms*-Simulation, learning, design methodology, motivation, instructional simulation, learner motivation, ARCS Motivation Model, John Keller.**

I. Introduction

A recurring truth weaves itself through the literature about effective learning: in order for learners to process energetically and to learn deeply, they must choose to engage [1-3]. Learner motivation is fundamental to learning. Detractors have dismissed the study of learner motivation as either a vague concept [4] or as lacking predictive power in respect to learning outcomes [5]. But, as instructional design practice moves towards a learner-centered view, it is clear that motivation is essential to learning [1].

Learner motivation can be divided into six concepts that make it easier to understand. Motivation can be defined as what initiates behavior, controls its intensity, maintains behavior, and mediates choice; and it may also be defined by the emotions that accompany behavior [4]. As it applies to learning, motivation is focused on the energy and choice that learners apply to learning. It concerns effort, but also concerns the learner's choice to engage effective metacognitive strategies [1]. It is useful to think about learner motivation in terms of both the energy and persistence of the learner towards the task of learning, but also about the learner's choices and employment of effective learning strategies.

John Keller's ARCS Model has been presented as the most comprehensive and practically accessible learner motivation model for instructional designers [3, 6]. In fact, Jonassen has proposed that the ARCS Model is the only effective practical model for incorporating motivation theory into the instructional design process. The model synthesizes motivation research into a "four-factor macro-level model" [2].

This paper explores the motivational design of instructional simulations using the ARCS Model as a framework. The paper will present a summary of the ARCS Motivation model, the motivational challenges and opportunities of instructional simulation, and offer practical approaches to designing simulations that engage learners and keep their interest. Motivational challenges will be analyzed from the perspective of the ARCS Model paradigm. In addition, limitations of the model will be discussed and proposals for expansion of the model—particularly as they relate to instructional simulations—will be offered.

II. The ARCS Model of Learner Motivation

Keller's ARCS Model is derived from many areas of motivation research [2]. He took motivation from a vague generalized idea and broke it down into four useful components: attention, relevance, confidence, and satisfaction [7]. According to Keller, there are four general requirements that must be met for learners before they will invest cognitive energy in learning. These requirements are related to two major questions, as depicted in Table I. Keller developed these four requirements into the four components of the ARCS Model.

A. *Attention*

Keller states that, "To be motivated, a learner's attention has to be aroused and sustained"[2]. The design goal here is to capture and sustain the learner's attention. This requirement includes things that relate to curiosity, sensation seeking, and other factors that explain how attention is

Manuscript received March 12, 2010

D.G. Robison is a Ph.D. student in the Instructional Design and Technology Program in the Darden College of Education, Old Dominion University, Norfolk, VA 23529 USA (email: drobi0365@odu.edu)

TABLE 1
COMPONENTS OF THE ARCS MODEL [2]

<table>
<tr><th colspan="2">Major Categories and Definitions</th><th>Major Process Questions</th></tr>
<tr><td>Attention</td><td>Capturing the interest of learners; stimulating the curiosity to learn</td><td rowspan="2">“How is this learning valuable and stimulating to my students?”</td></tr>
<tr><td>Relevance</td><td>Meeting the personal needs/goals of the learner to effect a positive attitude</td></tr>
<tr><td>Confidence</td><td>Helping the learners believe/feel that they will succeed and control their success</td><td rowspan="2">“How can I (via instruction) help students succeed and allow them to control their outcomes?”</td></tr>
<tr><td>Satisfaction</td><td>Reinforcing accomplishment with rewards (internal and external)</td></tr>
</table>

obtained. Keller believed that this motivation requirement is the easiest to satisfy. He proposed that attention is stimulated and sustained by events that are startling or unexpected which arouse curiosity or engage learners in inquiry oriented behavior that, in turn, stimulates a deeper level of curiosity.

In addition to creating a sense of inquiry, this attention component relates to Keller’s concept of “variability.” Variability is changing up the content presentation appearance or pattern. Care must be taken to maintain the flow of the experience even as some variety is introduced [7].

B. Relevance

Keller describes both an “ends” and “process” aspect to relevance [7]. The ends aspect is straight-forward: If the instructional content is perceived to be helpful to the learner in accomplishing important future goals, the learner will attach relevance to it. An example of relevance: A college junior desires to go to medical school and has signed up for a course to prepare to take the medical entrance exam. In this college student’s mind, the entrance exam is the key to getting into medical school, and the course to prepare is the key to the exam. In this case, the student would perceive a great deal of relevance in the preparatory course.

However, the end does not always have to be aimed at the future. The learner may attach a near-term value to the instructional content. Today’s class could relate to challenges the learner expects to experience tomorrow, causing the learner to attach relevance to the learning. Likewise, learning that is seen to relate to past challenges is viewed as relevant to learners.

The process aspect of relevance has to do with methods of teaching in relation to the student’s preferences or needs. For example, people who are high in their need for affiliation will be attracted by group learning contexts and may attach relevance to group learning experiences.

C. Confidence

A learner must believe that there is a reasonable probability of success before investing energy in learning [2]. This is related to Bandura’s [8] self-efficacy principle: an individual’s expectation for success strongly affects selection of tasks. If there appears to be no chance for success—either because of personal capability limitations or circumstances—motivation towards the task will be low.

On the other hand, if learners believe that a task challenge is too easy, they will not engage energetically. The optimal challenge is where a task is perceived by the learner as having moderate difficulty. A successful outcome should not be guaranteed, but the learner needs to be provide dwith some sense of challenge [9]. In addition to expectations for success or failure, this confidence dimension includes the perception that the learner has control over the outcomes. For optimal motivation, the learner must feel that success is possible, but that it will only come as a result of personal effort [9].

D. Satisfaction

The learner question that relates to satisfaction is a simple one: “Is the outcome of this effort desirable to me?” For example, the business school student who does not really want to go into business will likely not work as hard as the one who is passionate about a career in business. Learners will work for outcomes that are desirable to them. Behaviorist influences on the ARCS Model are most discernable in this satisfaction motivation component. Keller speaks of the use of reinforcement schedules in his description of satisfaction [2]. In addition to extrinsic rewards, Keller also discusses the importance of intrinsic motivation as part of learner satisfaction.

Intrinsic motivation is defined as motivation that is driven by interest or enjoyment of a task [10]. The artist who paints pictures and then hangs them in a studio where they are never seen, is intrinsically motivated to paint. Motivation that is intrinsic to the learner is more complicated than extrinsic motivation, but may be more durable. Intrinsic motivation is of particular interest in the design of instructional simulations as designers have the potential to create simulations that are both interesting in content and enjoyable in process.

III. MOTIVATIONAL OBJECTIVES AND AUDIENCE ANALYSIS

One valuable contribution of the ARCS approach is its systematic procedure for motivational design. Borrowing from the process models of instructional design, Keller articulates a process for developing motivational objectives, conducting audience analysis, and designing motivational strategies that is matches the audience and performance [7]. Table II presents Keller’s conception of the subcategories of motivational design strategies in the ARCS Model. In Keller’s process, the designer systematically focuses on each component of the model and develops motivational objectives for each subcategory. For example, for a training session with instructional designers, an objective in the confidence dimension might read: “Participants will indicate a higher degree of confidence in their ability to conduct instructional design [7].” After developing a list of motivational objectives, measures for achieving each objective are developed. This process is similar to the development of performance objectives and associated evaluation items.

The measures of achievement of a motivational objective is either a self-report or a direct observation of behavior that indicates the targeted performance [7]. For example, if the motivational objective is that “Participants will indicate a higher degree of confidence in their ability to conduct motivational design.” Then, the measure would be a survey item that has learners rate the improvement of their confidence in conducting motivational design.

Once the designer has specified the motivational objective and has defined the method of measurement for achieving it, an audience analysis is conducted. Since there is an endless list of possible strategies for improving the motivational draw of instruction, an audience analysis helps focus the number and types of motivational strategies that should be employed in a given context [2].

The audience analysis is conducted by assessing the audience motivation ranges for each of the four components of the ARCS Model. Based on the target audience, pre-instruction motivation ranges in each component will vary. The designer will vary strategies to bring motivation to the optimal level in each component area.

For example, a class of Anatomy students may have a final exam in a course that will dictate their acceptance to the next level of their program. In this case, the relevance motivation measures might be too high. Some of the learners would be fearful and anxious. The motivational strategy would then be to reduce anxiety by reducing the relevance variables. One motivational intervention may include a change in policy to make passing the test one of several checkpoints for progressing in the program, rather than making the one exam the key checkpoint.

Another example is high school English in an urban school setting. The students in this example say that they do not care for the subject. They do not believe it makes a difference in their lives, and the results of success are meaningless to them. In this case, the pre-instruction levels for relevance and satisfaction are very low. Motivation strategies would attempt to raise motivation as it relates to relevance and satisfaction with learning outcomes.

IV. Motivational strategies

Once motivational objectives and measures have been identified, and the audience has been described along motivational dimensions, the designer develops motivational strategies that aim at the specific motivational objectives and measures given the target audience characteristics. Keller held that these strategies would fall into one of the subcategories defined in Table II. For example, if a group of learners are nervous about working with computers, and computer-based learning is planned for them, a motivational strategy may be to spend a lot of time making the learning program user-friendly to boost learner confidence. Systematically thinking through the learners’ needs and characteristics, it is possible to determine where the greatest emphasis should be placed in terms of motivational interventions [2].

TABLE II
Subcategories of Motivational Design Strategies in the ARCS Model [2]

ARCS Dimension		Strategy Subcategory
Attention	A.1	Perceptual Arousal
	A.2	Inquiry Arousal
	A.3	Variability
Relevance	R.1	Familiarity
	R.2	Goal Orientation
	R.3	Motive Matching
Confidence	C.1	Learning Requirements
	C.2	Success Opportunities
	C.3	Personal Control
Satisfaction	S.1	Natural Consequences
	S.2	Positive Consequences
	S.3	Equity

V. Applying the ARCS model to instructional simulation

Researchers have studied the application of the ARCS Model in several instructional settings: self-regulated instructional text [11], classroom instruction [3], email supplements to instruction [12], and computer-based instruction [2] to mention a few. But, the ARCS model has not been systematically applied to instructional simulation. The following analysis will address the four main ARCS components as they apply to instructional simulations. Because there is no specific target audience, and because instructional simulations represent a broad spectrum of approaches, these discussions will address general motivational challenges and affordances.

A. *Attention Related Issues*

Keller held that concreteness, curiosity arousal, and variability were key to capturing and keeping a learner’s attention [7]. He also felt that the quality of initial presentation related to this motivational component. As these factors relate to instructional simulations, the quality of the interface is important, as is the initial layout of both the simulation process flow and the screens: they must represent the simulation in clear and compelling ways. One of the strengths of simulation is that a simulation allows for concrete presentation of sometimes abstract phenomena [13]. This means that simulations should provide designers with unique opportunities for attention-related motivational design.

Variability and novelty are key approaches to capturing and maintaining attention, but care must be taken to present content-appropriate changes and novelty[7]. For example, animations of missiles going off would be inappropriate for a simulation of cross-cultural communications. In simulations, variability and novelty may be introduced through scenario changes or through the model design itself.

The use of humor may be an effective way to capture a learner's attention, but again, the humor must match the context and content of the instruction. The effectiveness of the use of humor has been demonstrated in game-based learning [14] and could be effectively applied in instructional simulation as an attention strategy.

Curiosity arousal may be stimulated through the use of unanswered questions or presented mysteries. Drawing again on lessons learned from gaming, an instructional environment may be made challenging by introducing clear goals with uncertain outcomes [15]. The use of various levels of difficulty also helps to motivationally enhance the experience. Uncertainty may be provided through the use of hidden information, or randomness in presentation.

Instructional simulations are particularly well-suited for capturing and keeping the learner's attention: the interface must be designed well, goals must be clear, the simulation flow must be clear at the outset, and variation may be introduced through the use of uncertain outcomes, hidden information or randomness in presentation.

B. Relevance Related Issues

The relevance of an instructional simulation is a direct result of learner's perception of how the instruction matches learner goals and motives. Relevance is also related to how the instructional approach matches expectations or learning preferences. Instructional simulations based on models that relate to real-life challenges in the learner's current world of work or learning, or to the learner's desired world, would be motivational.

Instructional simulations should explicitly describe the projected outcomes of successful completion of the simulation. For example, for Coast Guard officer trainees, a ship-handling simulator would state that the skills required to successfully complete the simulation are the same skills that will be required to successfully handle an actual cutter. A simple way to tie relevance to a simulation is to state the real-world outcome related to success. A simulation would be perceived to be more relevant if the learner perceives high model and process fidelity.

Another way that instructional simulations can be motivationally enhanced is by including user options that allow learners to match their own goals or aspirations to simulation outcomes or activity. When appropriate, the use of like-respected role models is effective in modeling successful skill performance, but also in showing the effort it takes to achieve [16]. The actions of one that is perceived as being similar to a learner are relevant in that they give the learner information regarding their her own capacity to achieve what the model achieved. The most effective employment of a model in this capacity is to use one that the learner sees as similar in many respects, and achieves through effort and challenge [16]. If the model's achievement is seen as too easy, it will not have a motivating effect on the learner.

Keller presented the idea of using competition—particularly against self—as a means for increasing the relevance of an instructional activity [9]. Including a competitive element in a simulation may take the form of a score, an achievement criteria for advancement in difficulty levels, or even a head-to-head competition.

Relevance may be enhanced by basing models on real-life scenarios, explicitly describing the outcomes of successful learning, including user options that match personal motives and goals, using like-respected role models, and introducing a competitive element in the simulation.

C. Confidence Related Issues

To optimally engage, a learner must have an expectation that personal effort will lead to success [9]. This does not mean that there should be no challenge, because the lack of challenge is de-motivating in itself. Optimally, the level of challenge would be such that the learner is not sure of success, but believes that success is probable if effort is exerted. Keller posited three chief approaches to enhancing confidence [17]. It is important to clearly describe the learning requirements. If possible, provide an opportunity for the learner to describe personal learning goals. Allow the learner choice in difficulty, and provide opportunities to progress. It is helpful to include options for self-evaluation. Confirmational feedback for acceptable responses and corrective feedback for responses that do not meet criteria enhance instruction. In addition, it is good to give learners as much personal control and responsibility as possible.

The application of Keller's principles for enhancing learner confidence as described above to instructional simulation is apparent. Again, the interface to the simulation must be clear and provide learners with an understanding of the performance and learning requirements. If practical, allow learners choice in the level of difficulty and give them an opportunity to select a performance or learning goal. Whenever appropriate, give the learner control of path or next step.

Optimal learner motivation is achieved when the learner believes that personal effort will result in success. Providing clear learning requirements, allowing learners to select difficulty levels, allowing them to state personal performance or learning goals, and giving them control whenever possible will enhance learner confidence in instructional simulations.

D. Satisfaction Related Issues

Satisfaction relates to the learner obtaining a satisfying level of accomplishment upon completion of learning. The issues here relate to extrinsic rewards, intrinsic motivation, and equity. Extrinsic rewards are outcomes of the instruction that the learner values. At a basic level, state lotteries exist because players value the possible outcome of winning and being wealthy. In the case of lotteries, the expectations of success are very small, but the value of the final outcome outweighs those expectations, and people play. Keller suggested the use of scoring systems with computer games to enhance this sense of extrinsic reward. Congratulatory comments, reinforcements, and certificates of success are other example of extrinsic rewards.

Deci [10] demonstrated that intrinsic motivation is powerful, and Keller proposed strategies for encouraging intrinsic motivation. Keller believed that situating learning in real-world contexts was an effective way to encourage intrinsic motivation. He proposed that verbally recognizing and praising intrinsic motivation encouraged it. Giving learners who have mastered a task the opportunity to help others also stimulates intrinsic motivation towards a task.

The search for identity is a primary motivator, particularly for adolescents and young adults. Erikson proposed that this search for ‘who I am’ and ‘what I am capable of’ in young adults energized behavior and choice [18]. Perhaps providing learners with an opportunity to purposely associate achievement in a simulation with some aspect of identity would be an effective way to enhance learner satisfaction with instruction. In an instructional simulation, this may be as simple as making a level of attainment synonymous with a valued title or role. For example, in the ship-handling simulation example used earlier, one level of learner attainment may be “Qualified Deck Watch Officer” a title that matches a future goal for these learners. The key is to this approach is to use a title that would have be a part of valued self-identity descriptions. Another tactic for using this interest in identity is to allow learners to personalize avatars. Gamers use this tactic to draw game players into a game [19] .

Finally, for motivational satisfaction to be optimal, learners must feel that the process of learning was fair. Evaluations and real-world experiences should relate to the learning processes. The application in instructional simulation is related to both model and process fidelity. While there may be some simplifying of the model or process in training, the final learning product should relate directly to the real-world environment where the skill will be practiced.

Satisfaction in instructional simulations, therefore, may be enhanced through the use of extrinsic rewards: scoring systems, levels of difficulty, congratulatory feedback, or certifications. Intrinsic reinforcement may be achieved through situating the simulation in the real-world, through model or process fidelity, recognizing and praising intrinsic motivation, and giving master learners and opportunity to help others. One idea is to encourage identity formation in success by associating levels of simulation advancement with valued position titles. In addition, learners need to have a sense the learning experience was fair. This may be accomplished by making simulation experiences match actual experiences or tests.

TABLE III

APPLYING THE ARCS MODEL TO INSTRUCTIONAL SIMULATION

Strategies for Accomplishing Goal of ARCS Component	
Attention	• Well designed user interface • Clearly state instructional goals • Clearly describe simulation flow • Use of clear goals with uncertain outcomes • Use of hidden information • Use of randomness in presentation
Relevance	• Create model faithful to real-life scenarios • Explicitly describe the outcomes of successful learning • Include user options that match personal motives and goals • Use like-respected role models • Incorporate a competitive element
Confidence	• Provide clear learning requirements • Allow learners to select difficulty levels • Encourage learners to state personal performance or learning goals • Give learners control whenever possible
Satisfaction	• Use scoring systems • Build in levels of difficulty • Provide congratulatory feedback • Provide achievement certifications • Situate the simulation in the real-world • Recognize/praise intrinsic motivation • Give master learners opportunities to help others • Associate an identity position or role with a level of difficulty • Make simulation experiences match test and real-world experiences

VI. OTHER MOTIVATIONAL ELEMENTS

As described earlier, the value of Keller’s contributions in providing a comprehensive applied motivation model for instructional designers cannot be overstated. Yet, motivation is a complex study, and there are significant motivational elements that were either under-represented or not represented in the ARCS Model.

A. Key Motivation Elements Underrepresented in ARCS

Three examples of motivational elements that have a place in the ARCS Model but are underrepresented are the motivational impacts of identity formation, play, and fantasy. Intentionally incorporating elements that encourage ego identity discovery or exploration (particularly with adolescents and young adults) was discussed as part of the explication of the motivational component of satisfaction above. Since the search for identity is a broadly motivating issue, incorporation of identity facilitating elements would improve the motivational appeal of an instructional simulation, if appropriate to the simulation objective.

Gaming researchers have established the power of play and humor in keeping players engaged in instructional games [20] [14]. The lessons learned may be readily transferred to instructional simulation. The use of appropriate and rightly-timed humor, or the introduction of enjoyable competitions or quests, could enhance the motivational impact of instructional simulations. Care must be taken in the design to ensure that the use of humor or fun fits the objectives. Play has been found to be an effective mediator of learning for both children and adults [20].

The use of fantasy has been shown to be effective in gaming [15]. For example, having users assume the identity of a star baseball player and hit the World Series winning home run is motivating. Or, another fantasy approach that is motivating to players is to have the player find the secret weapon with which the world may be saved. These routine gaming approaches could be used effectively in instructional simulations, as well.

B. Key Motivation Components Not Represented in ARCS

Four important motivation dimensions relevant to instructional simulation that are not addressed by the ARCS Model are: social influences on motivation, the motivational impact of visual design, the motivational force of structured story, and the motivational effects of flow.

First, though the ARCS model addresses aspects of social learning, including vicarious learning and self efficacy, it does not address broader social learning variables related to learner motivation, including the effects of group affiliation, collaboration, optimal synergy, and presence. Each of these variables has a demonstrated impact on learner motivation and achievement, and yet are not explicitly addressed in the ARCS Model. It is possible that Keller was influenced to downplay social motivational influences by his colleague, Gagné, who held that social motivators were not predictive of learner engagement [21].

Researchers from various theoretical orientations agree on the basic principle that people are social creatures, and social factors influence learning motivation. Social learning theorists, for example, argue that much of what one learns is learned from the example of others [8, 22]. Cognitive learning theorists maintain that much of learning is devoted to "meaning making" which is essentially a collaborative and social process [22]. Palincsar, a social constructivist, goes as far as to say that learning is an entirely social act [23]. Interestingly, Hacker and Bol demonstrated that the mere potential presence of others affects individual cognition and behavior [24]. One does not have to fully adopt all of these perspectives to grant that social factors influence learner motivation.

Additionally, highly motivated individuals with common ends often find each other and work together to accomplish what they could not individually. Examples of this interaction are school study groups where students gather to learn difficult content together, and professional working groups which voluntarily assemble to address a challenging issue. Bennis [25] argues that these synergistic collaborations lead to the most creative innovations, and are the best way to accomplish difficult tasks. He describes a sort of "synergistic abrasion" that is foundational to innovation. Great work and great achievement are their own rewards. Bennis calls these groups "Great Groups." The motivational of impact of such groups on their participants is powerful.

Some examples of how instructional designers, simulation designers, and game designers have incorporated social elements in the motivational design of their products include: the creation of virtual social spaces and the use of personal avatars in micro-worlds like *Second Life* [26], using animated pedagogical agents to teach empathetic teaching skills in counselor training [27], and using animated human models in procedural demonstrations of heavy equipment maintenance [28]. The social dimensions of learner motivation may also be incorporated into an instructional simulation using collaborative, competitive, and cooperative approaches.

Second, the visual design and aesthetic quality of the simulation affects user perceptions of quality and credibility. This has been demonstrated in web design research [29, 30], but the extension of this principle to the design of instructional simulation is apparent. Instructional simulations are, by definition, experiential learning, and the designer must consider the quality of the entire experience and not just the goals and mechanics of it [31]. Keller places this design "excellence" dimension in the "attention" motivational component of his model [32], but it is possible that this aesthetic variable has a broad impact, also affecting learner perceptions of both relevance and satisfaction.

Third, learning experiences are best understood and processed in the context of a story [33]. Learning is enhanced by placing learners in the real-world context of the target task performance. One way to situate learners in context is to follow the lead of screen writers and book authors who have developed story patterns that reliably engage users in their productions [34]. Game designers currently use stories to enhance interest in their games. Incorporating proven story patterns in pre-training or in-training simulation design could enhance learner engagement. In a related development, neuroscience research in brain activity has shown that specific patterns of cinematic plot engage large percentages of audiences in predictable ways [35], opening the door for researchers to establish story plot patterns that instructional designers may utilize in their designs. A practical heuristic for applying the strength of story to instructional simulations is the introduction of a compelling "back-story" that could increase initial learner interest and serve to sustain engagement.

Fourth, flow theory is a way of describing a unique phenomenon where people are caught up in the flow of an activity or game and so enjoy themselves that they lose track of time and orientation. Czikszentmihaly defines flow as

"…the state in which people are so involved in an activity that nothing else seems to matter" [36]. Rieber argued that the most important element in the creation of a flow experience is optimizing the challenge [20]. This idea is consistent with Keller's ARCS Model as the optimal state for learner confidence [17]. One final point about the importance of flow in motivational design: it could be a primary motivator for many learners. People actively seek flow experiences, and will expend energy and resources to find them [19]. This idea of flow is compatible with the ARCS Model, and Keller discusses it [32], but there are few practical heuristics for designing for flow in the ARCS process.

To optimize for flow in instructional simulation motivation design, simulations should be designed in such a way that learners are challenged at every point, but not overwhelmed. This may be achieved by providing learners with options for selecting simulation difficulty, or monitoring learner achievement levels and then varying the difficulty of the simulation presentation based on that achievement. Rieber describes the simulation design goal as creating an opportunity for "ideas to expand as the learner is ready for them" [20].

According to Rieber [20], enjoyment results when an activity includes one or more of the following components:

1. Challenge is optimized.
2. Attention is completely absorbed in the activity;
3. The activity has clear goals;
4. The activity provides clear and consistent feedback as to whether one is reaching the goals;
5. The activity is so absorbing that it frees the individual, at least temporarily, from other worries or frustrations;
6. The individual feels completely in control of the activity;
7. All feelings of self-consciousness disappear; and
8. Time is transformed during the activity (e.g. hours pass without noticing) [20].

One conspicuous example of flow theory in action is the massive multi-player online role-playing game (MMORPG) *World of Warcraft (WoW)*. The worldwide subscriber base for *WoW* reached over 12 million in 2010 [37]. Each subscriber pays approximately $14 a month to play, and in addition, they pay $40 for each new version update [37]. Millions of users have been subscribers for more than two years. This means that *WoW* taps into significant motivators. The factors that initially engage and motivate *WoW* users and the factors that continue to sustain engagement and motivate *WoW* users are relevant to the optimal motivational design of instructional simulations.

Users report that they were initially drawn to *WoW* by either social influences (e.g., the invitation of a friend) or the initial challenge [19]. Some users report that they were also initially drawn by the detailed and extensive background story and the exceptional graphics. As users become more proficient, they continue to be drawn by social connections (both formal and informal), identity factors (they can tailor their personal avatar to a fine level of detail), and the increasing complexity of the game itself. New versions of the game with complex levels, more detailed story, and improved graphics and functionality are released regularly [19], which established players find motivating.

. *TABLE IV*

APPLYING SOCIAL AND FLOW MOTIVATION PRINCIPLES TO INSTRUCTIONAL SIMULATION

Strategies for Incorporating Social and Flow Principles	
Social Factors	• Include collaboration features • Provide opportunities for learners to cooperate • Create teamwork experiences • Provide team challenges • Use pedagogical agents in the simulation • Use avatars in the simulation
Visual Design	• Graphics should be designed by professional artists
Structured Story	• Develop an interesting story that complements the goals of the instructional simulation
Facilitating Flow	• Articulate clear goals for the simulation • Giver learners control of any optional functionality • Provide clear, consistent, and immediate feedback in relation to goals • Provide the opportunity for learners to expand their activities through new levels or new activity paths • Provide opportunities for competition • Allow learners to personalize avatars • Provide opportunities for formal and informal social networking

There are obvious differences between games and instructional simulations, but some of the features of games that may be used in the motivational design of instructional simulations include the articulation of clear goals, giving learners control, providing clear and immediate feedback, providing communication channels for learners, and allowing them to personalize avatars. The unique characteristics of simulation make it an ideal medium for the creation of flow experiences for users. Table IV summarizes design heuristics to optimize flow in a simulation.

VI. CONCLUSIONS

Motivation is a complex science, but the motivational aspects of a simulation can be designed in such a way that motivation is optimized. Energetic learner engagement cannot be guaranteed as the result of a particular design process, but, the ARCS Model provides an effective framework for addressing the motivational challenges of instructional simulation. Motivation may be intentionally and systematically influenced. Strategies for the motivational design of instructional simulations should include specific strategies for gaining and keeping learner attention, making the simulation relevant, encouraging an optimal sense of self-efficacy in learners, and pointing to the desirable outcomes—intrinsic and extrinsic—of successful accomplishment. In addition, designs may be enhanced by including design features that facilitate social elements in the simulation. Examples of social motivators that may be used in simulations include collaborative features and the use of pedagogical agents. Additionally, Instructional simulation designs may incorporate structured story, quality visual design, and flow designs to enhance motivation.

Learning is the result of actions and choices learners make, and therefore it is dependent upon the motivational engagement of the learner. Effective learning demands more than logical instructional presentation or high learner aptitude, it requires that the learner energetically engages in the learning act. Instructional simulations present unique motivational opportunities and challenges, but these challenges can be met with a motivational design that addresses the key components of motivation.

REFERENCES

[1] P. C. Blumenfield, *et al.*, "Motivation and cognitive engagement in learning," in *The Cambridge Handbook of The Learning Sciences*, K. R. Sawyer, Ed., ed New York, NY: Cambridge University Press, 2006.

[2] J. M. Keller and K. Suzuki, "Use of the ARCS Motivation Model in courseware design," in *Instructional Designs for Microcomputer Courseware*, D. H. Jonassen, Ed., ed Hillsdale, NJ: Lawrence Erlbaum Associates, 1988, pp. 401-434.

[3] D. H. Jonassen, "Designing motivating courseware," in *Instructional Designs for Microcomputer Courseware*, D. H. Jonassen, Ed., ed Hillsdale, NJ: Lawrence Erlbaum Associates, 1988, pp. 397-400.

[4] B. Weiner, *Human Motivation: Metaphors, Theories, and Research*. Newbury Park, CA: Sage, 1992.

[5] R. M. Gagné, *Conditions of Learning*. New York, NY: Holt, Rinehart, and Winston, 1965.

[6] J. M. Keller, "What is Motivational Design?," in *Motivational Design for Learning and Performance*, ed: Springer US, 2010, pp. 21-41.

[7] J. M. Keller, "Strategies for stimulating the motivation to learn.," *Performance and Instruction,* vol. 28, pp. 1-7, 1987.

[8] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change.," *Pscyhological Review,* vol. 84, pp. 191-215, 1977.

[9] J. M. Keller, "Using the ARCS Motivational Process in Computer-Based Instruction and Distance Education," *New Directions for Teaching and Learning,* pp. 39-47, 1999.

[10] E. L. Deci, *Intrinsic Motivation*. New York, NY: Plenum Press, 1975.

[11] H. Astleitner and P. Lintner, "The Effects of ARCS-Strategies on Self-Regulated Learning with Instructional Texts," *E-Journal of Instructional Science and Technology,* vol. 7, 2004.

[12] C. Kim and J. M. Keller, "Motivation, Volition and Belief Change Strategies to Improve Mathematics Learning," *Journal of Computer Assisted Learning,* vol. 26, pp. 407-420, 2010.

[13] S. M. Alessi and S. R. Trollip, *Multimedia for Learning*, 3rd ed. Needham Heights, MA: Allyn and Bacon, 2005.

[14] C. Dormann and R. Biddle, "Humour in Game-Based Learning," *Learning, Media & Technology,* vol. 31, pp. 411-424, 2006.

[15] T. W. Malone, "Toward a theory of intrinsically motivating instruction," *Cognitive Science,* pp. 333-369, 1981.

[16] D. H. Schunk, *et al.*, "Peer Models: Effects on Children's Achievement Behaviors," presented at the 1986 Meeting of the American Psychological Association, Washington, DC, 1986.

[17] J. M. Keller, "The systematic process of motivational design.," *Performance and Instruction,* vol. 26, pp. 1-8, 1987.

[18] E. Erikson, H., *Identity and the Life Cycle*. New York, NY: W.W. Norton and Company, 1980.

[19] J. P. Gee, "The ecology of games: Connecting youth, games, and learning," in *MacArthur Foundation Series on Digital Media and Learning,* K. Salen, Ed., ed Cambridge, MA: The MIT Press, 2008, pp. 21-40.

[20] L. P. Rieber, "Seriously considering play: Designing interactive learning envrionments based on the blending of microworlds, simulations, and games," *Educational Technology Research and Development,* vol. 44, pp. 43-58, 1996.

[21] R. M. Gagne, *The Conditions of Learning*, 2nd ed. New York, NY: Holt, Rinehart, and Winston, Inc., 1970.

[22] R. H. Bruning, *et al.*, *Cognitive Psychology and Instruction*, 4th ed. Upper Saddle River, NJ: Pearson, 2004.

[23] A. S. Palincsar, "Social constructivist perspectives on teaching and learning," *Annual Review of Psychology,* vol. 49, pp. 345-364, 1998.

[24] D. J. Hacker and L. Bol, "Metacognitive theory: Considering the social influences," in *Research on Sociocultural Influences on Motivationa and Learning*. vol. 4, S. Van Etten and D. McInerny, Eds., ed Greenwich, CT: Information Age Press, 2004, pp. 275-297.

[25] W. Bennis, *Organizing Genius: The Secret of Creative Collaboration*. Cambridge, MA: Perseus Books, 1997.

[26] K. Andreas, *et al.*, "Fostering Collaborative Learning in Second Life: Metaphors and Affordances," *Computers & Education*, vol. 55, pp. 603-615, 2010.
[27] A. B. Adcock, *et al.*, "The Impact of Content Area Focus on the Effectiveness of a Web-Based Simulation," *British Journal of Educational Technology*, vol. 41, pp. 388-402, 2010.
[28] L. Ieronutti and L. Chittaro, "Employing Virtual Humans for Education and Training in X3D/VRML Worlds," *Computers and Education*, vol. 49, pp. 93-109, 2007.
[29] B. J. Fogg, *et al.*, "How do people evaluate a web site's credibility?," Stanford UniversityOctober 29, 2002 2002.
[30] A. David and P. Gore, "The impact of design and aesthetics on usability, credibility, and learning in an online environment.," *Online Journal of Distance Learning Administration*, vol. 13, 2010.
[31] P. E. Parrish, "Aesthetic principles for instructional design.," *Educational Technology Research and Development*, vol. 57, pp. 511-528, 2009.
[32] J. M. Keller, *Motivational Design for Learning and Performance: The ARCS Model Approach*. New York, NY: Springer-Verlag, 2009.
[33] R. Schank, *Tell Me a Story: A New Look at Real an Artificial Memory*. New York, NY: Charles Scribner & Sons., 1990.
[34] R. McKee, *Story: Substance, Structure, Style, and Principles of Screenwriting*. New York, NY: Harper-Collins Publishers, 1997.
[35] U. Hasson, *et al.*, "Neurocinematics: The neuroscience of film.," *Projections*, vol. 2, pp. 1-26, 2008.
[36] M. Czikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York, NY: Harper Row, 1990.
[37] Blizzard. (2010, November 25, 2010). *World Of Warcraft Subscriber Base Reaches 12 Million Worldwide*. Available: http://us.blizzard.com/en-us/company/press/pressreleases.html?101007

Low versus High Intensity Emotion in Animated Pedagogical Agents: A Qualitative Study

Author(s): Enilda Romero

Keywords: Emotion, Instructional Simulations, Animated Pedagogical Agents, Intensity

Absract: This paper presents a phenomenological approach to study learners' perceptions of low intensity versus high intensity emotionally expressive animated pedagogical agents in computer–mediated environments. The researcher uses maximum variation sampling and snowball sampling for recruitment of the participants. Data was collected through individual interviews and guided reflection journals. The participants were eight university students that were randomly assigned to one of the possible treatments (i.e. low intensity emotion, high intensity emotion). Results indicated five main themes: perceptions of importance, perceptions of humanness of the agent, perceptions of enjoyment, perceptions of barriers for implementation, and suggestions for improvement. Contributions to the topic, implications for future research and limitations of the study are considered.

Development of Web-Based Engineering, Educational and Assessment Modules for Learning Numerical Methods

Subhash C.B.S.V Kadiam*, A. Mohammed, D T Nguyen
Civil and Environmental Engineering Department, Old Dominion University (ODU)

Abstract: Solving large (and sparse) system of simultaneous linear equations has been (and continue to be) a major challenging problem for many real-world engineering/science applications [1-2]. For many practical/large-scale problems, the sparse, Symmetrical and Positive Definite (SPD) system of linear equations can be conveniently represented in matrix notation as [A]{x}={b}, where the square coefficient (non-singular) matrix [A] and the Right-Hand-Side (RHS) vector {b} are known. The unknown solution vector {x} can be efficiently solved by the following step-by-step procedures [1-2]: Reordering phase, Matrix Factorization phase, Forward solution phase, and Backward solution phase.

With the growing popularity and possibilities of the internet, web-based learning has becoming more and more prominent these days. The new trend focuses on developing more effective learning methods based on large pre-existing scientific languages like FORTRAN, C etc. In this paper, a web-based environment is utilized as a means to introduce numerical methods concepts in civil engineering and other related fields of engineering. Software development and implementation is presented, including detailed descriptions of the techniques employed to link software written in high level computer languages such as FORTRAN and C to a web-based, user friendly interface for both input and output.

Finally, minimizing Fill-in-Terms module, structural matrix method module, online/web-based LU decomposition software for solving symmetrical and non-symmetrical linear equations and web-based symbolic input for integration software tools were described in this paper. Engineering students will have the necessary (online) tool to practice their problem solving skills, and have good assessment tool to self-evaluate their learning capabilities.

*contact author

I. Introduction

Knowing that "knowledge that is taught in only a single context is less likely to support flexible knowledge transfer than is knowledge that is taught in multiple contexts" [9] and to further improve the educational experience and performance, a number of new prototype tools have been developed, implemented and assessed for the topic of SLE under this project. These tools are:

- Online Minimizing Fill-In-Terms Game Based Learning (MFIT-GBL) approach for learning re-ordering algorithm and SLE [4].
- On-line Structural Matrix Method (SMM) software module for problem solving skills and self-assessment tests [6].
- Web-based LU decomposition method for solving symmetrical and non-symmetrical systems of linear equations [7] and
- On-line problem solving skills (related to symbolically input for integration topic) [10].

II. Online Minimizing Fill-In-Terms Game Based Learning (MFIT-GBL) approach [4, 5] for learning re-ordering algorithm and solving Simultaneous Linear Equations (SLE) [1, 2, 4, 5].

Solving large (and sparse) system of SLE has been (and continue to be) a major challenging problem for many real-world engineering/science applications [1-2].

In matrix notation, the SLE can be represented as:

$$[A]\{x\} = \{b\} \tag{1}$$

where [A] = known coefficient (non-singular) matrix, with dimension NxN

{b} = known right-hand-side (RHS) Nx1 vector
{x} = unknown Nx1 vector.

2.1 Symmetrical Positive Definite (SPD) SLE

For many practical SLE, the coefficient matrix [A] (see Eq.1) is SPD. In this case, efficient 3-step Cholesky algorithms [1-2] can be used.

Step 1: Matrix Factorization Phase

In this step, the coefficient matrix [A] can be decomposed into

$$[A] = [U]^T [U] \tag{2}$$

where [U] is a NxN upper triangular matrix.

Various terms of the factorized matrix [U] can be computed/derived as following (see Eq. 2):

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} u_{11} & 0 & 0 \\ u_{12} & u_{22} & 0 \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} \tag{3}$$

For a general NxN matrix, the diagonal and off-diagonal terms of the factorized matrix $[U]$ can be computed from the following formulas:

$$u_{ii} = \left(A_{ii} - \sum_{k=1}^{i-1} (u_{ki})^2 \right)^{\frac{1}{2}} \tag{4}$$

$$u_{ij} = \frac{A_{ij} - \sum_{k=1}^{i-1} u_{ki} u_{kj}}{u_{ii}} \tag{5}$$

Step 2: Forward Solution phase

Substituting Eq. (2) into Eq. (1), one gets:

$$[U]^T [U]\{x\} = \{b\} \tag{6}$$

Let’s define:

$$[U]\{x\} \equiv \{y\} \tag{7}$$

Then, Eq. (6) becomes:

$$[U]^T \ \{y\} = \{b\} \tag{8}$$

Since $[U]^T$ is a lower triangular matrix, Eq. (8) can be efficiently solved for the intermediate unknown vector $\{y\}$.

In general, one has

$$y_j = \frac{b_j - \sum_{i=1}^{j-1} u_{ij} y_i}{u_{jj}} \qquad (9)$$

Step 3: Backward Solution phase

Since $[U]$ is an upper triangular matrix, Eq. (7) can be efficiently solved for the original unknown vector $\{x\}$.
In general, one has:

$$x_j = \frac{y_j - \sum_{i=j+1}^{N} u_{ji} x_i}{u_{jj}} \qquad (10)$$

2.2 Re-Ordering Algorithms for Minimizing Fill-in Terms [1]

During the factorization phase (of Cholesky algorithms), many “zero” terms in the original/given matrix [A] will become “non-zero” terms in the factored matrix [U]. These new non-zero terms are often called as “fill-in” terms (indicated by the symbol F). It is, therefore, highly desirable to minimize these fill-in terms, so that both computational time/effort and computer memory requirements can be substantially reduced.

One would like to solve the following modified <u>s</u>ystem of <u>l</u>inear <u>e</u>quations (SLE) for $\{x^*\}$,

$$[A^*]\{x^*\} = \{b^*\} \qquad (11)$$

rather than to solve the original SLE (see Eq.1).

2.3 On-line MFIT-GBL For Reordering/Factorized Phase [2, 4, 5, 8, 16].

Based on the discussions presented in the previous sections, one can easily see the <u>similar</u> <u>operations</u> between the symbolic, numerical <u>factorization</u> and <u>reordering</u> phases of sparse SLE.

In this work, MFIT-GBL tool (shown in Figure 1 [4, 5]) has been designed with the following objectives:

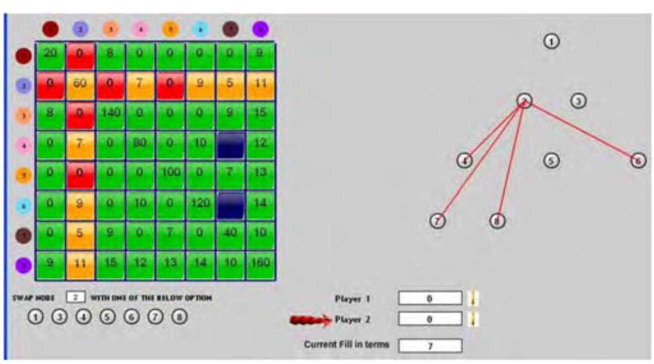


Figure 1: Online MFIT-GBL For Learning to Solve SLE.

(A) Teaching students the process how to use the reordering output for converting the original/given matrix [A] into the new/modified matrix $[A^*]$. This step is reflected in Figure 1, when the “Game Player” decides to swap node (or equation) “i” (say i =2) with another node (or equation) “j”, and click the “CONFIRM” icon!

Since node "i = 2" is currently connected to nodes j = 4,6,7,8; hence swapping node i = 2 with the above nodes j will "NOT" change the number/pattern of "Fill-in" terms. However, if node i =2 is swapped with node j=1, or 3, or 5, then the fill-in terms pattern may change (for better or worse) !

(B) Helping students to understand the "symbolic" factorization" phase, by symbolically utilizing the Cholesky factorized Eqs. (4,5). This step is illustrated in Figure 1, for which the "game player" will see (and also hear the computer animated sound, and human voice), the non-zero terms (including fill-in terms) of the original matrix [A] to move to the new locations in the new/modified matrix $[A^*]$.

(C) Helping students to understand the "numerical factorization" phase, by numerically utilizing the same Cholesky factorized Eqs. (4,5).

(D) Teaching engineering/science students to "understand existing reordering concepts", or even to "discover new reordering algorithms"

III. Structural Matrix Method (SMM) Software Module [6, 12, 13]

The SMM module [6] includes brief reading sections on various components of the SMM process and the theoretical backgrounds behind the developed formulas adopted for calculations. The reading sections are followed by an interactive application unit, which includes the computation of the structural responses (such as nodal displacements, member-end-actions and support reactions), visualization and animation (such as plots of un-deformed and deformed structures) [6], with highlighted observations to enhance students' learning. Students are then assigned exercises, which require both hand calculations and the use of the interactive unit. This pre-processing phase is followed by structural "analysis/computation" (to calculate the structural responses) and the "post-processing" (to display the structural responses in the "graphical" forms) phases (see Figure 2).

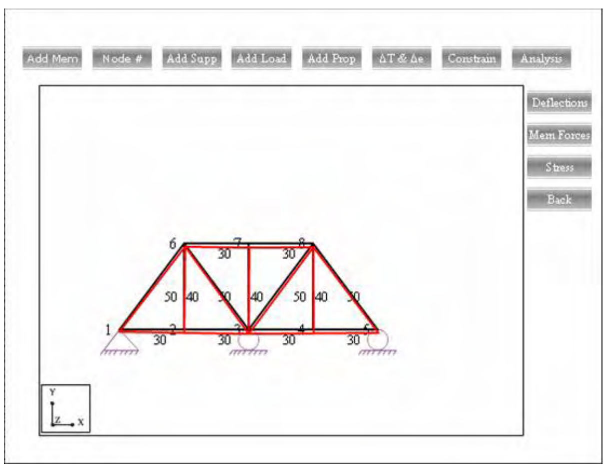


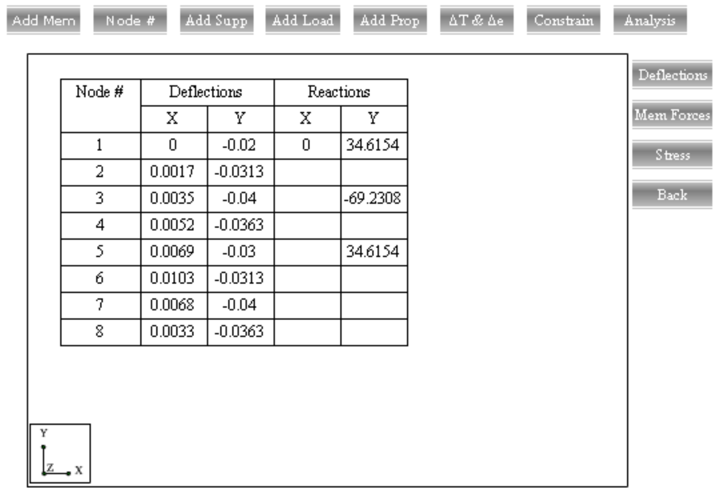
Figure 2: Deflected Shape for Support "Settlements" Bridge (Truss) Example.

Theoretical background for the Stiffness Matrix Method [6, 14]:

The entire Stiffness Matrix Method (SMM) will involve with the following major components:

(a) Element local matrices
(b) Element global matrices
(c) Assembly process
(d) Boundary conditions
(e) Solution of system of linear equations
(f) Structural responses (see Figure 3)

Details of the above key components has been explained and presented in the ODU website [6] (then click on the theoretical development module). More advanced treatments of the above item (e) can be found in Refs. [1, 2, 6].



Node #	Deflections		Reactions	
	X	Y	X	Y
1	0	-0.02	0	34.6154
2	0.0017	-0.0313		
3	0.0035	-0.04		-69.2308
4	0.0052	-0.0363		
5	0.0069	-0.03		34.6154
6	0.0103	-0.0313		
7	0.0068	-0.04		
8	0.0033	-0.0363		

Figure 3: Deflections and reactions for support "settlements" example.

Students Self-Assessment Test (in none_multiple choice style) [6]:

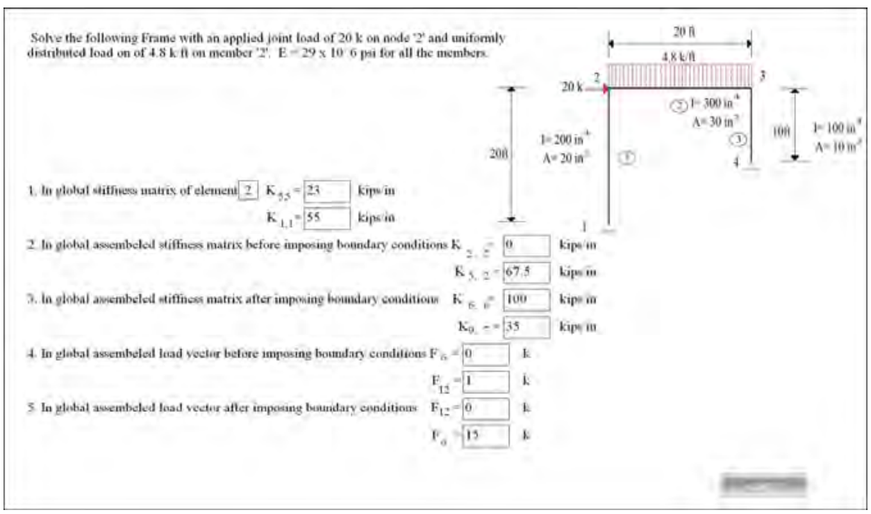


Figure 4: Self-Assessment Test: Frame problem (Students will enter his/her answers in the textbox provided).

Self-Assessment module is a friendly (and critically important) module where students can assess their performance themselves. In this module, separate set of (randomly generated) questions were designed for 2-D Truss, Beam and/or Frame problems. The students have the choice to take the self-assessment test on either 2-D truss, beam, or frame problems. The results of the test are automatically graded and sent to the instructor and student by E-mail. The main advantage of this module is that the student has to compute some "detail, intermediate" variables (see Figure 4) before getting their "final" answers. While the *final answers* could be obtained by the student through the developed "Interactive Simulation & Visualization Module" [6], the *intermediate answers* are "intentionally" unrevealed, for student's self-assessment test purpose! The grading policy adopted in this module is as follows: the student will be awarded 100% for each correct answer, otherwise only 35% for partial credit will be given to him/her for a wrong answer. At the end, the student's average test score is calculated and sent to students' email address, along with the student's answers and the correct answers (automatically generated/printed by the computer, as shown in Figure 5). The grading policy of this module can be changed according to the instructor's choice. This is also a very helpful module for instructor, since the instructor does not have to "painfully grade" students' tests (especially for those classes with high student enrollments!).


Self Test CEE 310 Inbox
Ahmed Ali Mohammed <ahmedali484@gmail.com> show details 2:43 pm (4 r
Name:subhash
UIN:-1234567
Test Type:frame

Student Answers
1:23
2:55
3:0
4:67.5
5:100
6:35
7:0
8:1
9:0
10:15

Actual Answers
1:7.55208333333333
2:3625
3:2416.66666666667
4:1.47664990112384e-13
5:241666.666666667
6:1208.33333333333
7:-1920
8:0
9:0
10:-1920

Score
1:35
2:35
3:35
4:35
5:35
6:35
7:35
8:35
9:100
10:35

Average Scored:41.5

Figure 5: Self-Assessment results and graded score are included in each student's email

IV. On-Line Software (Numerical Method Topics) for Problem Solving Skills [10, 11]

The developed on-line software has the following important/nice features:

- No software license is required.
- The users/learners are not even required to download any software to his/her desktop/laptop computer.
- For numerical methods topics, it was NOT easy/convenient to write the appropriated program for self-assessment tests/quizzes. This is especially true if the students are asked to find "many numbers" (instead of finding just one number) within a question.
- Most (if not all) of self-assessment computer software system do use "random generations" to randomly create "DIFFERENT" numerical values for the "SAME" style of question(s). The users usually are NOT allowed to provide his/her own input data. This is especially true for the "SYMBOLIC" data (such as function expression) in addition to the usual "NUMERICAL" input data.

4.1 LU Decomposition Method for Solving Symmetric and Non-Symmetric Systems of Linear Equations [7, 15]

A screen shot of the LU decomposition module is shown in Figure 6

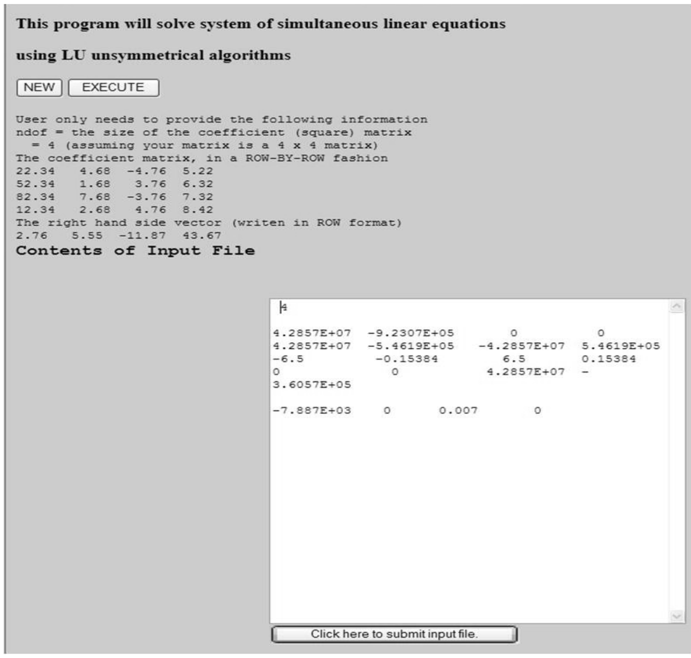


Figure 6: LU Decomposition Method for solving SLE.

4.2 On-Line Symbolic Input for Integration (by multiple traperzoidal rule) Problem Solving Skills [10]

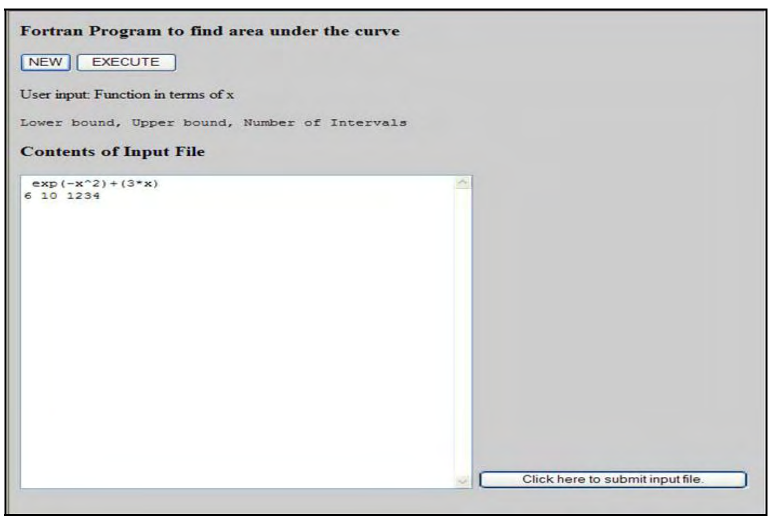


Figure 7: Advance Endless Quiz for “Multiple-Trapezoidal Rule for Integration” (with user’s input symbolic function)

A screen shot for the developed on-line “multiple trapezoidal rule for integration” with user’s specified/input “symbolic function” is shown in Figure 7.

V. Conclusions

A web-based educational software/tool/game intended to help students of engineering course dealing with numerical methods dealing with concepts concerning SLE was developed. The students, in general, do feel that the developed chess-like internet game does help them to understand the fill-in term process which occurs during the most time consuming factorization phase. A fairly long lists of new prototype tools/software/game have been tested to enhance the educational experience and examination performance of STEM (Engineering) students (for a “Structural Analysis” course), and (Engineering/Mathematical) students (for a “Numerical Methods” course). Freely available educational and assessment software's [4, 6, 7, and 10] provide engineering students a convenient learning environment. The developed software/tools/games are believed/expected to be helpful to both STEM students and educators.

VI. Acknowledgements

The authors would like to acknowledge the partial support, provided in this work through the National Science Foundation (NSF Grant #0836916). The authors would also like to thank the reviewers for their helpful comments and suggestions for improvement during the review process.

VII. References

1. Duc T. Nguyen. (2006). *Finite Element Methods: Parallel-Sparse Statics and Eigen-Solutions*, Springer Publishers.
2. Duc T. Nguyen. (2002). *Parallel-vector Equation Solvers for Finite Element Engineering Applications*, Kluwer Academic/Plenum Publishers
3. www.brothersoft.com/downloads/flash-animation-software.html.
4. http://www.lions.odu.edu/~amoha006/Fillinterms/FILLINTERMS.html
5. www.lions.odu.edu/~skadi002 (then CLICK/select CEE-305 course materials, and see YOUTUBE video items #23, #25).
6. http://www.lions.odu.edu/~amoha006
7. www.lions.odu.edu/~skadi002/work/demo1.php
8. http://youtube.com/civilengineeringprof
9. *How People Learn: Brain, Mind, Experience and School* , J. D. Bransford, et al, 1999, p. 66.
10. http://www.lions.odu.edu/~skadi002/work/symbolic_integration.php
11. Bradford G. Singley, Rollin H. Hotchkiss, Jason K. McDonald. *Computer-Based Instruction for Engineering Education in the Developing World*. (Internet source)
12. Andres W.C Oreta. *Developing a Web-Based Learning Module in a Basic Civil Engineering Course, Computer Applications in Engineering Education*, Vol 7, Issue 4.
13. Abi Aghayere. *A Web-Based Distance Learning Course in Structural Analysis*, Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition.
14. S.D Rajan, Introduction to Structural Analysis and Design, John Wiley & Sons, Inc, 2001.
15. Banks, C.M, Sokolowski, J A. *Meeting the Challenges of STEM Education: Teaching Modeling and Simulation with Real-World Applications*. In Proceedings of the Spring Simulation Conference, April 11-15 2010, Orlando, FL pp. 146-151.
16. Emily Oh Navarro, Andre van der Hoek, *SIMSE: An Interactive Simulation game for Software Engineering Education*.(Internet Source)

General Science

Track Lead: Dr. Joshua Behr, VMASC

Judge: Mr. Charles Turnitsa, VMASC

Towards an Appropriate Population Model for the United States: A System Dynamics Approach
Author(s): Mandar Tulpule and Rafael Diaz
Abstract Only
1st Place Gene Newman Award Winner, General Science Track

The Model of Distributive Mechanics
Author(s): Ryan Roberts
2nd Place Gene Newman Award Winner, General Science Track

Mobilizing Maternal Health: Modeling Rwanda's Plan to Reduce Maternal Mortality Rates
Author(s): Erika Frydenlund, David Earnest
3rd Place Gene Newman Award Winner, General Science Track

Towards an Appropriate Population Model for the United States: A System Dynamics Approach

Author(s): Mandar Tulpule and Rafael Diaz

Keywords: Population, System Dynamics, Structural Modeling

Abstract: Estimation of present and future population trends and characteristics is essential for public bodies in order design appropriate public policies. Population trends and characteristics are also of interest to private organizations from marketing strategy point of view. The population of any nation can be considered as one of the major driver of resource consumption other notable factors being per capita consumption and efficiency of consumption .This paper presents a system dynamics approach towards modeling the population dynamics in the United States. The model is constructed as a combination of the most popular cohort component method and an alternative method called structural modeling. Two major aspects namely age and citizenship status and taken into consideration. While age is the most recurrent factor in population models, immigration is also deemed critical in this particular case, since the dynamics of immigrant population are different than those of the natives. This model would enable the understanding of how the US population has evolved over the simulation period and forms a basis for reasoning about future trends. More importantly, this model serves as a foundation for modeling current and future demand in critical areas like healthcare and energy, which would be prerequisite for building models addressing specific issues in these areas.

The Model of Distributive Mechanics

Ryan J. Roberts
Old Dominion University, M.A. International Relations
University of Toledo, MBA International Business

***Abstract*— One of the most interesting characteristics in the pre and post-Westphalian eras has been the integration and fragmentation of the state system. Evolutionary processes have contributed to cooperative behavior among collectives as social clusters have internalized normative values and formed larger structures e. On the other hand, non-cooperative actions have resulted, in part, to the dissolution of these structures into smaller factions. The simultaneity of this phenomenon is the basis of the Model of Distributive Mechanics which incorporates the collective reconciliation between security and identity under the conditionality of two dichotomous environments. In this N-person constant-sum game, sub-national units, states, and the system of states interact on the assumption of individual and group rationality in order to maximize their utility over a set of given spatial preferences.

***Index Terms*— identity, integration, fragmentation, N-person game, security

I. Introduction

The Treaty of Westphalia in 1648 was a seminal moment in history that was a collective action response to security provision.[1] States were conceived through multilateral bargaining and an acknowledgement of mitigating the effects of a hostile European environment; chiefly the Thirty Year's War. As a result of establishing universally recognized sovereign boundaries, a plethora of sub-national units were incorporated into the state apparatus. Moreover, the inception of the state merged national identities and cultures under a single demarcated territory; which, according to Benedict Anderson (1983),[2] propagated the concept of nationalism. Moving away from the epicenter of the state model, European powers then exported and imposed their concept of delineable boundaries of sovereignty establishing arbitrary borders in large segments of the world. Thus, states can be observed as a myriad of collectives distinguishable by their sub-national identities.

Since the formulation and implementation of the tenets imbued in the Treaty of Westphalia, the state system has experienced the ebbs and flows of integration and fragmentation. On the one hand some states have chosen to integrate and form larger structures such as the European Union, while others seek national self-determination and attempt to secede from the state in order to assert their identity in the global order. The Model of Distributive Mechanics is based on an N-person game that draws upon bargaining and kernel models to explicate the integrative and fragmentative dynamics of the state system. The focus of this paper is centered on the interaction and behavior of three comprehensive entities using NetLogo, an agent-based modeling software. The entities considered are the state, the sub-national units within the state, and the system of states as a whole. Further, the simulation incorporates the trade-off between security and identity tested against two contrasting environments, Hobbesian and Kantian, which, in turn, determine state types.

Neorealists such as Kenneth Waltz (1979) postulate that states operate in an anarchic state system where cooperation is

[1] Lillian Goldman Law Library. "Lillian Goldman Law Library, The Avalon Project: Yale Law School." Lillian Goldman Law Library. http://avalon.law.yale.edu/17th_century/westphal.asp (accessed March29, 2011).

[2] Anderson, Benedict. *Imagined Communities*. London: Verso, 1983.

limited to balancing against uneven distributions of power. Under conditions of anarchy, states interact in a self-help system which causes them to converge on functionally similar poles. The mainstay of this theory focuses on the systemic level, largely ignoring subsystemic actors, and assumes security is the single most important concern for states. Thus, states will focus on relative gains and are less inclined to cooperate if other states are perceived to receive higher gains, or advance their power position more, as a result of the relationship.[3]

The constructivist argument, as posited by Alexander Wendt (1999), treats environment as intersubjective fact. A self-help system, for instance, is an outcome produced by the objectification and exteriority of state behavior as opposed to anarchy. Repetitive behavior, without a change in meaning, and behaving as if a condition is immutable fact, respectively, reinforces the nature of the state system as a self-help environment. However, social structural factors such as the distribution of knowledge, identity, and interests contribute to shifts in the perception of ‘self’ and ‘other’.[4] In other words, state interaction leads to self-conceptions and understanding. Thus, cooperation can be learned through iterative processes.

Conclusions about the validity of the hypotheses will be measured by the sum of the total payoffs received as a result of the strategic choices made by players. There are five sections that compose the body of this paper. The first section of the paper describes the Model of Distributive Mechanics followed by a methodological description of the simulation procedures and limitations. Section four highlights the relevant literature of renowned scholars that have contributed to this body of research; more specifically, the different models and approaches that have been used to simulate social clustering. The final sections present the simulation results and conclusions which include a discussion of the implications of the model relative to the neorealist and constructivist predictions of state behavior, and additional applications of the model.

II. The Model of Distributive Mechanics

The Model of Distributive Mechanics is compounded using bargaining and kernel models with three distinct entities whose payoff and utility distributions are impacted by individual and group rationality in accordance with predefined systemic parameters.

A. Bargaining and Kernel Models

Coalition formation can be observed as a series of bargaining strategies among actors for the purpose of earning a higher utility than would otherwise be attainable as an independent entity. The characteristic function form of a three person constant-sum game is represented by

$$v(\cancel{1}) = v(\cancel{2}) = v(\cancel{3})$$
$$v(\cancel{12}) = v(\cancel{13}) = v(\cancel{23})$$
$$v(\cancel{123}) = 1$$

where,

$$x_i \geq v(\cancel{i}) \text{ and } \sum_{i \in X_j} x_i = v(X_j),\ j = 1, 2, m$$

indicates the individual rationality payoff configuration. A winning coalition is one in

[3] Waltz, Kenneth. 1979. *Theory of International Politics*. New York: McGraw-Hill.
[4] Wendt, Alexander. 1999. *Social Theory of International Politics*. Cambridge: Cambridge University Press.

which there exists an equitable distribution of payoffs that has been attained by coalition members where no other party can object or intercede, and neither member has an incentive to deviate from the agreed distribution. In other words, if parties A and B enter into a winning coalition, then party C has nothing to offer either A or B that could further maximize their utility. The precept of this model is centered on individual rationality.

The kernel model, on the other hand, introduces decisions based on excess. When entering iterative play, excess is a determinant of an individual's willingness to seek higher payoffs by leaving an existing coalition to form an independent entity or enter into a new coalition. Excess with regard to the payoff configuration (x:X) is denoted by

$$e(Y) = v(Y) - \sum_{i \in Y} x_i$$

where the coalition structure X is abandoned for the new coalition Y. The maximum surplus of an entity k over l is

$$s_{k,l} = \max_{Y \in rk,\, l} e(Y)$$

for the set of all coalitions that contain k but not l

$$\mathscr{I}_{k,l} = \{Y \mid k \in Y,\ l \notin Y\}.$$ [5]

B. Agents

In this N-person game, there are three primary agents that make permissible the act of integration or fragmentation. One agent, the state, acts to increase both relative and absolute gains through integration to command a larger power position within the state system. At the same time, the state seeks to mitigate the national self-determination of its collectives to maintain its global status. When the cost of doing so, however, exceeds the benefits of preserving segments of the state, granting independence becomes a strategic option.

Another agent considered in this model are the sub-national units that compose each state. Sub-national units seek to increase their utility derived from the state as well as their relative power position within the state. When the utility of the state is reduced below a minimal threshold, secession attempts become more probable. Moreover, state integration hinges on the collective agreement of these sub-national units. The final agent, or set of agents, is the amalgam of states within the system that have the authority to veto or ratify the national self-determination of sub-national units.

To put the interaction of these agents into perspective, consider two examples; South Ossetia and South Sudan. Situated in the Tskhinvali region of Georgia, South Ossetia has been a contested territory whereby the Georgian government maintains control over the area in which Ossetians have a strong preference for sovereignty. In recent decades, conflict has escalated exacerbating hostilities between both sides as secession attempts have been thawrted. Independence referendums have been submitted twice on behalf of South Ossetia, and with the exception of a few states, have been rejected.[6] Thus,

---

[5] Kahan, James P. and Amnon Rapoport. 1974. "Test of the Bargaining Set and Kernel Models in Three-Person Games." In *Game Theory as a Theory of Conflict Resolution*, ed. By Anatol Rapoport, 120-160. Dordrecht, Holland: Reidel Publishing company.

[6] Mchedlishvili, Niko. 2006. "Georgian Rebel Region to Vote on Independence." *Kurdish Globe*,

illustrating independence requires both state approval as well as international recognition to legitimze a sub-national unit as a sovereign entity.

The Republic of Sudan, on the other hand, has historically been charcterized by civil war between the North and South regions. Civil unrest perpetuated by military conflict has only recently led to a cooperative agreement between both sides that will permit the secession of the Southern Sudanese from the Northern part of the country. In this example, the international community has approved the independence referendum of South Sudan that will establish the territory as an autonomous state.[7] While South Ossetia and South Sudan have sought self-determination, integration continues to proliferate within geographic spaces proximate to both nations as the European Union continues to include more states and East Africa seeks to sustain the free trade area with the East African Community (EAC), respectively.

C. Decisions, Utilities, and Payoffs

Each agent within the system is assigned a specific point value. Point values increase or decrease based on the strategic choices of individual entities as well as collective decisions that affect payoff distributions. Integration, for instance, is played as a modified version of the Divide-the-Dollar game. The first decision node begins with player one deciding to maintain the status quo and making no offer in which both players receive a payoff of zero, or making an offer to player two on the distribution of the total points from both players (the points are distributed to the sub-national units). Player two can then decide to reject the offer and preserve his status quo while player one incurs a small cost, make a counter-offer to player one, or accept player one's offer. If a counter-offer is made, player one has the option of reject in which both players will incur a small cost, or accept in which both players receive the agreed distribution of points minus a small cost for player one due to the extended duration of negotiations.

When an offer or counter-offer is accepted, the sub-national units maintain veto power. Based on their respective power positions, the sub-national units of each player will form coalitions that decide whether to accept or reject the offer. The total point value of the coalition assigned to the final decision determines whether integration is accepted or rejected and the subsequent payoffs assigned. For instance, the inclusion of Mexico into NAFTA was contingent upon the politicking of opposing sides whose outcome was finalized through democratic processes.[8] Figure 1 illustrates the extensive form N-person game of integration where each node represents a decision on the game tree for a specific player with termination points that indicate the subsequent payoffs for a given strategy configuration.

September 11. http://www.kurdishglobe.net/display-article.html?id=D24CB8C311CE6D3ABA82D1C52E948435 (accessed February 18, 2011).

[7] BBC News. 2011. "South Sudan Backs Independence – Results." *BBC News*, February 7. http://www.bbc.co.uk/news/world-africa-12379431 (accessed February 18, 2011).

[8] Skonieczny, Amy. 2001. "Constructing NAFTA: Myth, Representation, and the Discursive Construction of U.S Foreign Policy." *International Studies Quarterly,* 45(3), 433-454.

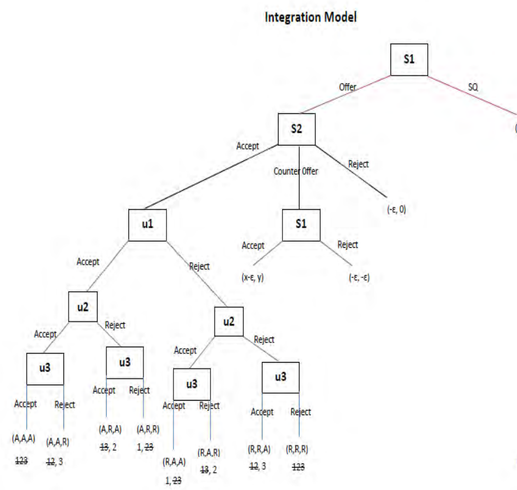


Figure 1

War as a possible mode of integration adds to the available pool of strategic decisions for each state. A state can either attack or not attack. If a state attacks, the defending state can either submit in which he retains a fraction of his total point value and is incorporated into the attacking state as a sub-national unit where the remaining points are absorbed, or defend his territory. When the defender chooses the latter strategy, the attacker can either press on or retreat. For a strategy of press on, an attacker is awarded the sum point total of the defending state less the cost of war, and the defender is eliminated from the game. If, however, the attacker retreats, then the defender earns a payoff of 10 for successfully defending his territory while the attacker receives a payoff of -5 to account for resource mobilization and reputational costs. Figure 2 illustrates the extensive form game of integration through the application of war.

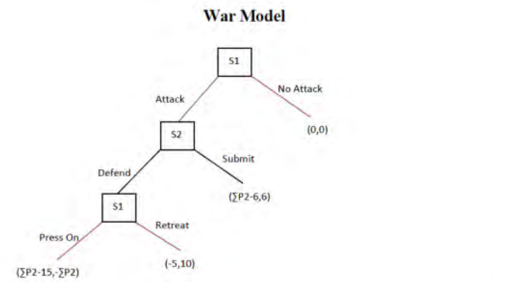


Figure 2

Unlike the integration model that begins externally with state interaction and confirmed or opposed internally through the coalition formation of sub-national units, the fragmentation model is initiated by the sub-national unit, decided by the state, and either ratified or vetoed by international recognition. Figure 3 illustrates the extensive form game tree of the fragmentation model (The decision to grant independence by the state is followed by state coalition formation which is not represented here, but follows the same logic as figure 1 where coalitions are formed by sub-national units. In this case the states within the system maintain veto power).

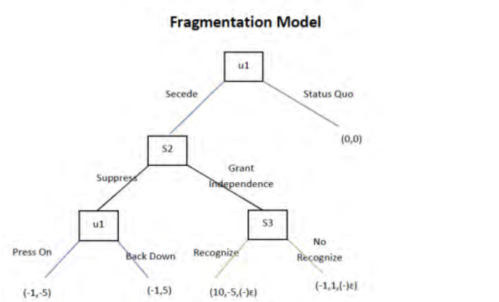


Figure 3

D. Variables and Environments

There are two immutable characteristics that remain omnipresent in social clustering; the concessions and weighted preferences given to security versus identity. Ultimately sub-national units must determine the cost-calculus of remaining a member of the state and receiving the benefits of public goods such as security, or pursuing national self-determination to assert their identity in the global order.

The two types of environments considered are Hobbesian and Kantian. A Hobbesian world represents a hostile environment and adds the dimension of war as a means of integration, or in some cases, for weaker entities, extermination. The Kantian environment, on the other hand, is more pacific omitting the possibility of war. Polarizing the environments serves the purpose of measuring the choices made by entities under duress, and eliminating that threat to determine if environment influences the decisions of integration or fragmentation.

Both environments have a constant competitive characteristic. Hobbesian types obligate entities to focus on self-preservation (physical security) while Kantian types compete for resources and markets (economic security). Further, environment contributes to state types the same as state types foster a specific environment. In a Hobbesian environment, for example, man lives in "continual fear and danger of violent death; life is solitary, poor, nasty, brutish, and short,[9]" which is reverberated through state interaction. Conversely, the Kantian Triangle of peace (international commerce, domestic democratic institutions, and international organizations) generates a more pacifying affect.[10] In short, state interaction acts as a determinant in establishing the structural conditions with which all other states are subjected and function.

[9] Hobbes, Thomas. 1651. *Leviathan*. Trans. A.R. Waller. Cambridge: University Press.

[10] Kant, Immanuel. 1795. *Perpetual Peace*. Trans. M. Campbell Smith. London: Swan Sonnenschein & Co.

III. Methodology

The Model of Distributive Mechanics was simulated using NetLogo agent-based modeling software. NetLogo is equipped with the tools conducive to executing the aforementioned commands in addition to providing the flexibility necessary for simulating dynamic systems where the agents are capable of making decisions within a programmable set of behavioral parameters that govern social interaction.

The initial conditions of the simulation include five states that are each assigned three collectives. Additionally, each state is assigned an initial point value which is allocated to their respective collectives. Interaction and decision-making was coded to the specifications provided by the extensive form game trees that appeared in the preceding section. As a control mechanism rewards (δ) and punishments (Є) for integration were adjusted using two sliders that each ranged from values one to five.

The crux of the simulation was executed using a genetic algorithm. When selecting a strategy, player decisions went through a three stage process; selection, cross-over, and mutation.[11] In the selection stage players randomly compared a combination of two strategy sets from the available decisions. The strategy sets that maximized a player's payoffs were retained and entered the cross-over stage where strategy sets were combined with other "winning" strategy sets from the selection stage. Finally, in the mutation stage, a string from a player's cross-over strategy set was randomly replaced with an alternative

[11] Axelrod, Robert. 1987. "The Evolution of Strategies in the Iterated Prisoner's Dilemma." Ed. by Lawrence D. Davis. *Genetic Algorithms and Simulated Annealing*. San Francisco: Morgan Kaufmann.

strategy choice. Each player in the simulation went through this process of selection, cross-over, and mutation for 100 generations of gameplay. Within each generation, 100 rounds of the model were executed, plotted, and recorded using NetLogo's Behaviorspace which tracked the subgame and outcome of the simulation. Thus, player performance was measured based on observing the final payoffs accumulated over the span of each generation.

There are some limitations of the software and model. Despite the applicability of the software, NetLogo is run by allowing entities to interact simultaneously. For the purposes of this simulation, however, the simulation was coded for sequential play by randomly selecting different states to make the first move. Further, the strategic choices made by the entities were indistinguishable. The reporting system of NetLogo, for example, was not programmed to determine the strategy preferences based on state power distributions.

Other than the integration model, the other two models constructed to execute the simulation were based on a set of fixed payoffs. Future trials of the simulation should incorporate payoffs to varying degrees to simulate the diverse set of spatial preferences and utilities assigned to outcomes by state and non-state actors. Finally, the resolution of the model should be increased to ascertain a more palpable outcome (i.e. assigning attributes to collectives to establish the strategic value of allowing or not allowing national self-determination).

IV. Relevant Literature

The purpose of agent-based modeling is to analyze the impact of micro-level phenomena in the production of macro-level outcomes.[12] In recent years, there have been several models constructed in the domain of identity to simulate social clustering. Robert Axelrod (1995) offers the inverse position of the distributive mechanics model with his Tribute Model. Whereas the Model of Distributive Mechanics takes, as a starting position, large structures that, under certain conditions, dissolve, the tribute model illustrates the effects of commitments in social clustering to create emergent political actors.[13] Bibb Latané, Andrzej Nowak, and James Liu (1994) focus on three parameters that contribute to the aggregation and disaggregation of human systems; dynamism, polarization, and clustering. The combination of these three measurements, according to their research, propagates social behavior over time and space.[14] Finally, Ian Lustick (2000) uses the Agent-Based Identity Repertoire Model to test the constructivist theory of identity by adjusting landscapes, payoff configurations, and incentive structures. His simulation focuses on the sustainability of identities given micro-level interactions among entities.[15]

---

[12] Pepinsky, Thomas B. 2005. "From Agents to Outcomes: Simulation in International Relations." *European Journal of International Relations*, 11(3), 367-394.

[13] Axelrod, Robert. 1995. "A Model of the Emergence of Political Actors." in Nigel Gilbert and Rosaria Conte (eds.) *Artificial Societies: The Computer Simulation of Social Life*. London: University College Press.

[14] Latané B., Nowak A., & Liu J. 1994. "Measuring Emergent Social Phenomena: Dynamism, Polarization and Clustering as Order Parameters of Social Systems. *Behavioral Sciences*, 39, 1-24.

[15] Lustick, Ian. 2000. "Agent-based Modeling of Collective Identity: Testing Constructivist Theory." *Journal of Artificial Societies and Social*

V. Simulation Results

After running ten trials of the simulation for each combination of δ and Є, players consistently improved their payoffs when comparing generation 1 to generation 100. Regardless of the reward or punishment of integration, the majority of simulation runs produced a zero-sum outcome; Pareto-improving when compared to the initial strategy choices of the players. In other words, over the course of the simulation, players learned to improve their payoffs. For instance, the highest outcome (trial 6 simulation 5) with an average payoff of 8.916, war was played 54% of the time in generation 1, but abandoned as a strategic option after generation 29. The lowest performing simulation (trial 9 simulation 4), on the other hand, generated an average payoff of -63.512 with war being the most preferred strategic choice, and the equilibrium solution after generation 24. Additionally, the shape of the learning curve for this trial run was declining.

VI. Conclusion

Overall, the simulation supports the arguments posited by Alexander Wendt. Social interaction through iteration contributes to conceptualizing 'self' in relation to 'other'. Insofar as the behavior of states establishes the intersubjective criterion that guides social interaction, cooperation *can be* learned to produce mutually beneficial outcomes. Furthermore, a diminished capacity for a Hobbesian environment in favor of the conditions outlined by Immanuel Kant incentivizes a more equitable distribution of integration and fragmentation. In a Hobbesian environment, however, integration is more ubiquitous suggesting that security concerns galvanize states to cooperate as a self-preservation mechanism. Finally, the results of the simulation indicate that, for the most part, these are games of common aversions where players "have a common interest in *avoiding* a particular outcome. These situations occur when actors with contingent strategies do not most prefer the same outcome but do agree that there is at least one outcome that all want to avoid.[16]" The Model of Distributive Mechanics contains within it multiple equilibria points for the various strategic choices. Players demonstrated, consistently with each generation, a capacity to learn and avoid receiving least preferred outcomes through coordination.

The model presented is not limited to the realm of political science. The same concepts, constraints, and conditions discussed here can be leveraged to examine firm behavior. It offers a way to conceptualize mergers and acquisitions, the dissolution of strategic business units to focus on core competencies, and emergent threats such as increased competition from market liberalization. Overall, how do firms manage the utility derived from security (merging with and/or acquiring other firms to create synergies and mitigate perceived threats) and identity (corporate image and recognition with which customers can readily associate) to maximize their payoffs?

Simulation 3(1). http://jasss.soc.surrey.ac.uk/3/1/1.html.

[16] Stein, Arthur A. 1983. "Coordination and Collaboration." Ed. by Stephen Krasner. *International Regimes*. Ithaca: Cornell University Press.

Mobilizing Maternal Health: Modeling Rwanda's Plan to Reduce Maternal Mortality Rates

Erika F. Frydenlund

***Abstract*—Under pressure to meet the United Nations Millennium Development Goal 5 to reduce maternal mortality rates by 75% and provide universal access to reproductive health care by the year 2015, the Rwandan government began to explore technologically savvy ways to reach its rural women. In addition to improving access to medical training for midwives and skilled birthing attendants, the government began to equip maternal community health workers stationed in villages across the country with mobile phones. These phones allow medical technicians to update expectant mothers with antenatal health care, track maternal health records in regional databases, and access emergency health services during labor complications. After a brief overview of Rwanda's journey to address its high maternal mortality rate, this study will examine the use of mobile phone technology to improve public health. Using Agent Based Modeling techniques, the paper will analyze the viability of expanding Rwanda's pilot program to all areas of the country with implications for use in other developing states.**

***Index Terms*—Agent Based Modeling, Maternal Health, Rwanda, UN Millennium Development Goals, mHealth**

I. Introduction

The United Nations Millennium Development Goal (MDG) 5 is broadly to "improve maternal health," though the specific target is to "reduce by three quarters, between 1990 and 2015, the maternal mortality ratio [and] achieve, by 2015, universal access to reproductive health care [1]." In 1990, at the start of implementing MDG 5, Rwanda's maternal mortality ratio was approximately 1,100 deaths per 100,000 live births [2]; this was among the highest in the world [3]. In 2005, a World Health Organization (WHO) report estimated that 35% of deaths among females of reproductive age were the result of maternal complications [4]. A 2006 report noted, "reducing the maternal mortality rate towards the Millennium Development Goal in 2015 is perhaps the greatest challenge for the [Rwandan] Government in the health sector owing to the high cost of delivering quality major obstetrical services…To achieve these goals, it will be necessary to massively increase access to basic health services for the indigent population [5]." Though great progress has been made in Rwanda since 1990 in improving maternal health, there are conflicting reports about whether the country is on track to meet MDG 5 by 2015 [6],[7].

Not deterred by the odds stacked against it, the Rwandan government began to implement policies to address the high maternal mortality rate. First, the shortage of women's health professionals was addressed by making training more accessible. As of 2000, over 90% of deliveries in Rwanda took place in the home, overseen by a traditional midwife whose training was largely a combination of experience and knowledge passed from previous generations [8]. In 1996, the Kigali Institute of Health opened its Department of Midwifery to train skilled maternal health professionals that could be posted in clinics across the country. To date, approximately 100 midwives have received an "advanced diploma in midwifery" in this program; though this is inadequate to cover the nation's need. According to the Kigali Institute of Health, there should be approximately one midwife to a population of 500 women of reproductive age; the current ratio is closer to 1:29,000 [9]. Despite increased accessibility to midwifery training, women in rural areas still must travel great distances to reach proper medical care, and often suffer complications and die along the way. In response to this, academic studies and grass roots efforts have begun looking into ways to bring antenatal care and clinic access to rural women, though these initiatives are still in their infancy [10].

Realizing that specifically focusing on trained midwives was not enough, particularly because of the three year time commitment to coursework, the Rwandan government, in 2008, launched a program to enhance the skill of Traditional Birthing Attendants (TBAs). Though they are prohibited by law from directly delivering babies, the newly certified TBAs are trained to monitor pregnancies and educate women about antenatal care [11]. The role of trained women to oversee antenatal care cannot be underestimated. Beyond providing basic healthcare to expectant mothers, TBAs also can provide critical safety precautions to prevent the spread of HIV from infected mothers to infants [12]. In October of 2009, Rwanda made further progress toward MDG 5 by establishing the National Nurses and Midwives (NCM) Board "that is charged with regulating 6,000 nurses and midwives across the country" which involves ensuring professional standards of nursing and midwife education and certification [13]. Despite these advances, progress in accumulating skilled health

Manuscript received February 26, 2011.
Erika F. Frydenlund is a doctoral student at Old Dominion University, Norfolk, VA 23329 USA. (email: efryd002@odu.edu)

professionals is slow, and retaining them at clinics in remote areas is difficult. Though there are now more health professionals available to the community, the Rwandan government struggles with incentivizing posts in rural areas; after an initial post at a small clinic, trained health professionals often migrate to higher paying jobs at hospitals in the larger cities [14]. Though 24% of graduating nursing students claim they would like their first post to be in a rural area in order to "help the poor," one study approximates that the salaries at rural clinic postings would require an 80% increase to ensure retention [14].

Despite the many logistic obstacles facing the Rwandan government in reaching the UN defined targets for MDG 5, the country is committed to transforming into a "middle income country" which requires a change from a subsistence agricultural economy to a "knowledge-based society." In order to achieve this goal, known officially as *Rwanda Vision 2020*, the government publicly acknowledges the need for "pro-poor" plans to address the needs of Rwanda's most economically disadvantaged citizens. One of the pillars of the *Vision 2020* is "Comprehensive human resources development, encompassing education, health, and ICT (Information Communications Technology) skills aimed at public sector, private sector and civil society…to be integrated with demographic, health and gender issues [15]." Though *Vision 2020* covers a wide array of goals for the economy, health, education, and gender equality, this pillar makes clear the government's commitment to use technology to address some of its major issues. To this end, the government has adopted an innovative strategy to address the high maternal mortality rates in Rwanda through the use of mobile phones and mHealth technology.

II. mHealth and MDG 5

The "m" in mHealth stands for mobile-based or mobile-enhanced, referring to the use of mobile devices to deliver health care solutions [16] and is a subset of the larger field of eHealth. eHealth refers to the use of information and communications technologies (ICT) in administering health care. According to some, "eHealth is the single-most important revolution in healthcare since the advent of modern medicines, vaccines, or even public health measures like sanitation and clean water [17]." mHealth allows a direct connection between healthcare providers and patients regardless of geographic proximity, making it ideal for reaching out to rural populations [18]. Additionally, mHealth can allow regions with a limited supply of trained healthcare professionals to reach a larger number of patients. According to the UN Foundation, "Better data on health worker supply and demand can help allocate human resources appropriately across geographies and specialties. This is particularly important where there is a shortage of healthcare workers and only a portion of those are adequately trained [19]." The mHealth Alliance website states, "The UN's Global Strategy for Women's and Children's Health Innovation Working Group has endorsed mHealth and Public-Private Partnerships as the key innovations to improving maternal health outcomes [20]." With the UN's endorsement, mHealth has increasingly become an integral tool for providing healthcare in developing countries.

Rwanda's plans for future healthcare provision include ensuring that each village has at least three trained healthcare providers with specialties in maternal, child, and community health [21]. Eventually, each of these healthcare providers will be equipped with a mobile phone. In a country where 85% of women live in rural communities, far from hospitals with staff trained in obstetrics [22], mobile phones provide an invaluable link to limited health care resources. In addition to the provision of mobile phones, Rwanda instituted a requirement that expectant mothers report directly to local community health providers. Additionally, they are encouraged to participate in antenatal classes three months before their due date.

Though other countries such as Peru run mHealth initiatives to improve maternal and infant health, "Rwanda has emerged an e-health leader [23]." And the Rwandan government has high hopes for the expansion of its mUbuzima, its mHealth program, as is evident from the Ministry of Health's website:

> mUbuzima is an application that builds on Rwanda's mobile phone infrastructure to show how this infrastructure can support Community Health Workers. It allows Community Health Workers to enter and transmit CHIS [Community Health Information System] indicators in real time- even in remote parts of the country using only a mobile phone…The Ministry of Health can also send educational messages to Community Health Workers. In the future, it is envisioned that the system will provide decision support directly to Community Health Workers to enhance the delivery of maternal and child health services [24].

In one pilot program, several private companies established a mHealth capability in Mayange, about twenty kilometers from Kigali. The project features thirty phones, a computer database, and a solar charger. This clinic's mHealth capabilities allowed them to retrieve medical records from an online database and assist local caregivers with instructions during labor emergencies. Additionally, the phones were equipped with a maternal and childcare training manual complete with visual and audio directions available to expectant mothers and families [25].

In August 2008, the Rwandan government began a pilot mHealth program by distributing 432 phones to community health workers in Musanze district [21]. These phones help health workers provide antenatal care, report data about antenatal health and birth outcomes, and report complications or emergencies to health institutions nearby [21]. During a press release for this pilot program, Dr. Richard Sezibera, Health Minister, explained, "Community Health Workers will ensure that all expecting mothers especially those in rural areas get skilled care during pregnancy and childbirth, including access to emergency obstetric and newborn care, postnatal care, preventing HIV and malaria during pregnancy [26]." The healthcare workers are provided with different numbers to use during different situations: 912 to report an emergency case and access one of the standby ambulances; 115 to submit health and birthing outcome reports; 114 to report negligence at health centers and district hospitals [26]. During this pilot study, only two maternal-related deaths were reported, compared with ten in the same time period the previous year. The study notes that these two women did not report to the community health provider during pregnancy [22].

Though there is much work left to be done in the area of maternal health in Rwanda, it serves as a role model to other developing countries for how to utilize technology to maximize resources and overcome geographical obstacles to healthcare provision. The following study models Rwanda's mHealth initiative to reduce maternal mortality to rural and dispersed populations using Agent Based Modeling techniques.

III. The Method

When considering the research question of how well mHealth accessibility addresses maternal mortality rates in Rwanda, an Agent Based Modeling (ABM) approach is an appropriate tool. Joshua Epstein states that agent-based models are artificial societies where agents exhibit heterogeneity, autonomy, bounded rationality, and interact in a local, explicit space [27]. Rather than assume that all women of reproductive age will behave in exactly the same way, ABM allows the experimenter to consider a set of behaviors relevant to an individual's reproductive choices. Epstein suggests that to conduct an ABM experiment one should:

> Situate an initial population of autonomous heterogenous agents in a relevant spatial environment; allow them to interact according to simple local rules, and thereby generate – or "grow" – the macroscopic regularity from the bottom up. (p 7)

In this case, the "initial population of autonomous heterogeneous agents in a relevant spatial environment" represents a population of reproductive age women within a geographical space- perhaps a village or district- that, according to simple probabilistic rules, become pregnant and have children. In this particular model, there is also a population of Birthing Attendants (BAs) to assist pregnant women outside of health facilities. Maternal death rates vary based on the type of delivery accessible to the woman. This is simplified in the model to three types of delivery: in a hospital (low maternal mortality), with a mobile phone equipped Birthing Attendant (relatively low maternal mortality), and all other cases (alone or with unskilled birthing attendants).

When addressing experiments involving healthcare to the public, questions of ethics arise as one group becomes privy to treatments that another group does not. In the case of mHealth, an additional consideration is cost effectiveness of implementation. Before a government can begin implementing mHealth to reduce maternal mortality, it is most efficient to experiment with its effectiveness and the required resources to run such an effective system. ABM allows for these types of considerations. Since it is only a simulated population being experimented upon, ethical questions fade away. Additionally, ABM provides the researcher with opportunities to repeat experiments and recreate interesting dynamics in the simulation. In fact, Miller and Page list a number of reasons for utilizing ABM, most of which apply to this case. In particular they list: the dynamism of the model compared with static alternatives; scalability, in this case to explore the effects of mHealth on a village, a country, or a continent; repeatable and recoverable, as mentioned above; constructive, in that ABM requires one to grow a population from the ground up, thus providing some proof of the relationship between micro and macro phenomena; and low cost, as mentioned above. As they describe it, agent based models "are a fertile playground from which theories can be created, refined, and tested [28]."

Given these considerations, using ABM to model mHealth effectiveness was a natural fit. In order to conduct this study, the modeling software NetLogo was used [29]. NetLogo is a free program based on a relatively simple coding language that offers a user-friendly visual interface [30].

IV. Model Development

As mentioned above, the model contains several types of agents. A population of "Mothers" forms the majority of agents in the model. Eleven percent of the starting population is pregnant at varying stages (between 0 and 280 gestation days). This eleven percent was determined from 2010 estimates of children born per year and number of women within reproductive age in Rwanda [31]. A variable number of birthing attendants can be entered into the model, some or all of whom may have access to mHealth technology. Additionally, the model is seeded with a number of health centers where it is assumed women would have access to trained medical staff.

The model algorithm requires Mothers to wander around the geographic space, representative of a village or a district. The model updates day by day, called a "tick." As days pass, women progress through gestation, up to the average 280 days of human gestation [32]. If they are not currently pregnant, agents are either on break between pregnancies or are attempting to conceive. The break between pregnancies, though not applicable to all women in the real world, is set in this model at six months. This is based on the assumption that most women do not conceive while breast feeding during the first six months after their child's birth, and recent studies that show that women are less likely to encounter complications during birth, including maternal and infant death, if they wait

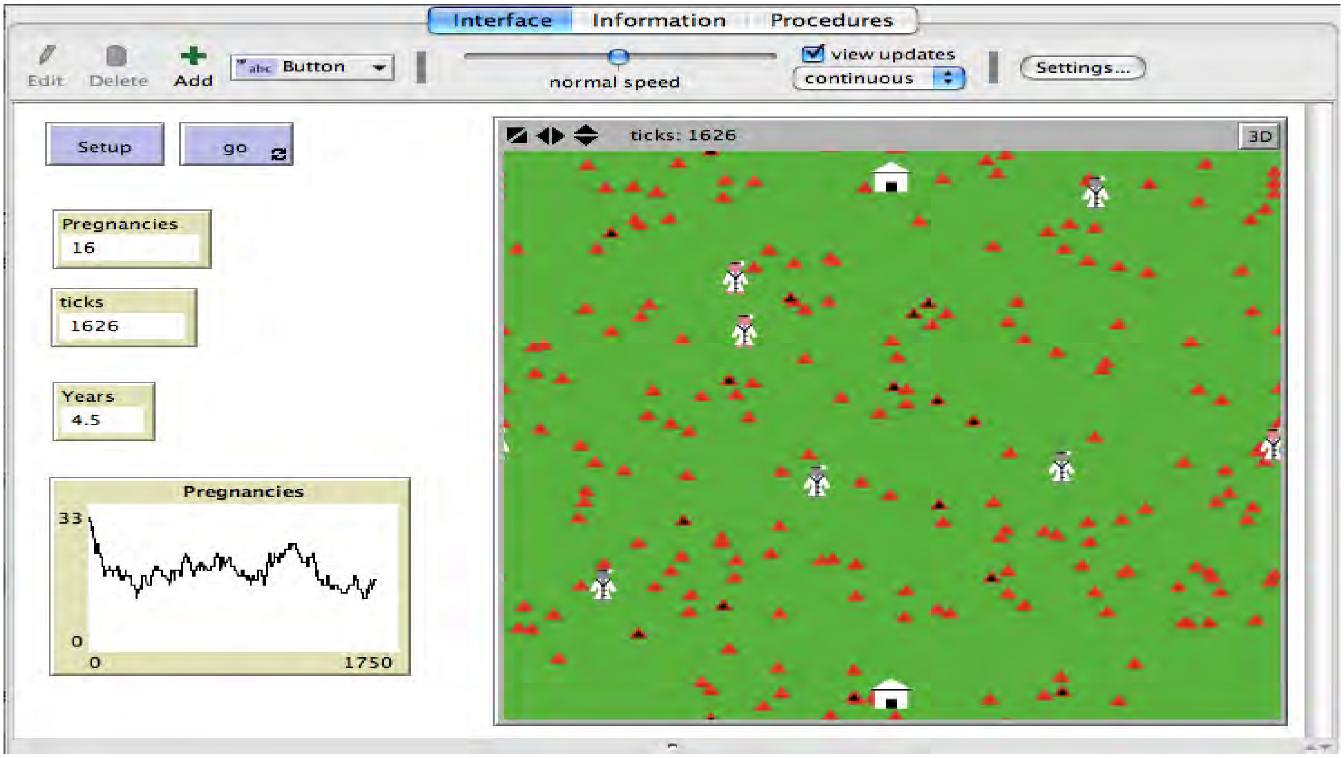


Figure 1: NetLogo Interface for Rwanda mHealth Model

at least six months to conceive the next child [33]. Though this practice may not hold for women in Rwanda, and certainly is not practiced by *all* women *everywhere*, this 'break' was incorporated into the model in order to simplify the maternal mortality probabilities for this particular simulation. It should be noted that with the spread of mHealth, this practice of spacing children to avoid complications would be a good example of a health text message the government could issue to local community health workers stationed throughout the country that are equipped with mobile phones. Mothers in the model are visually depicted in the graphical interface (see Figure 1: NetLogo Interface for Rwanda mHealth Model above) as red triangles; red triangles with black centers represent currently pregnant Mothers.

As mentioned in the previous section, there are two additional types of agents in the model: clinics and birthing assistants. As seen in Figure 1: NetLogo Interface for Rwanda mHealth Model, these agents are visually depicted as white housing structures and persons wearing white doctor's coats respectively. Pink persons wearing white doctor's coats are those equipped with mobile phones, grey are not. As Mothers travel the region and progress through gestation, they may encounter either clinics or birthing attendants. If they are near a clinic at the time of birth, their chance of maternal mortality reflects that in a developed country, in this case Norway's probability of maternal mortality of 1 in 7,700 was used as the benchmark [34]. If they encounter a birthing attendant who has access to mHealth technology, then the maternal mortality rate is halved compared with the World Health Organization's 2009 estimates for Rwanda of 540 deaths in 1,000 live births [35]. If no one is around, or the proximate birthing attendant does not have access to mHealth, the probability of maternal mortality will be 540 in 1,000 live births.

This initial prototype model does not require complex decision criteria or procedures because it reflects only the issue of access to health providers with mHealth. The value of mHealth is not the main question, rather how many mHealth providers are necessary to reduce maternal mortality in a given community in order to reach the MDG 5.

V. Analysis

Forthcoming based on anticipated field data collected through Memorandum of Understanding between Old Dominion University and the Rwanda Ministry of Health.

VI. Validation

A. *Verification*

During the initial phases of this study, a "metacode" version of the model was developed and extensively reviewed. Metacode serves as a logical road map for the actual NetLogo programming code of the model. The actual code is only a broad view of the dynamics of the Rwandan population and thus is relatively simple. With increased data supplied through the Rwanda Ministry of Health, this model could be more accurately verified. This metacode can be viewed in the Appendix of this paper.

B. *Sensitivity Analysis*

Sensitivity analysis for this model was not pursued due to computational cost and time constraints. It should be noted that the most obvious source of sensitivity in this mHealth model is that of the probabilities of sexual intercourse and conception built into the model. In order to achieve the desired 11% of pregnant mothers on any given day, as justified earlier, the rates of sexual intercourse and subsequent conception were determined somewhat arbitrarily. Since these data were not readily available, it was determined that probability of conception would be based on known birth rates [36]; probability of sexual intercourse was set at 2% to balance the model approximately at 11% pregnant Mothers on any given day. This value is somewhat arbitrary and will be updated with actual data if it becomes available. Additionally, because of computational cost constraints, the actual ratio of skilled birthing attendants to expectant mothers could not be explored. Exploring the true ratio of one midwife to 29,000 people in the population is a goal for future studies. Because the true ratio could not be explored, it is unknown if extensively large population changes will be more or less sensitive to other variables such as number of clinics or birthing attendants available.

C. Replication

As noted above, the mHealth model in this study was seeded with the most recent data available on Rwanda. Though some arbitrary measures were incorporated for lack of data, the use of genuine census values to seed the starting model and probabilities of agents' decision criteria lends itself to replication by future researchers. This model is seeded with the most recent data, but a future test might recode the model to begin with data at the start of the implementation of MDG 5 in 1990 in order to explore how well the model predicts the data seen today. This would increase the study's validity by reproducing trends that have actually occurred from a historical perspective.

D. Empirical Calibration

The fact that actual data on Rwanda was used to seed this model makes the empirical calibration strong. There is, however, room for improvement. Field research conducted on how many women at any given time are attended by a birthing attendant with mobile phone connection to a mHealth system versus those without would greatly enhance this study. Additionally, it would be helpful to know how many women actually choose to use mHealth equipped community health workers to aid in childbirth when they are available. The Musanze District government website notes that in 2008, the year the implementation of the pilot program described above began, only 15% of women went in for antenatal checkups, up from 10% in previous years [37]. The model, as currently written, does not account for women's agency in seeking out antenatal health care despite mandates from the government. Future studies might include a measure of how many antenatal visits each woman goes to and the relationship between the attendance at these visits and the outcome of the pregnancy. In future versions of this model, this researcher would like to account for fertility based on better data on reproductive age ranges and availability of choices to limit fertility, such as birth control and information about pregnancy spacing. As written, this model groups women into a large age range, 15-64 years of age, as this was the only data available. Since the life expectancy of Rwandan women is only 58.91 years as of 2010 [36], this range is obviously insufficient for accounting for changing fertility rates during different stages of women's lives. Knowing that complications during childbirth are correlated with diseases such as malaria and HIV/AIDS, future models might also use data that separates direct and indirect causes of maternal mortality [38] as this data becomes available. Of note, none of the factors in this study were chosen simply because data were available. The changes for future models suggested above are possible extensions of the current pilot model.

VII. Conclusion

This study provided only a preliminary examination of the effectiveness of mHealth in reducing maternal mortality rates. Most notably, the model does not currently accommodate variable inputs that would allow for experimentation on the effects of various government policies regarding reproductive health. Rather than aim to provide a conclusive analysis on the effects of government policies, this model served as a prototype to illustrate to the Rwanda Ministry of Health the value of Modeling and Simulation as a methodology. If the Rwanda Ministry of Health in fact chooses to enter into an agreement with Old Dominion University in the future, and provides the applicable data, this researcher will be better equipped to assess the likely parameters that affect the real-world system. Without this closely guarded empirical data, the parameters that matter are anyone's best guess.

Based on discussions between this researcher, Dr. Jennifer Fish of Old Dominion University Women's Studies Department, and the Director of Maternal and Child Health at the Rwanda Ministry of Health during a visit to Kigali, Rwanda in March 2011, the prospects of moving toward a jointly authored study based on the actual mHealth data looks very promising. This Memorandum of Understanding between the Rwanda Ministry of Health and Old Dominion University would serve not only to provide other developing countries with an analysis on the effectiveness of mHealth initiatives to reach MDG 5, but would also allow Ministry of Health officials to experience the value of Modeling and Simulation as a tool for analyzing available data and experimenting with future policy options.

In future versions of this model, this researcher hopes to incorporate a cost benefit analysis for those states interested in pursuing a similar public health initiative. In particular, this researcher would like to focus on the feasibility of instituting mHealth initiatives in vulnerable, displaced populations such as refugee camps along the Rwanda/Democratic Republic of Congo and Rwanda/Tanzania borders. This would require additional research as ownership of refugee public health provision is not necessarily undertaken by the host nation. Future versions of this study will then focus on the cost effectiveness of implementing this program in addressing health and or security concerns of forcibly displaced populations in remote geographic regions.

APPENDIX

Metacode for this model:

Establish three types of agents:
  Women of reproductive age (called "Mothers")
  Birth assistant (called "BAs")
  Clinic (full medical treatment available)

To set-up
  [
  Create a population of Mothers
    Set some to be pregnant at various stages, measured by days up to 280 days – average human gestation in days
    Set some to be on post pregnancy break, measured by days up to 180 – average time from birth until restoral of ability to conceive
    All other Mothers are fertile and pregnable
  Create a population of BAs
    Scatter in stationary positions throughout the community
    BAs should track a variable (Y/N) of whether they are tied in to the mHealth network via mobile phone
  Create a population of Clinics
    Scatter in stationary positions throughout community in realistic proportions based on Ministry of Health data; model generally based off population vs. clinic availability in the pilot district of Musanze
  ] end

To go
  [
  The Mothers wander around the world, representative of a district
  Get-pregnant
  Time moves by day (called "tick")
  If ticks mod 365 = 0 then [restore-population]
  Model automatically stops after 50 years
  ] end

To get-pregnant
  [
  If gestation days are greater than 280
    have-baby
      If successful pregnancy, begin tracking the infertile break period between children (called "break")
  If count of gestation days = 0 (i.e. not pregnant)
    And if break = 0 (i.e. not in between babies)
      Try for conceive
    If 0 < break <= 180
      Move the count of days on break + 1 – remain in the infertile period between pregnancies
    If break > 180
      Set break = 0 – next time around, this Mother can attempt to get pregnant
  ] end

To have-baby
  [
  If a medical facility is within a given radius, set maternal mortality rate to that of a developed country with good health infrastructure – say Norway = Lifetime risk of maternal death is 7,700 [34]

  If a BA with a mobile phone is within a given radius, set maternal mortality rate to half the current mortality rate

  For all other cases, mortality is that of the current mortality rate= 540 per 100,000 births [2]
  ] end

To conceive
  [
  ;; This is probabilistic based on Rwandan birth rates and approximation of chance of intercourse

  Determine randomly the probability of intercourse
    If intercourse takes place, determine stochastically if it results in pregnancy
      If there is a pregnancy, change pregnancy indicator variable (carried by each Mothers), change shape, set gestation counter to 0 days.
  ] end

To restore-population
  [
  Each year, restore women to the population at the country's 3% growth rate
  ] end

ACKNOWLEDGMENT

E. Frydenlund thanks Dr. David Earnest for his support developing a model for maternal health initiatives and the idea to explore mHealth in Rwanda. She also thanks Dr. Jennifer Fish for the opportunity to research the progress of this mHealth initiative in Rwanda.

REFERENCES

[1] United Nations. 22 September 2010. "Goal 5: Improve Maternal Health- Factsheet". *United Nations Summit*. High-level Plenary Meeting of the General Assembly. New York. [Online]. Available: http://www.un.org/millenniumgoals/pdf/MDG_FS_5_EN_new.pdf
[2] World Bank Group. 2010. "Millennium Development Goals." World Development Indicators database [Online] Available: http://ddp-ext.worldbank.org/ext/ddpreports/ViewSharedReport?&CF=&REPORT_ID=1305&REQUEST_TYPE=VIEWADVANCED&DIMENSIONS=177
[3] World Health Organization. 2011. "Making pregnancy safer: Maternal mortality." WHO. [Online]. Available: http://www.who.int/making_pregnancy_safer/topics/maternal_mortality/en/index.html
[4] World Health Organization. 2007. "Maternal Mortality in 2005: Estimates developed by WHO, UNICEF< UNFPA, and The World

Bank." WHO Press. Geneva, Switzerland. [Online]. http://www.who.int/whosis/mme_2005.pdf
[5] International Monetary Fund (IMF). 2006. "Rwanda poverty reduction strategy paper annual progress report- Joint Staff advisory note." IMF Country Report 06/246. Washington, D.C. [Print].
[6] Great Britain Government. 2010. "DFID in 2009-10: response to the International Development (Reporting and Transparency) Act 2006 : presented to Parliament pursuant to Section 1 of the International Development (Reporting and Transparency) Act 2006." Stationery Office. London. [Print]
[7] United Nations. 2007. "MDG Monitor: Rwanda." United Nations Development Program. [Online]. Available: http://www.mdgmonitor.org/country_progress.cfm?c=RWA&cd=646
[8] Greenfeld/Kibungo, Karl Taro. Sunday April 9,2000. "Rwandan Sorrow." *Time*. [Online]. Available: http://www.time.com/time/magazine/article/0,9171,42679-2,00.html
[9] Kigali Institute of Health. "Department of Midwifery." Kigali Institute of Health. Kigali, Rwanda. [Online]. Available: http://www.khi.ac.rw/midwifery.html
[10] Nelson, Libby. March 28, 2011. "In Rwanda, a Professor Works to Save Mothers from Death During Childbirth." *The Chronicle of Higher Education* [Online]. Available: http://chronicle.com/article/Professor-Works-to-Reduce-M/64856/
[11] Twahirwa, Aimable. September 30, 2010. "Retraining Rwanda's Traditional Birth Attendants." *Inter Press Service (IPS)*. [Online]. Available: http://ipsnews.net/news.asp?idnews=53018
[12] Bulterys, Marc, Mary Glenn Fowler, Nathan Shaffer, Pius M. Tih, Alan E. Greenberg, Etienne Karita, Hoosen Coovadia and Kevin M. De Cock. January 26, 2002. "Role of Traditional Birth Attendants in Preventing Perinatal Transmission of HIV." BMJ Publishing Group. BMJ: British Medical Journal, Vol. 324, No. 7331 pp. 222-224. [Print]
[13] Gathoni, Peninnah. 10 October 2009. "Rwanda: Nurses and Midwives Board Inaugurated." *allAfrica Global Media*. [Online]. Available: http://allafrica.com/stories/200910120153.html
[14] Lievens, Tomas. 2010. *Diversity in career preferences of future health workers in Rwanda: where, why, and for how much?* Washington, D.C.: World Bank. [Print].
[15] Republic of Rwanda Ministry of Finance and Economic Planning. *Rwanda Vision 2020*. Available: http://www.minecofin.gov.rw/squelettes-dist/vision2020.html
[16] mHealth Alliance website. 2010. "Frequently Asked Questions." [Online]. Available: http://www.mhealthalliance.org/about/frequently-asked-questions
[17] Iakovidis, I., P. Wilson, and Jean Claude Healy. 2004. *E-health current situation and examples of implemented and beneficial e-health applications.* Amsterdam: IOS Press.[Print]. p 3.
[18] Dwivedi, A.N. , Rajeev K. Bali, Raouf N.G. Naguib, and Nahy S. Nassar. "The efficacy of the m-health paradigm: Incorporating technological, organizational, and managerial perspectives" in Istepanian, Robert S. H., Swamy Laxminarayan, and Constantinos S. Pattichis. 2006. *M-health emerging mobile health systems*. New York: Springer. [Print]. pp20, 26.
[19] United Nations Foundation, Vodafone Foundation, and mHealth Alliance. n.d. "Health Information as Health Care: The Role of Technology in Unlocking Data and Wellness - Factsheet." [Online]. Available: http://www.unfoundation.org/press-center/publications/infoascare.html
[20] mHealth Alliance. n.d. "Maternal mHealth Initiative – MMI Factsheet." [Online]. Available: http://www.mhealthalliance.org/ourwork/maternal_and_newborn
[21] Rasmussen, Mariko. October 12, 2010. "Celebrate Solutions: Training and Mobile Health Technology in Rwanda." Women Deliver website. [Online]. Available: http://www.womendeliver.org/updates/entry/celebrate-solutions-training-and-mobile-health-technology-in-rwanda/
[22] Holmes, Dave. 18 September 2010. "Rwanda: an injection of hope." *The Lancet.* Volume 376, Issue 9745, pp 945-946. [Print]. doi:10.1016/S0140-6736(10)61436-5
[23] Gerber, Ticia; Veronica Olazabal; Karl Brown; Ariel Pablos-Mendez. February 2010. "An Agenda for Action on Global E-Health." *Health Affairs, 29, 2.* p 223. [Print].
[24] "mUbuzima." Rwanda Ministry of Health website. [Online]. Available: http://mubuzima.gov.rw/mUbuzima/core/modules/pagelayout/web/showpage.aspx?menukey=1
[25] Kimani, Mary. April 2008. "Better health at the click of a button: Information technology helps Rwandan clinics reach out." *Africa Renewal*. Vol. 22 # 1 [Print]. p 5.
[26] Mugume, Nathan. 23 July 2010. "Rwanda: Mobile Health Conquers Maternal Death." *AllAfrica Global Media.* [Online]. Available: http://allafrica.com/stories/201007230306.html
[27] Epstein, Joshua M. 2006. *Generative Social Science: Studies in Agent-Based Computational Modeling.* New Jersey: Princeton University Press. [Print]. pp 4-8.
[28] Miller, John H and Scott E. Page. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. New Jersey: Princeton University Press. [Print] pp 78-89.
[29] NetLogo homepage. http://ccl.northwestern.edu/netlogo/
[30] Gilbert, Nigel. 2008. *Agent Based Models.* Thousand Oaks, CA: Sage Publications, Inc. [Print]. pp 48-49
[31] U.S. Central Intelligence Agency. January 20, 2011. "The World Factbook: Rwanda." [Online]. Available: https://www.cia.gov/library/publications/the-world-factbook/geos/rw.html
[32] Lipsky, Martin S. 2006. *American Medical Association concise medical encyclopedia.* New York: Random House Reference. [Print]. p 306.
[33] Gordon, Serena. December 16, 2009. "For Babies' Sake, Pregnancy Spacing matters: study finds 6- to 11-month interval lessens complication risks." Bloomberg Business Week. [Online]. Available: http://www.bsuinessweek.com/lifestyle/content/healthday/633948.html
[34] UNICEF website. "At a glance: Norway." [Online]. Available: http://www.unicef.org/infobycountry/norway_statistics.html
[35] World Bank Group. 2010. "Millennium Development Goals." World Development Indicators database [Online] Available: http://ddp-ext.worldbank.org/ext/ddpreports/ViewSharedReport?&CF=&REPORT_ID=1305&REQUEST_TYPE=VIEWADVANCED&DIMENSIONS=177
[36] U.S. Central Intelligence Agency. January 20, 2011. "The World Factbook: Rwanda." [Online]. Available: https://www.cia.gov/library/publications/the-world-factbook/geos/rw.html
[37] Musanze District website. n.d. "Health care." [Online]. Available: http://www.usanze.gov.rw
[38] Cross, Suzanne, Jacqueline S. Bell, and Wendy J. Graham. February 2010. "What you count is what you target: the implications of maternal death classification for tracking progress towards reducing maternal mortality in developing countries." *Bulletin of the World Health Organization* 88, no. 2. [Print]. pp 147-153.

Immersion / Virtual Reality

Track Lead: Dr. Yuzhong Shen, VMASC

Judges: Dr. Hongwei Zhu, Department of Information Technology and Decision Sciences, Mr. Hector Garcia, VMASC

Procedural Modeling and Animation of Faces
Author(s): Rifat Aras and Yuzhong Shen
1st Place Gene Newman Award Winner, Immersion/Virtual Reality Track

Level of Presence in Team-Building Activities: Gaming Component in Virtual Environments
Author(s): Gianluca Deleo, Koren Goodman, Elena Radici, Scott Sechrist and Thomas Mastaglio
2nd Place Gene Newman Award Winner, Immersion/Virtual Reality Track

Procedural Modeling and Animation of Faces

Rifat Aras and Yuzhong Shen

Abstract – Human brains have specific regions devoted to recognizing faces. So it is not surprising to claim that we distinguish people by their faces. Facial modeling and animation has been studied in computer graphics literature for about half a century and we see state of the art results in film industry. However, current approaches require offline and labor intensive procedures that cannot be applied to real-time systems such as virtual reality applications and video games. In this work, we discuss an automatic procedural modeling framework for faces. Individual facial features and expressions are represented as vertex displacement maps for real-time GPU based processing. Although, randomized faces can be generated instantly, we also provide an intuitive user interface to blend various facial features to obtain a customized face model.

Index Terms – procedural modeling, facial features, facial animation, vertex displacement map

I. INTRODUCTION

We distinguish people by their faces. Containing distinct features like nose, ears, eyes, mouth, lips, chin, and hair, and being the main sensory hub, we communicate through our faces using senses like olfaction, taste, hearing, and vision. Naturally, we experience this important functionality of faces in computational domains where we interact with virtual avatars representing humans.

The first computer generated facial animation dates back to 1972 [1]. Since that date, numerous studies have been performed and researchers have achieved significant advancements on the topic. The results can be best observed through the recent Computer Graphics (CG) assisted motion pictures.

Although CG-assisted feature movies present very realistic state of the art facial modeling and animation, it should be noted that this process is a labor intensive and offline process, which is not suitable for general purpose virtual reality applications and games, where the underlying engine needs to support many virtual characters at real-time operation. As the number of virtual entities increase, it becomes practically impossible to model each character individually. A common alternative is the repetitive usage (cloning) of characters with the same or similar appearances, which in turn results in the decreased immersion of the user.

The solution to this problem lies in the procedural modeling of virtual characters. Procedural modeling is the creation of the assets on-the-fly according to some predefined rules. Landscapes, plants, game levels, or an entire city with buildings can be generated using procedural modeling techniques (Figure 1).

Figure 1: Manually creating a 3D model like this tree is a huge burden on modelers. Procedural modeling techniques can create such models with programmed rules [2].

To satisfy the real-time operation requirement, procedural modeling parameters and animation descriptors have to be stored and processed in an efficient way. In this work, we are utilizing a content creation and consumption pipeline that encodes the modeling parameters and animation descriptors as vertex displacement maps. Such an approach has two apparent advantages. First, we only need to store a single displacement map for each animated expression regardless of the underlying model's resolution rather than storing the entire "key mesh". This property improves the storage efficiency especially with high resolution models. Also, this approach makes it possible to harness the computational power of graphics processing units (GPUs) to process vertex displacement maps and apply the corresponding deformations on the face mesh instead of using the precious CPU cycles that can be used for other sequential tasks.

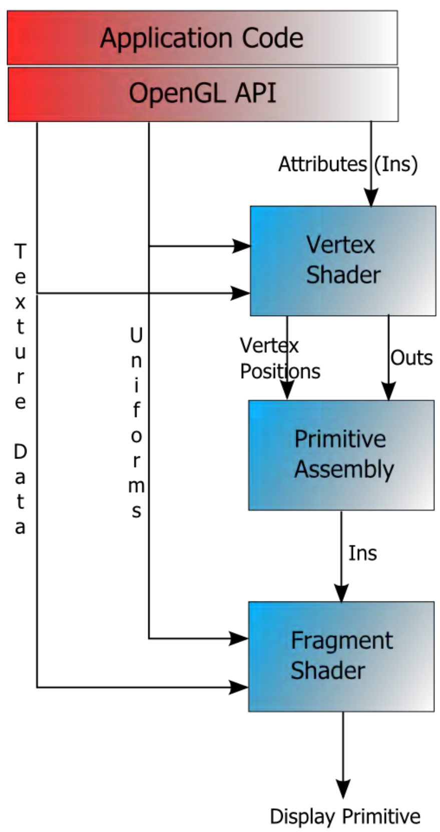


Figure 2: OpenGL's basic graphics pipeline.

Since 2003, GPU designs adopted many-core strategy that try to maximize the floating point calculation throughput by devoting more processor area to data processing units (typically in the order of 120-240 cores) instead of flow control logic elements and large data caches [3]. In this work, we used OpenGL as the underlying graphics library, which can be defined as a software interface to graphics hardware [4]. In a typical graphics library like OpenGL, objects on the screen are formed of primitives that are simply collection of vertices arranged together in a predefined way. The primitives are processed by the graphics pipeline (Figure 2) and displayed on the screen.

The basic graphics pipeline is managed in client-server architecture. Client side resembles the application code and the graphics driver that resides on CPU space. On the other hand, server side lies on the hardware and memory of the graphics card [4].

Shaders are program pieces that run on graphics hardware and form the main processing steps of the graphics pipeline. Vertex shader is responsible for processing the incoming client data like applying transformations, lighting, calculating displacements, and color values. The output of the vertex shader is input to the primitive assembly stage that, as its name implies, put the vertices together to form primitives. Finally, the rasterized primitives are filled in by the fragment shader, which outputs the final color value that is seen on the display device [4].

The remainder of the paper is organized as follows: Section II provides a brief literature review of the previous work. Section III discusses the proposed content creation workflow and the individual steps. Finally, Section IV concludes the paper and discusses the future work.

II. PREVIOUS WORK

Modeling and animation of faces is a well studied field in computer graphics world. The work of Blanz and Vetter [5] is considered to be the pioneer of the parametric face modeling based on a database of 3D face models. In this work, a linear combination of 3D prototype faces is formed to create new faces and expressions. This framework lets users to control complex facial features like gender, attractiveness, and body weight that are learned from the database of faces. Kähler et al. [6] present a framework to deform a canonical head model to fit 3D scan data and animate the model by employing pseudo facial muscles and elastic skin properties. This work also supports simulation of growth of a human head by utilizing anthropometric measurements on the model that results in animatable head models of an individual at different ages. One drawback of this work is the lack of high level control over facial animation. Facial motion is controlled by specifying muscle contraction parameters over time.

When an application employs large number of virtual avatars, it will be inevitable to use same models for the characters due to the huge effort needed to model them. However, for virtual reality applications and games, this approach would result in the decreased realism of the application. In the work of McDonnell et al. [7] a series of experiments are conducted to test user's perception of variety in such applications. The experiments showed that even smallest variations can increase the immersive nature of the simulation. Maïm et al. [8] attack the problem by proposing a human shape variation system by attaching accessories to the models. In addition to this, the model textures are also modified in terms of color and pattern to increase the effect.

III. CONTENT CREATION PIPELINE

In the proposed content creation pipeline, we aimed at using in-house developed scripts, applications and readily available third party tools to define the facial features, encode the features and emotion descriptors as vertex displacement maps, provide an intuitive user interface to blend vertex displacement maps in order to create the desired model, and bake the final set of displacement maps to be consumed by a client application.

A. *Defining Facial Features and Emotion Descriptors*

We used Blender [9], which is a free open source 3D content creation application, to define facial features and emotion descriptors. The reason behind choosing an open source alternative is to keep the access to the tool relatively easy and to promote the community to add further features and descriptors to the readily available database.

A canonical mesh, M, of a human bust is used as the base model from which deformations of facial features are obtained. This mesh is basically a list of vertices and triangles formed by these vertices along with their uv-2D texture coordinates (Figure 3). By placing Blender's lattice based free form deformation (FFD) tool around specific facial features, those features are deformed and the resulting deformed mesh, M', is saved as a new mesh (Figure 4).

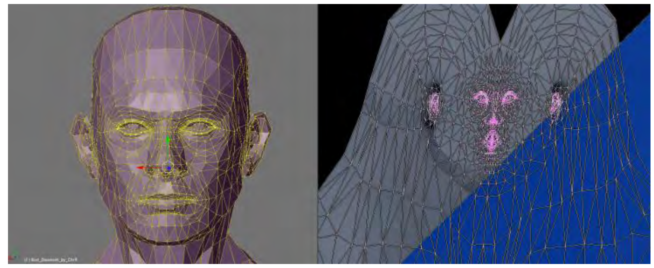
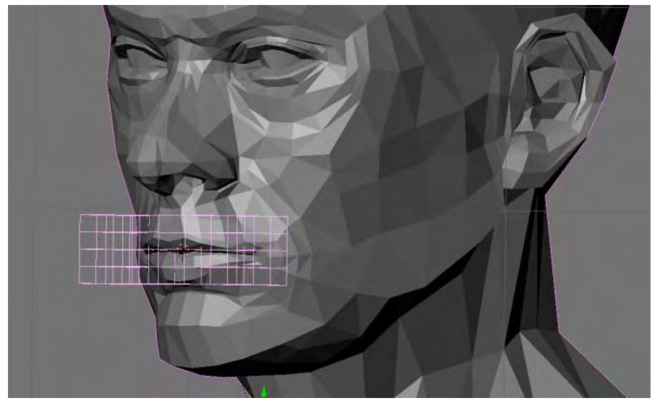
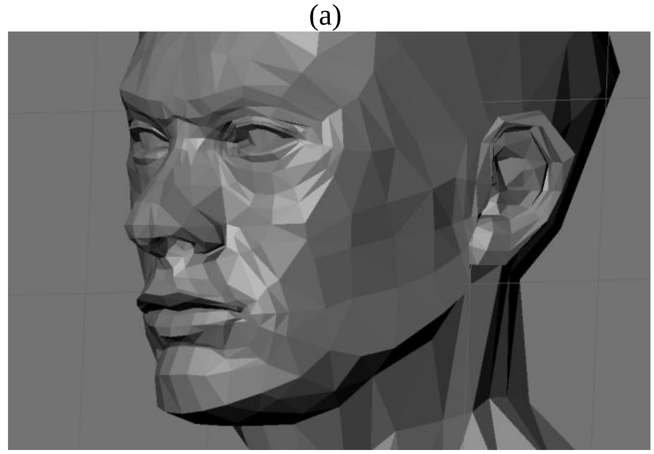


Figure 3: The canonical head model and the corresponding texture UV map.



(a)



(b)

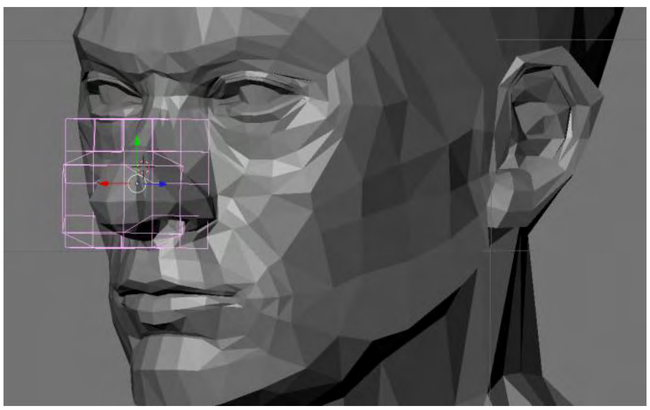
Figure 4: (a) The free form deformation lattice is placed to mouth region to define "lip thickness" facial feature. (b) The result of the deformation.

B. *Extracting Vertex Displacement Map*

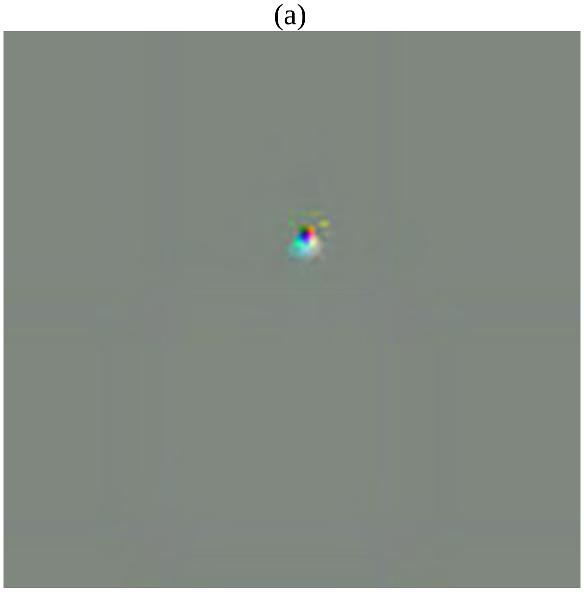
After deforming M with free form deformation techniques to obtain M', the two meshes are fed to a script to compute the spatial difference and encode the difference as a vertex displacement map (Figure 5), which will be used and processed by the vertex shader phase of the 3D engine while rendering the facial model.

Algorithm 1: The extraction of vertex displacement map from the original model and the displaced model.

```
function ExtractVDM(M,M'):
// K is resolution of final texture map
  mapSize := K;
// 3-channel texture
  vdm = zeros[K][K][3];
  for each vertex v in M:
   vdm[v.texV*K][ v.texU*K]=M'.v - M.v;
// Texture UV coordinates vary from
// 0 to 1. So, we need to convert them
// to integer array indices.
  end for;
  return vdm;
end ExtractVDM.
```



(a)



(b)

Figure 5: (a) The nose of the model is made larger with free form deformation. (b) The resulting vertex displacement map of this facial feature.

OpenGL is capable of storing the texture information in various precision and format. The color component can have 1, 2, 3, or 4 elements corresponding to red, green, blue, and alpha channels. Furthermore, these components can be stored internally as different formats such as integers, floating point numbers, and bytes. For the purpose of representing vertex displacements as texture information, we chose to store displacement vectors as 3-component floating point numbers. This approach let us to skip the normalization process, which we would face in case of using 3-component byte representation, in the expense of using more storage space.

C. *User Interface for Blending Features and Emotion Descriptors*

The stored facial features and emotions are presented as high-level controls that can be blended together to obtain complex final models. The interface is presented to the user by using GLUI, which is an OpenGL based platform independent user interface library. The high-level feature and emotion descriptors are provided as a list that can be selected and applied to the model on-the-fly using interactive blending weights.

After the user obtains the desired face model, the blended facial features are baked into a final vertex displacement map that is to be consumed later on by any application implementing the specialized vertex shader steps.

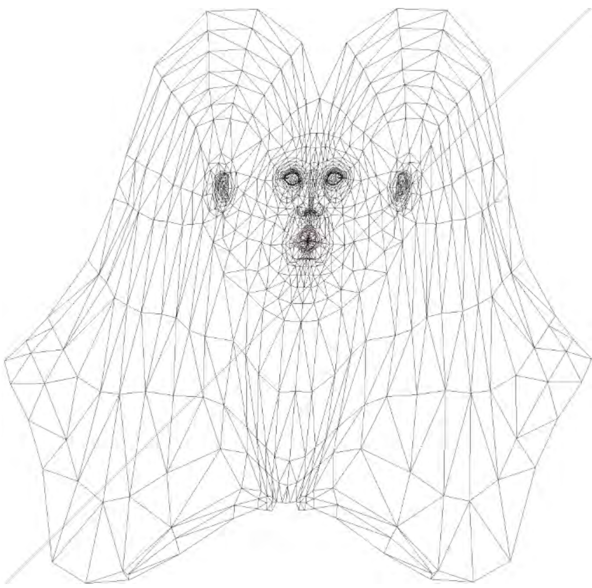
D. *Rendering*

Since the introduction of programmable graphics pipeline, lots of custom vertex and fragment shaders have been implemented to create sophisticated onscreen effects rather than using the flat or Gouraud shading that the graphics libraries provide by default. In this work, we employ Phong illumination model that computes the final color information per pixel by interpolating surface normals between vertices. Phong illumination model involves ambient, diffuse, and specular light and material components and is defined by:

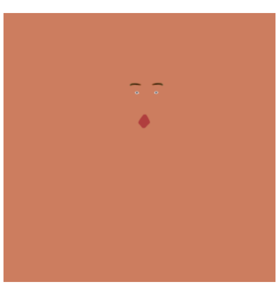
$$I = k_a i_a + k_d (L.N) i_d + k_s (R.V)^{\alpha} i_s \qquad (1)$$

Where k_s, k_d and k_a are specular, diffuse, and ambient reflection constants of the skin respectively. Likewise, i_s , i_d and i_a are specular, diffuse, and ambient light intensities. α is defines as the shininess constant for the skin. L is the direction vector from the point on surface to the light source, N is the normal of the surface at this point, R is the perfect reflection vector, and V is the direction vector towards the viewer. In our model, material's diffuse color is obtained from the diffuse texture that was manually created for the canonical mesh from reference real-life human face images.

In order to increase the visual realism of the model, ambient occlusion map, which is the pre-computation of the attenuation of the light due to occlusion, is created and overlaid on top of the diffuse texture. Although, it is only an approximation of the global illumination model, application of ambient occlusion map on the model boost the visual realism at virtually no price (Figure 6).

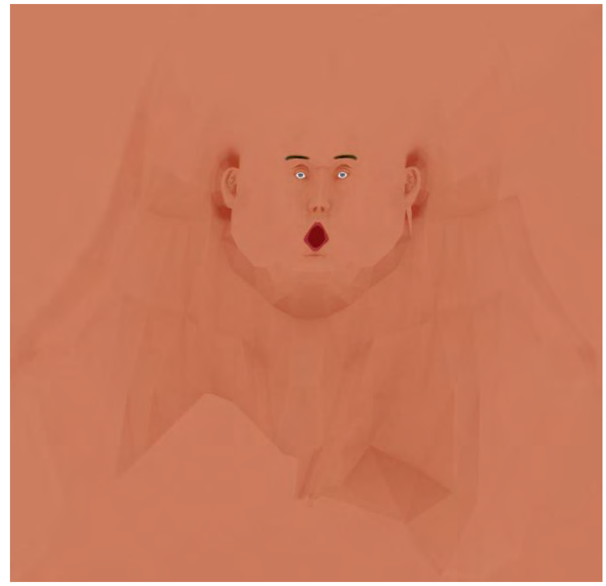


(a)



(b)

(c)

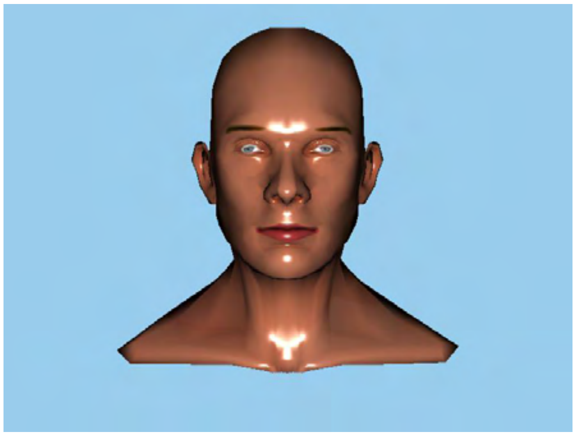


(d)

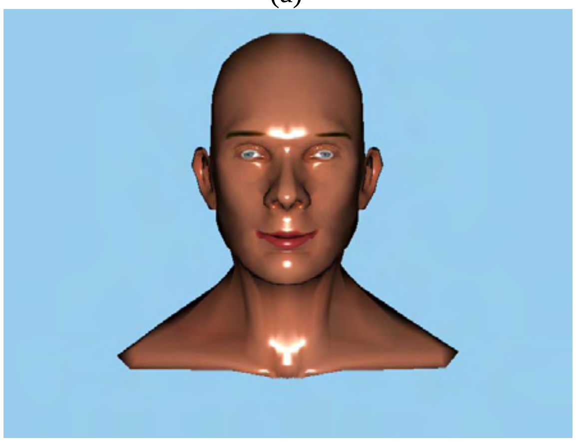
Figure 6: (a) The UV layout of the model displayed in wireframe mode. (b) The diffuse color map of the model. (c) The ambient occlusion map of the model. (d) The resultant diffuse map obtained from overlaying (c) over (b).

IV. CONCLUSION AND FUTURE WORK

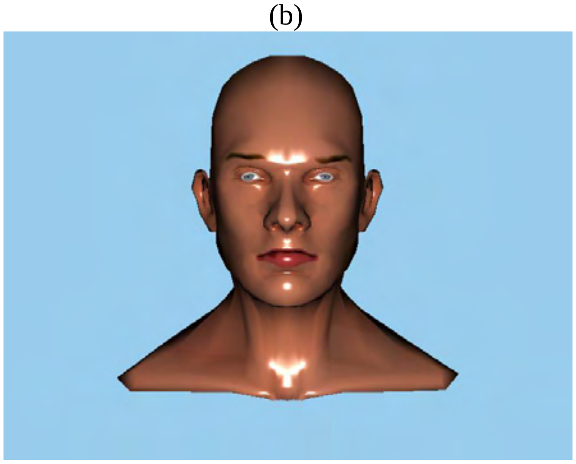
In this work, we demonstrated a procedural modeling pipeline for 3D face models. In order to make realistic crowd simulations, the entities in the simulation have to be as different as possible, but this requirement puts an enormous burden on 3D modelers. Procedural modeling techniques can be used in such circumstances to differentiate virtual characters from each other. In our pipeline, we encoded the facial features and expressions in vertex displacement format (Figure 7), which made it possible not only to process features and expressions in real-time by graphics hardware, but also to store these in a compact way regardless of the 3D mesh resolution. A near-term goal in this work is to extend this technique to whole body modeling and animation rather than limiting it only to facial modeling. Also, we would like to implement an animation infrastructure to utilize this technique and create complex animation sequences.



(a)



(b)



(c)

Figure 7: (a) The neutral expression. (b) "Smiling" expression parameter weight is set to 1. (c) The "angry face" expression parameter weight is set to 1.

REFERENCES

[1] F. I. Parke, "Computer generated animation of faces," presented at the Proceedings of the ACM annual conference - Volume 1, Boston, Massachusetts, United States, 1972.

[2] M. Cepero. (2010, *A Voxel Tree*. Available: http://procworld.blogspot.com

[3] NVIDIA, "NVIDIA CUDA C Programming Guide," 2010.

[4] R. S. Wright, N. Haemel, G. Sellers, and B. Lipchak, *OpenGL SuperBible: Comprehensive Tutorial and Reference*: Addison-Wesley Professional, 2010.

[5] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," presented at the Proceedings of the 26th annual conference on Computer graphics and interactive techniques, 1999.

[6] K. Kähler, J. Haber, H. Yamauchi, and H.-P. Seidel, "Head shop: generating animated head models with anatomical structure," presented at the Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation, San Antonio, Texas, 2002.

[7] R. McDonnell, M. Larkin, S. Dobbyn, S. Collins, and C. O'Sullivan, "Clone attack! Perception of crowd variety," *ACM Trans. Graph.,* vol. 27, pp. 1-8, 2008.

[8] J. Maïm, B. Yersin, and D. Thalmann, "Unique character instances for crowds," *IEEE Comput. Graph. Appl.,* vol. 29, pp. 82-90, 2009.

[9] Blender Foundation. (2009, April). *blender.org - Home* [Online]. Available: http://www.blender.org

Transportation

Track Lead: Dr. Mike Robinson, VMASC

Judges: Dr. Guzin Akan, ODU, Mr. David Boss, Skranska Infrastructure Development

Do Trip Productions Vary in Space? Capturing Unobserved Heterogeneity in Travel Demand Models

Author(s): Xin Wang, Asad Khattak

Evaluating Vehicle Queue Length Estimations at the HRBT Based on Probe Vehicle Data Using Micro-Simulation

Author(s): Craig Jordan and Mecit Cetin

1st Place Gene Newman Award Winner, Transportation Track

Freeway Lane Utilization Behavior

Author(s): Sanghoon Son, Mecit Cetin, Asad Khattak

Model Transferability in the Re-identification of Trucks Based on AVC and WIM Data

Author(s): Ilyas Ustun, Mecit Cetin

1st Place Gene Newman Award Winner, Transportation Track

Vehicle Re-identification using Neural Networks

Author(s): Tanweer Rashid, Mecit Cetin

1st Place Gene Newman Award Winner, Transportation Track

Do Trip Productions Vary in Space?
Capturing Unobserved Heterogeneity in Travel Demand Models

Xin Wang & Asad Khattak

Department of Civil and Environmental Engineering, Old Dominion University, Norfolk, USA

Abstract: The accuracy of travel demand forecasts is critical to properly anticipating future transportation problems that include traffic congestion and pollution. The frequency of trips produced by households is a key first step in the four-step travel demand forecasting process, as practiced in the United States. Traditional trip productions use regression models or multiple classification analysis to calculate how many trips will be produced by each zone, typically using socio-economic variables such as household size, household vehicle count, and household income to forecast trip frequency. One important drawback of traditional trip generation models (referred to as global models) is that they use one set of estimated parameters to forecast trips—an approach characterized as "one size fits all." However, for spatially diverse metropolitan regions, trip frequency and its associations with socio-economic variables can vary across space, causing a problem known as unobserved spatial heterogeneity. To capture spatial variations, this paper uses Geographically Weighted Poisson Regression (GWPR)—a locally-based model as an alternative. A behavioral dataset for Hampton Roads, Virginia collected in 2008-2009 (N=2,889) is used to understand spatial variations in how trip productions relate to socio-economic factors. The results indicate that the local models provide a more accurate estimation and better picture of how trip rates vary across space. The broader implication is to consider moving toward local models in order to improve the accuracy of travel demand forecasts.

Keywords— geographically weighted poison regression, trip generation, global model, local model, spatial variance, travel demand forecasting.

INTRODUCTION

Anticipating future traffic congestion and pollution are critical to a regions growth, livability and economic development. Currently, travel demand forecasting models are mandated to be used for regional demand forecasting in order to anticipate the impacts of various types of regional development and transportation capacity enhancements. For example a recent study in Hampton Roads used the regional travel demand model to predict the impacts of six large roadway enhancement alternatives on traffic congestion in 2030[1]. The accuracy of predictions in such studies is important to properly anticipate and address future problems that include traffic congestion and pollution, and to inform the planning process.

A critical step in the conventional four-step travel demand modeling process is trip generation, which captures the trips produced by households. These trips are typically associated with household level socio-economic variables, e.g., household size, vehicle ownership, and income. In this step, often, there is no consideration of the spatial and geographical nature of trips. However, it has been known that travel behavior is partly based on spatial constraints. For instance, residents living in the same spatial cluster may share similar lifestyles, attitudes and preferences, due to human, social or other environmental reasons. It has been known that “Everything is related to everything else, but near things are more related than distant things” [2]. This shapes some intangible patterns surrounding human activity participation. People/households living in close proximity exhibit similar behaviors in terms of trip decisions and outcomes, e.g., activity/trip frequency, trip durations, and vehicle-miles traveled [3]. Therefore, this paper explicitly explores whether trip rates and associations with socio-economic variables vary across space. This will allow assessing if households are different in terms of their trip productions in ways that are not accounted for in the explanatory variables of the models. More technically, the study explores if there is unobserved heterogeneity due to spatial location of residents, and how to improve the current travel demand models to capture this unobserved heterogeneity.

Unobserved spatial heterogeneity cannot be accounted for directly through standard models, referred to as global regression models, because the estimated parameters are fixed and represent a mean association between dependent and explanatory variables. This approach can be characterized as “one size fits all” modeling. Although unrestricted models can be estimated for various spatial units and compared with a global or pooled model, this is often cumbersome and rarely done—partially because the problem of fixed parameters still exists within the spatial units; moreover, the definition of the boundary of those spatial units will influence the estimation of the parameters, referred to as the (undesirable) “boundary effect”[4]. In this paper, a finer local model—a Geographically Weighted Poisson Regression (GWPR) is estimated to relax the fixed association assumption between travel demand and socio-economic characteristics and therefore capture the spatial heterogeneity and improve the accuracy of travel demand forecasts.

BEHAVIORAL DATA

The behavioral data analyzed in the paper are from the Virginia Add-on survey of NHTS (National Household Travel Survey) conducted in 2008—Figure 1 shows the study region. The survey period was from April 2008 through May 2009. The sampling plan included both geographic and demographic goals to ensure that the survey is representative of the region’s population (approximately 1.7 million) and their activity-travel patterns.

The database contains three different levels of data: personal data, household data, and trip data. The survey relied on the willingness of households to provide: 1) demographic information about the household, its members and its vehicles; 2) have all household members record travel-related details for a specific 24-hour period, including address information for all locations visited, trip purpose, mode, and travel time information.

The study area is located in southeastern Virginia-Hampton Roads Area, including Norfolk, Gloucester, Isle of Wight, Jamestown, Yorktown, Chesapeake, Hampton, Newport News, Poquoson, Portsmouth, Suffolk, Virginia Beach, and Williamsburg. The area studied in this paper is also consistent with the Hampton Roads Transportation Planning Organization (HRTPO) transportation study area. Due to large areas of water-bodies, creating and maintaining adequate infrastructure has long been a major challenge. The Hampton Roads Bridge-Tunnel (HRBT) and the Monitor-Merrimac

Memorial Bridge-Tunnel (MMMBT) are major harbor crossings of the Hampton Roads Beltway which links the largest Seven Cities of Hampton Roads.

After error-checking and geocoding the residential and all the trip destination locations at the TAZ level and removing the cases that did not have geographical information, the final dataset contained 2,889 households, with a total of 7,107 persons who made 24,810 trips on their assigned travel days.

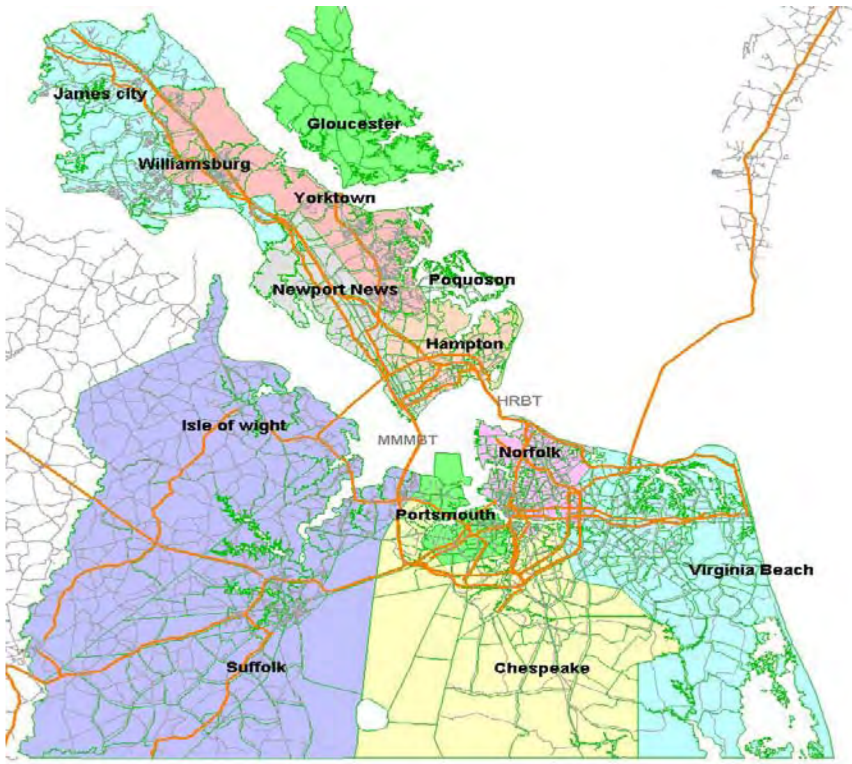


Figure 1: Study region (Hampton Roads Area)

DESCRIPTIVE STATISTICS

The households' descriptive statistics are shown in Table 1. Variables were selected based on the typical variables used in household trip production models. Due to privacy considerations, limited geo-location information was available. As a result, only those households with at least one trip on their assigned travel day are included in the database (i.e., zero-trip households are not included in the database analyzed). Therefore, the analysis of trip productions reflects only households who make at least one trip.

The reported trip frequency at the household level is 8.6, with an average household size of 2.46. Average vehicle ownership is 2.22 vehicles per household. Average annual household income is nearly $69,000, based on coding the middle value of income categories to calculate the mean of household income. Overall, these numbers are reasonable and in-line with national statistics for similar urban areas.

Table 1: Descriptive statistics by cities (N=2889)

NO.	CITY	AV. TRIPS	AV. VEHICLES	AV. HHSIZE	AV. INCOME
0	Gloucester	8.21	2.45	2.34	63.73
1	Isle of Wight	8.02	2.55	2.43	65.50
2	Jamestown	9.27	2.13	2.36	77.42
3	Yorktown	9.29	2.41	2.69	80.93
4	Chesapeake	9.21	2.41	2.65	72.99
5	Hampton	8.07	2.22	2.28	61.94
6	Newport News	8.57	1.96	2.37	59.79
7	Norfolk	7.13	1.86	2.20	57.51
8	Poquoson	7.69	2.62	2.41	79.02
9	Portsmouth	7.99	2.12	2.22	60.94
10	Suffolk	8.02	2.30	2.56	69.02
11	Virginia Beach	9.06	2.25	2.59	74.02
12	Williamsburg	8.41	2.07	2.07	67.41
Pooled	Mean	8.59	2.22	2.46	68.95
	Std. Dev.	5.93	1.12	1.24	37.50
	Min	1.00	0.00	1.00	2.50
	Max	47.00	10.00	10.00	168.60

To explore the relationship between these variables and space, a finer level of descriptive statistics based on county level can show which area has higher/lower levels of certain characteristics. The statistics for different counties indicate substantial diversity in trip frequency as well as socioeconomic characteristics over space. For instance, households in suburban areas like Jamestown and Yorktown make more daily trips compared with areas with more compact land use like the City of Norfolk. This makes sense when households with higher income and more household members prefer to live in compact areas. This partially indicates the existence of unobserved spatial heterogeneity because the variables show variation across the study area. This issue is further explored using the GWPR models.

MODEL STRUCTURE AND RESULTS

Application of Geographically Weighted Regression

GWR is a non-parametric methodology for the investigation of geographical drifts in regression parameters. It aims to explore spatial deviations of associations between dependent and explanatory variables by relaxing the assumption that estimated parameters hold globally. In GWR, the regression model is calibrated based on data that are geographically proximate to a household's location. That is, it is within a certain distance (bandwidth) and these data are weighted from a regression point—the weight calculation is shown later in equation (2). For a given regression point, the weight of an observation is at a maximum when it shares the same location. This weight decreases as the distance between the two points increases. In this way, a regression model is calibrated locally simply by moving the regression point across the entire region[5]. For each location, the data are weighted differently so the results of a calibration are unique to a particular location. By plotting the results of these local calibrations on a map, surfaces of parameter estimates, or any other display which is appropriate, can be generated. Therefore, when GWR is applied, key decisions must be made regarding 1) a weighting function (the shape of the kernel), and 2) the bandwidth of the kernel. The weighting function usually has a minimal effect on results, while bandwidth may affect results markedly [6]. In this study, a random location is assigned to the household based on available TAZ information. Considering trip frequency is a non-negative count variable, Poisson or negative binomial regression are appropriate. Due to limitations of the available software (GWRx3.0), only GWPR was estimated. In its most basic form, the GWPR model is described as follows [5]:

$$\ln\,(y_i) = \beta_{i0} + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \tag{1}$$

y_i = dependent variable (trip frequency) for household i (i = 1, 2, . . . , n, where n is the number of observations);
β_{i0}=the constant at household i;
β_k= the parameter at household i for explantory variable x_{ik};
x_{ik} = explanatory variables of the kth parameter for household i,
ε_i = error term at location i,
p = number of parameters.

Note that for each household location, the β parameter can be different. The model is fitted using a technique known as iteratively reweighted least squares [5].

Adaptive kernel, the bi-square kernel, is used to calculate the weights. Adaptive kernels are useful in the case there is a large variation in the geographical density of the observed data [5]. Then the kernels have larger bandwidths where the data are sparse and have smaller bandwidths where the data are plentiful. The weights are defined by:

$$w_{ij} = \begin{cases} \left[1 - \left(||\mathrm{u_i}-\mathrm{u_j}||\,/G_i\right)^2\right]^2 & if||\mathrm{u_i}-\mathrm{u_j}|| < G_i \\ 0 & otherwise \end{cases} \tag{2}$$

Where the parameter G (called the bandwidth) regulates the kernel size, and $||\mathrm{u_i}-\mathrm{u_j}||$ is the distance between locations i and j. When calibrating the model, the kernel bandwidth is determined by minimization of Akaike Information Corrected Criterion (AIC), which is a measure of goodness of fit and allows finding a model that best explains the data with fewer estimated parameters. The AIC of the model with bandwidth G is given by:

$$\mathrm{AIC(G)} = \mathrm{nln(RSS(G))} + 2\ \mathrm{K(G)} \tag{3}$$

$$\mathrm{RSS(G)}=\sum_{i=1}^{N} \hat{\varepsilon}_i^2 \tag{4}$$

Where RSS(G) is the residual sum of squares with bandwidth G; N is the number of observations; K(G) denotes the effective number of parameters in the model with bandwidth G, respectively. However, since the degrees of freedom for GWPR models are typically small, a small sample bias adjustment in the AIC calculation is done. The Corrected Akaike information Criterion (AICc) is then used to address this bias [5, 7]. AICc is defined as follows:

$$\mathrm{AICc(G)} = \mathrm{AIC(G)} + 2\frac{\mathrm{K(G)(K(G)+1)}}{N-\mathrm{K(G)}-1} \tag{5}$$

AICc can be used to evaluate whether GWPR provides a better fit than a global Poisson model [5-6]; a smaller AICc value reflects a better the model. If the effective number of parameters K is small relative to the number of observations N, then the difference between AIC and AICc is negligible [7].

Modeling Results

In this study, a global (conventional) Poisson regression model is compared with GWPR to examine which model is more appropriate statistically. The OLS (Ordinary Least Squares) global model is also provided for comparison. The model results are presented in Table 2.

Table 2: Global model (Poisson and OLS) for household trip frequency (N=2889)

	Poisson				OLS	
Variable	Coef.	Exp (coef.)	Marginal	T-stat.	Coef	T-stat.
Constant	1.233	3.432	-	66.229	0.592	2.400
HHVEH	0.062	1.064	0.495	10.362	0.467	4.990
HHSIZE	0.215	1.240	1.73	47.426	2.307	28.770
INCOME	0.003	1.003	0.02	14.752	0.019	6.950
Summary	Pseudo R Square =0.165 LR chi2(3) = 3557.640 Prob > chi2 = 0.000				R Square=0.330 F(3, 2885) = 475.410 Prob > F = 0.000	

Both of these two models are statistically significant overall, represent reasonable fit and provide similar estimation (based on marginal effects). As expected the number of trips is statistically significantly (5% level) associated with household income, vehicles owned and household size. The exponentiation of parameters, e^{β}, provides Incident Rate Ratios (IRR). The marginal effects of the coefficients are also provided. This facilitates interpretation of the parameter, i.e., whether it is associated with an increase or decreases in the expected number of trips, when the value of the explanatory variable changes. The ratios show that larger household size, higher vehicle ownership, and higher income are associated with more trips. Each additional person in the household is associated with higher expected number of trips by a factor of .215 in Poisson model, with the marginal effect of 1.73. A simpler interpretation is given by the Ordinary Least Squares regression, which shows that, on average, one more household member is associated with about 2.3 more trips. The global regression model provides associations between trip productions and correlates from an overall perspective, i.e., it represents the average relationship between correlates in the entire study region. More specifically, on average, one more vehicle in the household is associated with additional 0.5 trips (the marginal effects are similar in both Poisson and OLS model). However, this relationship may be not the case across all of Hampton Road, which means that the potential spatial variation of these parameters is unknown. GWR can detect spatial non-stationarity, allowing the parameters to vary across space, and it can therefore uncover possible local spatial deviations of explanatory variables.

The Local Model

A GWPR model was estimated based on the maximum likelihood method. But a distinct difference between global and local estimation is that the global estimation has one set of model parameters for all observations in the sample, while the local model estimates a set of parameters for explanatory variables for each location. Thus a parameter surface can be generated based on estimation which varies continuously in space. Table 3 shows the goodness for fit measures for both global and local models. By comparing AICc between the global and local models, the local model outperforms its counterpart, as the AICc values for local model is substantially lower than the global model. As a general rule, improvements in the AIC that are less than 3 in value could easily arise as a result of sampling error [5], while here the difference between the global and local models is substantially greater than 3, indicating that the local models are statistically better than the global model. Furthermore, MAD (Mean Absolute Deviation) and RMSE (Root Mean Square Error) are calculated to compare the accuracy of the two models. The results show that GWPR model has lower MAD and RMSE, which is better compared with the global Poisson model.

Table 3: Goodness of fit measures for global vs. local model (N=2889)

	Global model	Local model
AICc (Corrected Akaike Info Criterion)	7,017	6,904
MAD (Mean Absolute Deviation)	4.89	3.54
RMSE (Root Mean Square Error)	5.83	4.78

Table 4 shows the parameter summary for GWPR model, providing information about the “width” of a parameter’s spatial variance. Theoretically all parameters vary in space when GWR is used. It is important to determine if the spatial variance is significant enough to be captured by using the more complicated GWR model. It will be reasonable to use the global model when the spatial variation is modest, known as stationarity in space. To decide whether the spatial variance is significant, the difference between the lower quartile and upper quartile of a parameter is compared with the standard error. If the difference is larger than two standard errors, then the parameter is non-stationary in space [5], indicating that spatial variance is statistically significant. From the results, all three explanatory variables show significant spatial variance, among them, the number of vehicles and household size have larger spatial variance.

Table 4: Parameter summary for GWPR model (N=2889)

Variable	Minimum	Lwr Quartile	Median	Upr Quartile	Maximum	Upr-Lwr	2*Std. Errors	(Upr-Lwr)> 2*Std. Errors
HHVEH	0.015	0.050	0.069	0.081	0.150	0.031	0.012	Yes
HHSIZE	0.121	0.206	0.218	0.232	0.276	0.026	0.009	Yes
INCOME	0.001	0.002	0.003	0.003	0.004	0.001	0	Yes

Based on the local parameter estimates for 2,889 households, an Inverse Distance Weighted (IDW) interpolation algorithm is used to assign values to unknown points, thus a continuous parameter surface covering the whole region is created. IDW assumes that each measured point has a local influence that diminishes with distance. It weights the points closer to the prediction location greater than those farther away. As in most travel demand models, the predictor variables, e.g., population and number of vehicles are stored at TAZ level, an average value of the parameter is calculated for each of TAZ, which can to be used to predict travel demand in future years (if desired, parameter estimation at a finer spatial level is also possible with GWR).

Figure 2 shows the variation in parameter for household vehicles owned and household size respectively. In contrast to the unified parameters across space in the global model, both parameters vary substantially across the study area. While the global model shows that the household vehicle count parameter to be 0.062 with IRR of 1.063 ($e^{0.062}$), it varies from 0.04 to 0.15, with IRR between 1.04 and 1.16 across zones in the local model. Also, while the global model shows that the household size parameter is 0.215 with IRR of 1.24, it varies from 0.13 to 0.27, with IRR between 1.14 and 1.31 across zones in the local model. This indicates that the global model may obscure information about spatial variations [5]. The magnitude of household vehicle count is higher in cities like Norfolk and Suffolk; the magnitude of household size, which can be interpreted as personal trip rate, is higher in Chesapeake, Williamsburg, and Yorktown.

Also, the figure shows that simply defining an area as urban, suburban or rural and assigning it a higher or lower trip rate can be arbitrary, as the map does not show substantial variation of parameters with density. Thus, the commonly used cross-classification tables by urban or rural area can be considered archaic when richer data are available for spatial analysis. Practically, by using the parameter map created by the local model, a cross-calculation table can be created for each TAZ, providing a finer forecast of trip productions.

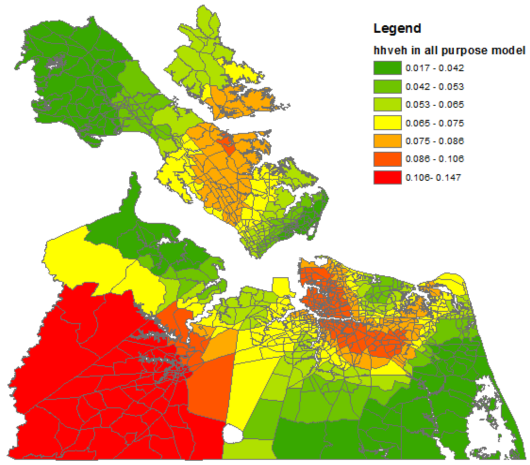


Figure 2-1: Parameter estimates for HHVEH in GWPR model

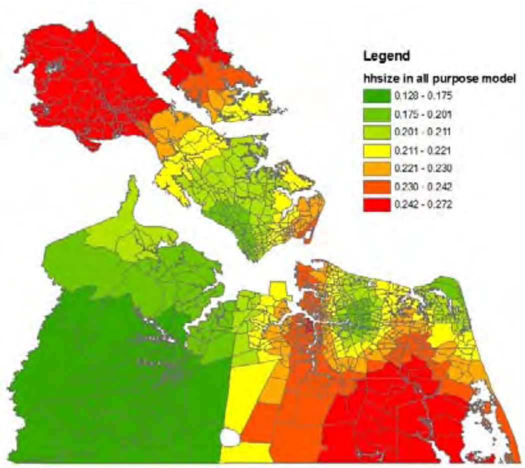


Figure 2-2: Parameter estimates for HHSIZE in GWPR model

The study attempted to estimate global Poisson and GWPR models by purpose (home-based work, other, and non-home-based). However, the GWPR failed to converge due to a substantial number of zeroes in the dependent variable; 60% of households made no home-based work trip and one-half of the sampled household made no HBO trips.

Model Application and Comparison

Comparisons between the global and local models can be obtained by computing the difference in predicted trip frequency using the models calibrated in this study. For the global model, the fixed equation is applied to every TAZ; for the local model, different equations are applied because parameters vary by TAZ. Figure 3 shows the difference between predicted daily trip frequency from these two models, using the 2030 population, vehicle ownership, and households available in Hampton Roads travel demand model. Since the global model is a spatial average of local models, both of these models, when applied for prediction purposes, should predict the same trip totals. Nevertheless, there is small difference (2%) in total trips predicted by the global Poisson model and GWPR model.

A positive value in Figure 3 implies that more trips are predicted using the local model. This indicates that using the global model for prediction can overestimate the travel demand in certain areas, e.g., the city of Norfolk, while the demand in areas such as northern Virginia Beach and areas north of Hampton Roads Bridge Tunnel (HRBT) is underestimated. Relocating the trip productions will change the trip distribution in this region significantly, and it will have a substantial impact on subsequent steps, i.e., trip distribution, mode split, and traffic assignment.

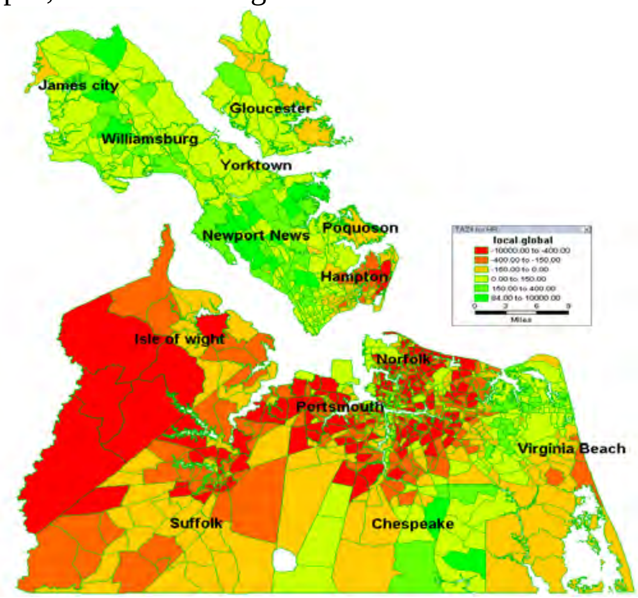


Figure 3: Differences in daily trip productions between global and local models.
(Note: A positive value implies that more trips are predicted by the local model.)

LIMITATIONS

A key issue is whether appropriate model specification was used. The model specifications are consistent with the ones used in the field, e.g., the Hampton Roads travel demand model. However, the study is limited by use of cross-sectional data from a single metropolitan area. The transferability of the findings from the Hampton Roads Area to other urban regions is somewhat limited. That is, the spatial relationships uncovered through GWPR are only valid for this region, due to the fact that the model itself—the weights used in the model are based on spatial locations, which are unique to the area. However, the methodology used in this paper can be transferred to other regions. Successful application of the GWPR method in this context can also help in designing similar efforts in other mid-sized cities that have a relatively strong private auto culture. Furthermore, the Hampton Roads is reflective of metropolitan areas with beltways and planning decisions that have directed growth, making it somewhat of a typical city among moderately-sized metropolitan areas. The data and models developed in this study cannot fully explore and reveal all the forces at work in trip production. Nevertheless, the findings are useful for fleshing out the tools that may be used to better predict travel demand taking into account the spatial context, which is critical for travel.

CONCLUSIONS

Much of the focus has been placed on understanding links between non-spatial travel activities and socio-demographic factors by researchers as well as practitioners; few studies have explored on understanding and modeling the spatial variations in socio-economic factors associated with household travel decisions. This paper fills the gap, and makes a timely contribution, given the need to improve the accuracy of travel demand models and emerging spatial analysis methods. A critical finding is that unobserved spatial heterogeneity exists in trip production models, since the mean and the variance of trip frequency and associated socio-demographical factors are not identical across the Hampton Roads metropolitan area; GWPR, a local model, provides a more accurate representation of travel demand since it can captures the unobserved spatial heterogeneity in trip production models. The results show that the association between household trip frequency and household characteristics such as household size and vehicle count varies significantly over the Hampton Roads Area. Statistical analysis shows that the local models provide a better statistical fit.

The methodology used in this paper can enhance the ability to forecast travel demand by quantifying the associations that vary across space for various factors (household size, vehicle ownership, income). This study demonstrates how new methods that incorporate spatial information can be applied to improving travel demand models. The spatial pattern of parameters shown by this study provides a clearer understanding of the link between space and travel behavior. The detailed association maps obtained from these models are valuable in quantifying spatial variations and making them available in easy to understand format to policy makers, engineers, and planners. Moreover, this paper is of interest to people who use travel demand forecasting models and also to policy makers who rely on these models for making their decisions. The research can help advance the state-of-the-art in travel behavior research by using rigorous analysis techniques to incorporate spatial variations into travel demand models. Practitioners may capitalize on the greater spatial variability in parameters to develop locally-based strategies for trip reduction, especially where trips are particularly numerous. Future explorations may delve into land-use and built environmental variables associated with trip generation, analyze trips by purpose and mode, and also look into trip attractions.

ACKNOWLEDGMENTS: We are thankful to Mr. Paul Angello for providing the data. We have benefited from input provided by Dr. Mecit Cetin. This work was supported by the Virginia Department of Transportation and conducted at the Transportation Research Institute, Old Dominion University.

References

1. Robinson, R., et al., *Hampton Roads Transportation Alternatives Traffic Simulation, Final Report to Virginia General Assembly*. 2008, Virginia Modeling, Analysis and Simulation Center, Old Dominion University: Norfolk.
2. Tobler, W., *A computer movie simulating urban growth in the Detroit region.* Economic Geography, 1970. **46**(2): p. 234-240.

3. Buliung, R.N. and P.S. Kanaroglou, *Activity-Travel Behaviour Research: Conceptual Issues, State of the Art, and Emerging Perspectives on Behavioural Analysis and Simulation Modelling.* Transport Reviews, 2007. **27**(2): p. 151-187.
4. Miller, H.J., *Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation (GIS-T).* Geographical Analysis, 1999. **31**: p. 373-399.
5. Fotheringham, A.S., C. Brunsdon, and M. Charlton, *Geographically Weighted Regression, the analysis of spatially varying relationships*, ed. U. University of Newcastle. 2002: John Wiley & Sons, Ltd.
6. Lloyd, C.L., *Local Models for Spatial Analysis*. 2007, Boca Raton, FL: CRC Press, Taylor & Francis Group.
7. Nakaya, T., et al., *Geographically weighted Poisson regression for disease association mapping.* Statistics in Medicine, 2005. **24**(17): p. 2695-2717.

Evaluating Vehicle Queue Length Estimations at the HRBT Based on Probe Vehicle Data Using Micro-Simulation

Craig Jordan and Mecit Cetin

Abstract—**Accurate and up to date traffic condition estimations are vital to assisting travelers in making educated route and mode choices, improving mobility, increasing safety, and helping the environment. This paper is part of ongoing research that examines analytical models for the real-time estimation of vehicle queue lengths using probe vehicles. The purpose of the research is to understand how information communicated from probe vehicles can be used to improve current methods for estimating traffic parameters such as volume and density. This paper compares analytical methods for queue length estimation at the HRBT with the simulated "actual" queue lengths from VISSIM, a micro-simulation software program.**

***Index Terms*—IntelliDrive, Micro-simulation, Vehicle Queues, Transportation.**

I. Introduction

There has been an increase in demand for research in Intelligent Transportation Systems (ITS) in recent years. In 2009, the US Department of Transportation developed a five year strategic research plan with the purpose of improving safety, mobility, and the environment. The foundation for the plan is a program called IntelliDrive, which uses a wireless communication system between vehicles and infrastructure to provide a connection for the transmission of traffic flow information. For simplicity, the term "probe vehicle" will be used throughout this paper to describe a simulated IntelliDrive vehicle that has been equipped with a Global Positioning System (GPS) and a wireless communication device.

The purpose of this paper is to support the development of IntelliDrive systems by investigating how the information obtained from probe vehicles can be utilized to estimate queue lengths and traffic patterns more accurately. In particular, this research spotlights the understanding of the relationship between the accuracy of the queue length estimates and the market penetration (the percentage of probe vehicles in the vehicle population) of probe vehicles.

In the Hampton Roads Region of Virginia, reoccurring daily bottlenecks occur at specific locations during peak hour traffic flows. One of these locations is the Hampton Roads Bridge Tunnel (HRBT) that connects the Hampton Roads Upper Peninsula with the Norfolk, Virginia Beach, and Chesapeake areas. During the AM and PM peak hours, the HRBT experiences high traffic volumes in both the northbound and southbound directions (outer and inner loops, respectively). Operational disruptions at the HRBT caused by disabled vehicles or vehicle collisions cause significant vehicle queuing and traffic delays.

Because of the reoccurring traffic congestion at the HRBT and the availability of traffic flow rates from loop detectors and vehicle queue lengths from surveillance cameras to support future work, this study will use the location along I-64 in the vicinity of the HRBT as the study site for this research.

II. Methodology

The objective of this study is to determine the accuracy of the estimated vehicle queue lengths based on the location and the time of probe vehicles at different market penetration levels. The steps outlined below provide an overview of the methodology used in this study to analyze the estimated queue lengths.

Step 1: Develop a micro-simulation traffic model
Step 2: Run micro-simulation to model vehicle queues
Step 3: Output queue data from micro-simulation that will be used as the "actual queue" for comparison with queue length estimate
Step 4: Randomly assign vehicles in the queue as "probe vehicles"
Step 5: Use conditional probability formula to estimate the length of the queue based on the probe vehicle data
Step 6: Compare the estimated queue length with the actual queue length through the use of the root mean square error technique

Each of these steps is explained in more detail in the following sections.

III. Assumptions and Clarifications

In order to perform the analysis in this study certain assumptions were made and are listed below.

1) The storage capacity of the roadway at queue locations is infinite
2) The location of all probe vehicles in the queue is observable on the roadway network
3) The time of entry at the back of queue for each probe vehicle is observable

In order to simplify calculations, the location of the vehicles in the vehicle queue and the total queue length is measured in terms of the number of vehicles rather than in feet.

For clarification purposes, all reference to "estimated queue length" is defined by the estimated queue length calculated by the data from probe vehicles.

IV. Analytical Method of Queue Length Estimation

The method for estimating queue lengths uses a conditional probability formula. A conditional probability is the probability that an *event A* will occur given the fact that *event B* occurs. In the queue length estimation, *event B* is comprised of two events. The first being the location of the last probe vehicle as it enters a queue and the second being the time the probe vehicle enters the queue. This conditional probability formula was determined in the previous work of Comert and Cetin[1].

The formula is provided below and a description follows:

$$E(N|L=l,\ T=t)=l+\theta\delta,\ l \geq 0,\ t \geq 0 \qquad (1)$$

E = expected value

N = total number of vehicles in queue

L = location of probe vehicle

T = time instance probe vehicle enters queue

δ = elapsed time since time T

$\theta = \lambda(1-p)$

λ = vehicle arrival rate

p = market penetration

This equation states that if the following conditions are met,

1) The location of a probe vehicle L is equal to a known location l and

2) The time the probe vehicle enters the queue T is equal to a know time t

then the expected length of the vehicle queue is equal to l (the location of the probe vehicle) plus θ (the vehicle arrival rate multiplied by (1 minus the market penetration percentage)) multiplied by δ (the length of time period elapsed since the arrival of last probe vehicle).

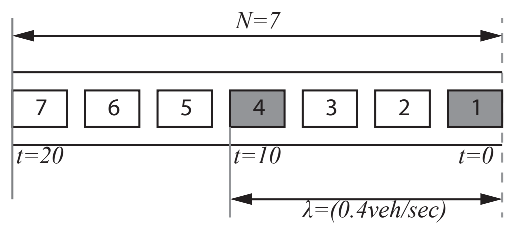


Figure 1: Vehicle Queue Snapshot

As an example, Figure 1 provides a visual representation of a vehicle queue with a probe vehicle market penetration of 10%. Each box in the figure represents a queued vehicle. The solid boxes represent probe vehicles and the hollow boxes represent non-probe vehicles. At time t = 0 seconds, probe vehicle 1 enters the queue and at t = 10 seconds, probe vehicle 2 enters the queue. The numbers in the boxes indicate the vehicles' position in the queue. The time between the arrival of the first probe vehicle and the second is 10 seconds. The position of the second probe vehicle is the fourth vehicle. Therefore, the arrival rate (λ) is 0.4 vehicles per second. To estimate the queue length at time = 20 seconds the equation is as follows:

$$E(N|L=4,T=20)=4+(0.4veh/sec)(1\text{-}0.1)(20sec\text{-}10sec)$$
$$E(N|L=4,T=20)=7.6$$

The partial vehicle is rounded up to 8 vehicles. This is the estimate for the vehicle queue length at time 20 seconds.

V. Micro-simulation

Due to the limited market penetration of probe vehicles and the lack of infrastructure to obtain probe vehicle data, a transportation model was created to simulate the probe vehicles in traffic flow and vehicle queues at the HRBT. The software program VISSIM was used for the model development.

VISSIM is a microscopic simulation (micro-simulation) software program that applies time step and behavior based simulation to model traffic operations. The program models every vehicle in the simulation as a separate object with a specific set of car following and lane changing behaviors. The program keeps track of the unique attributes for every vehicle in the model at each time step. The unique driving characteristics and flexibility of the software allow complex transportation models to be developed.

VISSIM has a COM (communication) interface that provides users the capabilities to control certain functions of the micro-simulation with outside programs. For this research, MS Excel and Visual Basic for Applications (VBA) was used. This allowed for information to be passed back and forth between the programs.

To evaluate the analytical methods for queue length estimation, a model of the Interstate 64 (I-64) northbound (the outer loop) at the HRBT was created in VISSIM, a microscopic simulation tool. The model included existing roadway geometry, lane widths, number of lanes, speed limits, and lane changing permissions. To simplify the simulation, only the mainline and mainline traffic flow rates were modeled. In addition, vehicle composition included passenger cars only.

A simulated incident was modeled in the HRBT to create vehicle queues for evaluation. The duration of the incident lasted for five minutes during which both lanes are blocked. After the incident, vehicles resumed traveling at their desired levels. The analysis is performed only for the duration of incident.

VI. Market Penetration Assignment

The market penetration of vehicles was determined after the completion of the simulation run and all vehicle queue data was exported into MS Excel. The market penetration assignment was determined by extracting the unique identification numbers for each of the queued vehicles and assigning them a binary designation ("1" for probe vehicle and "0" for non-probe vehicle). The rand() function was used to assign the numbers to each vehicle. This function uses a uniform distribution to create random numbers thus providing an equal chance for each vehicle to be a probe vehicle. This was done for each of the nine market penetration levels. In addition, the probe vehicles were determined independently for each market penetration level. Therefore, the probe vehicles in the 10% market penetration were not necessary the same vehicles as the probe vehicles in the 20% market penetration. The following section provides more information on the scenarios that were run.

VII. Analysis Scenarios

Multiple analysis scenarios were evaluated at varying market penetration percentages. These scenarios included two traffic flow rates (1800 veh/hr and 2700 veh/hr) for which nine market penetration levels were evaluated. This provided a total of 18 analysis scenarios. The results of the scenarios were compared to determine accuracy of queue estimation for the different market penetration values.

VIII. Data Collection and Analysis

As stated earlier, real-world data that includes varying degrees of market penetration of probe vehicles are not available; therefore, data was collected from the VISSIM micro-simulation tool. The micro-simulation was used exclusively for the simulation of traffic flow and the formation of realistic vehicle queues. It was not used in the determination of which vehicles were probe vehicles. As stated in Section VI. Market Penetration Assignment, this step was performed outside of VISSIM.

Data was collected for any vehicle determined to be in a queue for each simulation second throughout the duration of the simulation. A vehicle was determined to be in a queue if its speed was 4 miles per hour or less.

A VBA script was used to extract data from VISSIM. A loop function was created that paused VISSIM at each simulation second to check for any vehicles meeting the queue requirement. If a vehicle was found to be in a queue, the simulation time, vehicle queue position, vehicle identification number, vehicle speed, and the lane number that the vehicle was in was exported from VISSIM into MS Excel.

After the vehicle data was exported from VISSIM another VBA script was used to perform the queue length estimate calculations using the conditional probability function. The following logical flow chart illustrates how the script runs.

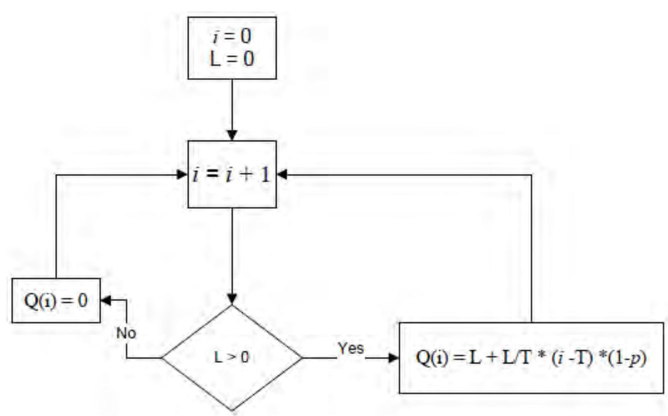


Figure 2: Queue Estimation Logic Flowchart

i: Simulation step

p: Percent of probe vehicles (known)

L: Location of the last probe vehicle in queue (measured in terms of number of vehicles)

T: Arrival time of the last probe vehicle to the back of the queue measured in simulation steps)

$Q(i)$: Estimated queue length at time step i in terms of number of vehicles

IX. Results

For each time step during the analysis period, an estimate of the queue length was determined for the time between the arrivals of probe vehicles at the back of queue. Figure 3 provides a graph comparing the actual queue length, location of the probe vehicles, and the estimated queue length over real-time at the 10% market penetration level.

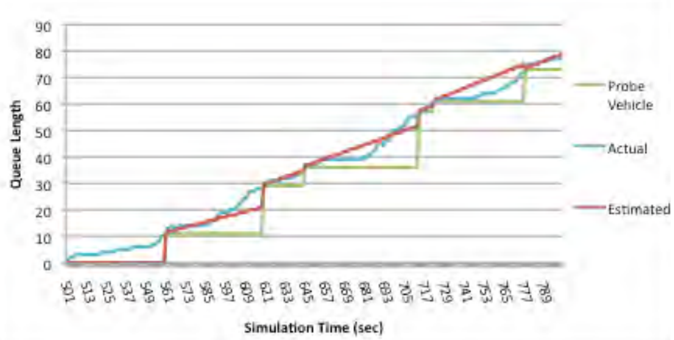


Figure 3: Actual vs 10% Market Penetration Estimate Lane 1

This graph illustrates how the limited number of probe vehicles doesn't provide a very accurate estimate of the queue length. This can be seen at time 609 where the actual queue length is 28 vehicles while the estimate is 20 vehicles.

In contrast to the 10% on lane 1, the 30% market penetration level on lane 2 shows a better estimate for the queue length. There are only a few locations where the estimate is significantly different than the estimated.

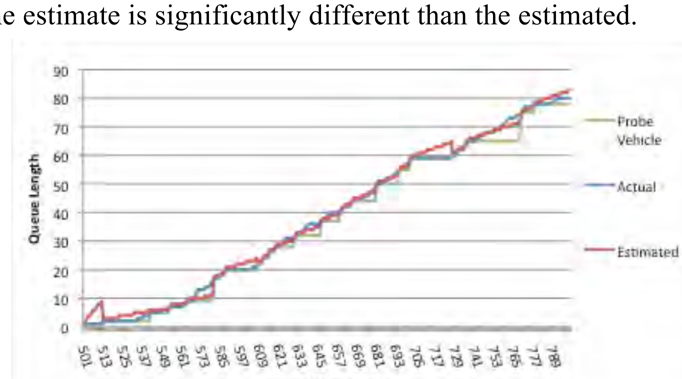


Figure 4: Actual vs 30% Market Penetration Estimate Lane 2

In addition to a visual inspection of the estimation graphs, the root mean square error (RMSE) was calculated for each of the market penetration levels. The RMSE is used to analyze the accuracy of the estimates. It is calculated by squaring the difference between the estimated queue length and the actual queue length produced in VISSIM. The squared values are then averaged and the square root is taken. The RMSE indicates how close the estimated values are to the actual values. The lower the value of the RMSE, the closer the fit is to the actual value. A perfect fit would have an RMSE value of 0.

Tables 1 and 2 are charts of the RMSE values for the two traffic flow scenarios. Each chart illustrates the differences between RMSE values for each of the market penetration percentage rate. It is more clearly shown in Table 2 at 2700 veh/hr that the RMSE values decrease with the increase in market penetration.

Table 1: RMSE Values for 1800 veh/hr at Specific Market Levels

Market Penetration Levels	RMSE Values		
	Lane 1	Lane 2	Combined
10%	3.6	3.0	3.3
20%	2.5	2.0	2.3
30%	5.4	1.9	4.0
40%	3.6	1.2	2.7
50%	1.3	1.1	1.2
60%	1.5	1.1	1.3
70%	1.2	1.0	1.1
80%	1.0	1.0	1.0
90%	0.9	0.9	0.9

Table 2: RMSE Values for 2700 veh/hr at Specific Market Levels

Market Penetration Levels	RMSE Values		
	Lane 1	Lane 2	Combined
10%	6.5	5.8	6.1
20%	6.8	4.6	5.8
30%	3.5	4.2	3.9
40%	2.2	3.8	3.1
50%	1.5	2.2	1.9
60%	1.1	1.6	1.4
70%	0.6	1.0	0.8
80%	0.9	0.4	0.7
90%	0.3	0.4	0.3

X. CONCLUSIONS AND FUTURE WORK

The results presented in this paper are preliminary and have not modeled the specific aspects of vehicle-to-vehicle communication. In addition, these results only represent one model run for each traffic flow rate scenario. Although the results indicate that market penetration rates around 30% provide accurate vehicle queue estimations, more research is needed to fully understand the relationship. Also, when the results from the low volume scenario are compared to high volume scenario, it is observed that the error (measured in terms of RMSE) is lower for all market penetration levels less than 70%.

In the future, we will investigate how quickly and how far upstream the information propagates based on limited communication range of current IntelliDrive technology. In addition, the accuracy of the estimates will be investigated with respect to time and space at different market penetration levels.

REFERENCES

[1] G. Comert, M. Cetin, “Analytical Evaluation of the Error in Queue Length Estimation at Traffic Signals From Probe Vehicle Data”, to appear in the IEEE Transactions on Intelligent Transportation Systems, 2011.

[2] W. Mendenhall, T. Sincich, “Statistics for Engineering and the Sciences” 5th ed., Upper Saddle River: Prentice Hall, 2007

[3] *VISSIM Com Guide*, Version 5.3, PTV AG, 2009-2010.

Freeway Lane Utilization Behavior

Sanghoon Son, Mecit Cetin, and Asad Khattak, Old Dominion University

***Abstract*— Variation in flow and speed exists across freeway lanes. This paper investigates and quantifies drivers' lane utilization behavior on the freeway. A lane utilization factor, representing uneven distribution of traffic flow, has been applied in the study. This paper analyzes traffic volume data (N=11,015) and roadway geometric data (11 basic freeway segments and 8 ramp sections) in high flow situations in Hampton Roads, Virginia. Lane use behavior is explored through time series and flow proportion plots. Descriptive analysis of the data is provided, followed by statistical models. Ordinary Least Squares regression models are estimated to identify and quantify key factors associated with lane utilization behavior. Results show that traffic flow distribution on freeway lanes is less likely even if average speeds and posted speed limits are high. When there is the presence of major ramps (freeway-freeway) in upstream or downstream and the presence of minor ramps (freeway-arterial) in upstream or downstream, traffic flow distribution is less likely equal. However, it is not consistent to ramp sections where an entry or an exit ramp is placed within a half mile. It is found that lane flow distribution in ramp sections is more likely uniform, indicating that the shoulder lane in ramp sections is where lower speed with less vehicle space are observed.**

Keywords-lane utilizaiton behavior, lane utilizaiton factor, basic freeway segment, ramp section, regression model

I. INTRODUCTION

Variations in traffic flow across freeway lanes can occur over time and space [1-2]. For each traffic condition, usually varying by time of day, the lane that is heavily used on the freeway fluctuates. Specifically, under the low flow situation in the three-lane freeway, the middle lane is generally preferred more than other lanes. As flow increases from moderate to high level, the median lane is more likely to be utilized. After congestion occurs, depending on traffic situations in downstream, it varies; however, the median lane is mostly dominant over other lanes. This is mainly because of the habitual driving rule. In other words, slower vehicles tend to drive in the shoulder lane on the freeway while relatively faster vehicles drive on the middle or median lane. However, the typical lane use pattern is not always constant over the entire freeway.

In addition to the driving habit, what else influences lane use behavior? Above all, the origin and destination (OD) of individuals and their corresponding desired paths are important to their lane choices [3]. Conceptually, it is logical; however, collecting individuals' OD is difficult for the given freeway segment. Instead, if demand is sufficient enough, roadway geometry may be able to explain lane utilization patterns. This is because drivers may prefer to use certain lanes and switch lanes in response to their entry and exit locations along the ways. In this regard, it would be interesting to investigate the relationships between lane utilization behavior and roadway geometry in high demand where freeway ramps are frequently used. Of course, there might be other factors impacting on the drivers' lane utilization behavior. They might be traffic characteristics: volume, speed, and density.

Understanding lane utilization behavior would be beneficial. It can be applied for better traffic operation and management. Also, traffic flow can be more accurately simulated in a micro-simulation from understanding lane utilization behavior. To do this, the study aims to explore and model the freeway lane use behavior and relationships between lane utilization behavior and corresponding factors. The study begins with introducing a lane utilization factor for explaining lane utilization behavior. Lane use behavior is explored through time series and flow proportion plots. Next, the study area and empirical traffic data with roadway geometric inventories are presented. After that, regression models are specified and estimated to quantify the factors associated with lane use variation based on roadway characteristics and traffic characteristics. Finally, results are interpreted and conclusions are drawn.

II. LANE UTILZATION BEHAVIOR

Lane use behavior can be represented by the lane utilization factor. The lane utilization factor captures non-uniform lane use; that is, variation in flows or speed across lanes. The higher lane utilization factor indicates that lane flow is evenly distributed and vice versa across lanes. The lane utilization factor is usually defined as follows: [4-7]

$$f_{LU} = V_g / (V_{g1}N) \qquad (1)$$

where,

V_g = flow rates for all lane (veh/h)

V_{g1} = flow rates on single lane with highest volume (veh/h), and

N = number of lanes.

This manuscript is submitted for 2011 Capstone Conference at Virginia Modeling, Analysis, and Simulation Center.

Sanghoon Son is with the Department of Civil and Environmental Engineering, Old Dominion University, Norfolk, VA, 23529, USA (email: ssanghoon.son@gmail.com).

Mecit Cetin is with the Department of Civil and Environmental Engineering, Old Dominion University, Norfolk, VA, 23529, USA (phone: 757-683-6700, email: mcetin@odu.edu).

Asad Khattak is with the Department of Civil and Environmental Engineering, Old Dominion University, Norfolk, VA, 23529, USA (phone: 757-683-5515, email:akhattak@odu.edu).

This study is to explain lane utilization behavior with the lane utilization factor. Figure 1 presents an example of the average traffic flows in a time-series plot for where there are three lanes, and a ramp is located within 1 mile upstream and downstream, respectively. Traffic flow profile is plotted based on the average of 15-min observed volumes for three months, April to June in 2007, to show the overall lane choice patterns. In the plot, lane 2 (middle lane) is most frequently used while lane 1 (shoulder lane) and lane 3 (median lane) are less likely to be used. On the other hand, lane 3 is more heavily used than other lanes during the morning peak.

Figure 2 shows the average travel speed in a time-series plot for the same site in Figure 1. In the plot, average speed in lane 3 (median lane) is always higher than that of lane 1 (shoulder lane) and lane 2 (middle lane). Interestingly, in the morning peak, the profile of the lane utilization factor seems to follow the speed profiles in a similar way. It is interesting to consider average speed as an associate with unbalanced lane use.

Figure 3 displays the proportion of traffic flows for each lane over all traffic flows in terms of the overall flows in an ascending order. When flows ranges from 2500 to 5500, all lanes are more likely to be used equally whereas when the flows are greater than around 5500, lane utilization factor decreases. It is interesting to investigate where demand is high and there may be an association with potential factors to cause uneven lane flow distribution.

Overall, for this site, the morning peak is where high demand can be observed and f_{LU} drops as there is more variation in flow. For the remaining study, lane utilization variation, measured by the lane utilization factor, is assumed to be impacted by roadway geometric conditions as well as traffic characteristics.

III. STUDY AREA AND DATA

To understand lane utilization behavior in high flow situations and to develop a statistical model for the lane utilization factor as a function of roadway characteristics, basic freeway segments and ramp segments that have vehicle detectors within the Hampton Roads network are identified. Hampton Roads is a metropolitan area in Southeastern Virginia and has an extensive transportation network. I-64, the major east-west route, connects to I-264, I-464, I-564, and I-664. Especially, 3-lane basic freeway segments, 3-lane with non-separated HOV lane segments and 3-lane ramp sections are focused for this study.

Study sites are listed with roadway inventory data in TABLE I. The data covers a wide range of geometric characteristics. Figure 4 shows the geographic locations of the study sites. Overall, 19 sites (11 basic freeway segment and 8 ramp sections) were selected, and at each site, geometric attributes are collected such as the distances to the nearest minor ramp (arterial-freeway) and major ramp (freeway-freeway) with posted speed limits. Even though one can consider other ramps in the vicinity of the segment (e.g., the second nearest ramps), the analysis in this paper are limited to the nearest ramps which are expected to affect lane utilization behavior more significantly.

For each study site, traffic volume data were processed. The Virginia Department of Transportation maintains a Traffic Monitoring System (TMS) which collects classified speed and counts from each continuous count station using loop detectors. The TMS raw data includes 15-minute intervals with 21 speed bins categorized at 5 mph intervals. Also, vehicles are classified into three categories: passenger car, single unit, and tractor trailer. Vehicle count data collected from January to June, in 2007, excluding weekends and holidays, are analyzed. The data was error-checked and cleaned.

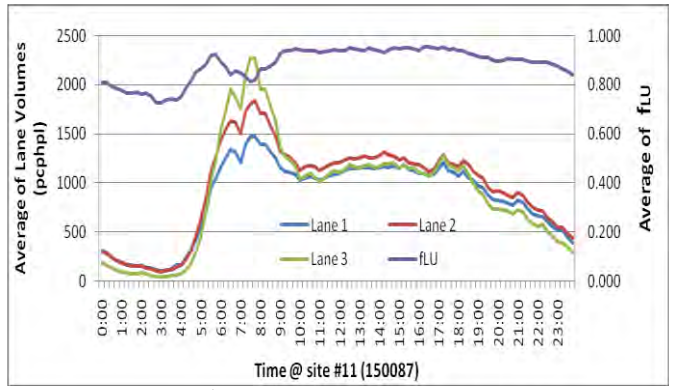


Figure 1 Illustration of time-series plot showing average lane volumes and average fLU

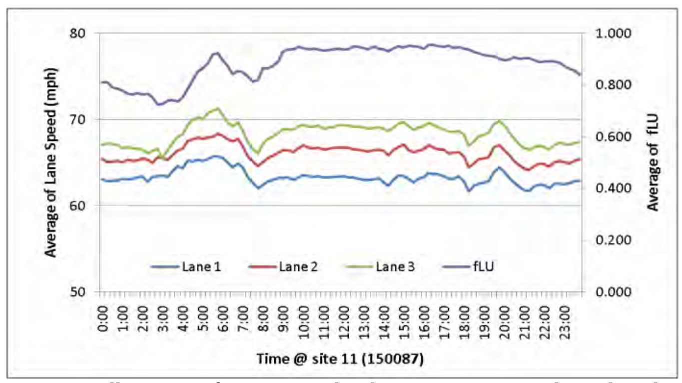


Figure 2 Illustration of time-series plot showing average travel speeds and average fLU

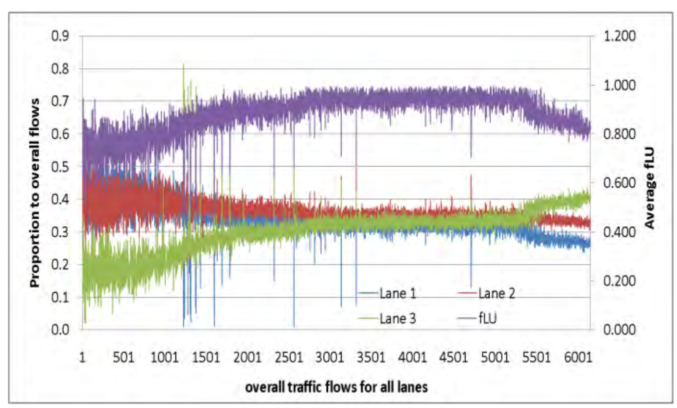


Figure 3 Illustration of proportion of each lane and average fLU along with overall traffic flows

TABLE I. SUMMARY OF STUDY SITES AND GEOMETRY INVENTORY (DISTANCES ARE IN MILES)

#	ID	Interstate	Nearest Arterial-Freeway Ramp		Nearest Freeway-Freeway Ramp	
			Up Stream	*Down stream*	*Up stream*	*Down stream*
1	50163	I-64	0.80	0.48	*	6.69
2	50169	I-64	1.55	0.32	1.80	5.23
3	50306	I-64	0.44	0.48	14.72	4.89
4	50616	I-264	1.13	0.9	3.65	1.92
5	150012	I-664	0.58	0.34	12.44	7.65
6	150028	I-464	0.29	1.32	2.41	3.22
7	150033	I-464	1.05	0.36	3.17	2.44
8	150051	I-64	0.9	0.29	7.23	*
9	150065	I-64	0.43	0.44	5.77	*
10	150070	I-64	0.35	1.70	5.85	1.80
11	150087	I-264	0.70	0.95	1.67	3.18
12	50165	I-64	0.22	0.37	6.5	*
13	50216	I-264	1.24	0.2	5.5	*
14	50305	I-64	0.18	0.44	5.5	*
15	150067	I-64	0.91	0.23	*	3.5
16	150068	I-64	0.23	0.36	4.25	4.6
17	150071	I-64	2.04	0.19	4.6	2.8
18	150079	I-264	0.34	0.25	4	2.7
19	150321		0.22	1.7	3.2	4.3

Note: *The distance is longer than 20 miles.

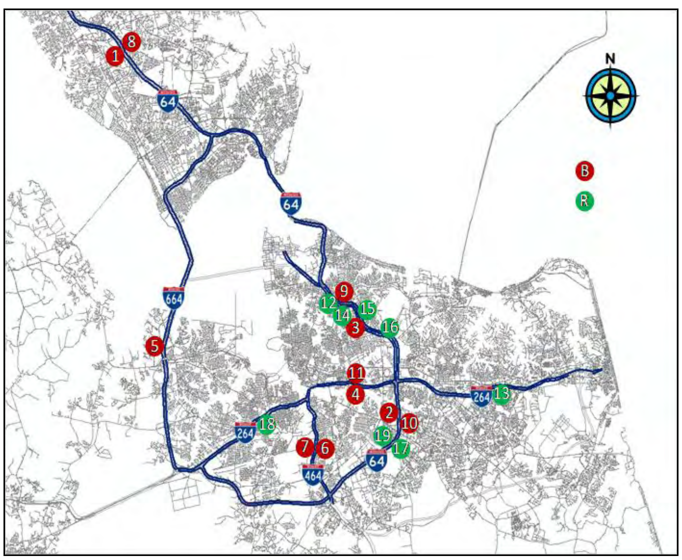


Figure 4 Location of the study sites

IV. STATISTICAL MODELING AND RESULT

A. Statistical Modeling

To quantitatively test the association of upstream and downstream roadway characteristics with lane use behavior, Ordinary Least Squares (OLS) regression models are estimated in high flow situations. The models are adopted because of the simplicity and intuitive interpretation of the parameters. Statistical software SPSS is used for estimation of the models. Traffic operating and roadway geometric factors are added to the models as explanatory variables. Among traffic variables, average speed is used to control traffic condition. Additionally, posted speed limits are added as an additional independent variable because they may produce more variation across freeway lanes. For roadway geometric factors, basically four categorical variables ($UpMinorRamp_{distance}$, $DwMinorRamp_{distance}$, UpMajorRamp, and DwMajorRamp) are created as shown in Equation 2. Subsequently, two variables (UpMinorRamp0-1 and DwMinorRamp0-1) out of four variables are replaced by four more detailed category variables (UpMinorRamp0-0.5, UpMinorRamp0.5-1, DwMinorRamp0-0.5, and DwMinor Ramp0.5-1) in order to specify the impact of ramp types and their distances to each study site more accurately. Overall, three different statistical models are estimated with the different set of variables. The detailed OLS regression models are as follows:

$$f_{LU} = \beta_0+\beta_1*AveSpeed+\beta_2*UpMinorRamp_{distance}+\beta_3*DwMinorRamp_{distance}+\beta_4*UpMajorRamp+\beta_5*DwMajorRamp+\beta_6*PostedSpeedLimit+\varepsilon \quad (2)$$

where,

f_{LU}	lane utilization factor ($1/3 \leq fLU \leq 1$)
AveSpeed	average speed at given time and site (mph)
UpMinorRamp0-1	arterial-freeway on or off-ramp within 1mile upstream (1=yes, 0=no),
UpMinorRamp0-0.5	arterial-freeway on or off-ramp within 0.5mile upstream (1=yes, 0=no),
UpMinorRamp0.5-1	arterial-freeway on or off-ramp within 0.5 to 1mile upstream (1=yes, 0=no),
DwMinorRamp0-1	arterial-freeway on or off-ramp within 1mile downstream (1=yes, 0=no),
DwMinorRamp0-0.5	arterial-freeway on or off-ramp within 0.5mile downstream (1=yes, 0=no),
DwMinorRamp0.5-1	arterial-freeway on or off-ramp within 0.5 to 1mile downstream (1=yes, 0=no),
UpMajorRamp	freeway-freeway on or off-ramp within 3mile upstream (1=yes, 0=no),
DwMajorRamp	freeway-freeway on or off-ramp within 3mile downstream (1=yes, 0=no), and
PostedSpeedLimit	posted speed limits (1=60mph, 0=55mph)
ε	error term.

Two criteria are adopted in order to capture high demand with no traffic breakdowns. This is because once breakdown takes place drivers are restricted in their choice of lanes. Samples, therefore, are screened by criteria that volumes in passenger car unit should be greater than 5,400 pcph (=1,800 pcphpl*3 lanes) and average speed should be greater than 50 mph. Overall, 11,015 observations are extracted from over 19 sites. TABLE II presents the descriptive statistics for samples. Average lane utilization factors, average speed, and densities are summarized. The average lane utilization factor varies across sites and it is relatively high, 0.886. The average flows for the selected sample are 5,944 pcph, which are reasonably high.

TABLE II. DESCRIPTIVE STATISTICS

Variables	N	Min	Max	Mean	Stdv
f_{LU}	11,015	.753	.998	.886	.046
TotalFlow	11,015	5,402	7,540	5,944	397
Density	11,015	26.7	48.1	32.6	3.5
AveSpeed	11,015	50.0	68.4	61.2	3.8
NearUpRamp	11,015	0	1	.80	.402
NearDwRamp	11,015	0	1	.91	.283
NearUpRamp-05	11,015	0	1	.41	.491
NearUpRamp05-1	11,015	0	1	.39	.488
NearDwRamp-05	11,015	0	1	.72	.448
NearDwRamp05-1	11,015	0	1	.19	.393
UpInter	11,015	0	1	.09	.289
DwInter	11,015	0	1	.31	.464
Speed	11,015	0	1	.17	.372

B. Results

Three lane utilization behavior models are presented and compared in TABLE III. All models have relatively good statistical fits (R^2 and adjusted R^2) and overall significance (F-test). Variance inflation factor for each variable is checked (<10), indicating that collinearity among explanatory variables is not problematic. The models indicate that variables have the expected signs and relatively reasonable magnitudes. All variables are also significant based on their p-values (<0.05).

Here, the higher lane utilization factor indicates that evenly distributed lane flow is expected and vice versa. For all models, expectedly, average speed has negative impact on the lane utilization factor, indicating as average speed increases freeway lanes are more likely to be utilized unevenly. In addition to average speed, posted speed limit is negatively correlated with the lane utilization factor. It means that the higher posted speed limit produces more variation in flow across lanes.

Each estimate of roadway geometric indicator variable, 0 or 1, implies the association of its presence with lane utilization behavior, i.e., the magnitude of the coefficient indicates the association while holding other variables constant. For all models, major (freeway-freeway) ramps within 3 miles of upstream and downstream in freeway segments are negatively correlated with the lane utilization factor. The presence of the major ramps in upstream and downstream is associated with decreases in the lane utilization factor by 0.042 from 0.023 in Model 1, respectively. Model 2 and Model 3 also have similar negative coefficient values: -0.029 and -0.021 in Model 2 and -0.028 and -0.020 in Model 3, respectively.

However, minor (arterial-freeway) ramps show different results depending on their distances to the corresponding ramps. Minor ramps within 1 mile of upstream and downstream in freeway segments are positively correlated with the lane utilization factor. The presence of minor ramps is associated with increases in the lane utilization factor, resulting in more uniform traffic flow distribution across lanes. This result is counter-intuitive; therefore, further analysis has been conducted by breaking down the distance.

As a result, more even flow distribution can be found if there is the presence of minor ramps within 0.5 mile of upstream or downstream. The presence of the minor ramps is associated with increases in the lane utilization factor by 0.020 from 0.032, respectively, in Model 2. In contrast, it is not the same if minor ramps are located further in upstream or downstream. In this case, the presence of them is associated with decreases in the lane utilization factor by 0.004 and 0.052, respectively, in Model 3

Overall, two multiple regression models (Model 2 and Model 3) indicate the impact of ramp types and ramp locations on the freeway in the urban area in terms of lane utilization behavior. This field-measured data shows variations in lane utilization by roadway geometric conditions, explaining traffic flow dynamics on basic freeway segments and ramp sections. Summing up, lane utilization behavior in major ramps is consistently associated with roadway geometry, whereas that of minor ramps is dependent upon the distances to the subject ramps.

TABLE III. STATISTCAL MODELING RESULTS (ALL DENSITY REGIME)

Independent Variables	Model 1	Model 2	Model 3
Constant	1.111	1.046	1.013
AveSpeed	-.004	-.003	-.002
Upstream Minor Ramp0-1	.033	-	-
Upstream Minor Ramp0-0.5	-	.020	-
Upstream Minor Ramp0.5-1	-	-	-.004
Downstream Minor Ramp0-1	.014	-	-
Downstream Minor Ramp0-0.5	-	.032	-
Downstream Minor Ramp0.5-1	-	-	-.052
Upstream Major Ramp	-.042	-.029	-.028
Downstream Major Ramp	-.023	-.021	-.020
PostedSpeed Limit	-.020	-.014	-.027
N	11,015	11,015	11,015
F	1035.652	1313.399	1330.028
R^2	.361	.417	.420
Adjusted R^2	.360	.417	.420

Notes: A dependent variable is fLU. All coefficient values are statistically significant.

V. LIMITATIONS

Results should be carefully interpreted. The main limitation of this study has to do with the relatively small number of sites analyzed. A larger sample size is needed to create models that can be transferable to other regions. Also there might be more factors like lane discontinuity. This does not take into account in the study.

VI. CONCLUSTION AND IMPLICATIONS

This study contributes by comprehensively exploring and quantifying lane utilization behavior in high flow situations in the urban area with regards to traffic and roadway geometric characteristics. The primary purpose of this study is to investigate factors associated with lane utilization on basic freeway segments and ramp sections and to develop statistical models that can help to understand lane use behavior. Using relatively good detector data from Hampton Roads, Virginia, and rigorous statistical models, the study identifies correlates of the lane utilization factor. This study found that average speed, ramp types, ramp locations, and posted speed limit are associated with variation in lane utilization. Specifically, the more variation in flow across freeway lanes in high demand can be found where average speed or posted speed limit is higher, indicating high speed allows a median lane to process more traffic, whereas a shoulder lane still handle almost same amount of entry and exit traffic. In addition, the more variation in flow across lanes can be seen if a major ramp is present within 3 miles of upstream or downstream, or a minor ramp is present from 0.5 to 1 mile of upstream or downstream. It indicates that traffic flow on basic freeway segments avoids using a shoulder where through movement are disturbed by entry or exit movement; therefore other lanes are preferred. Unlike basic freeway segments, the less variation can be observed in ramp sections where a minor ramp is located within 0.5 mile of upstream or downstream. This is partly because in ramp sections a shoulder lane where usually process less amount of traffics on basic freeway segments can handle more traffic. This is also because of less vehicle space and many of entry and exit traffic in high demand. Overall, the results are consistent with expectations and provide better understanding of traffic flow across freeway lanes.

REFERENCES

[1] V. Hurdle, *et al.*, "Study of Speed-Flow Relationships on Individual Freeway Lanes," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1591, pp. 7-13, 1997.

[2] M. Amin and J. Banks, "Variation in Freeway Lane Use Patterns with Volume, Time of Day, and Location," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1934, pp. 132-139, 2005.

[3] C. F. Daganzo, "A behavioral theory of multi-lane traffic flow. Part I: Long homogeneous freeway sections," *Transportation Research Part B: Methodological*, vol. 36, pp. 131-158, 2002.

[4] Transportation Research Board, Highway Capacity Manual 2000. Washington, D.C.: Transportation Research Board, National Research Council, 2000.

[5] J.-J. Lee, et al., "Models for Lane Utilization Prediction for Lane Drop Intersections," Transportation Research Record: Journal of the Transportation Research Board, vol. 1912, pp. 47-56, 2005.

[6] J. Banks, "Part 1: Traffic Flow Theory and Characteristics: Effect of Time Gaps and Lane Flow Distributions on Freeway Bottleneck Capacity," Transportation Research Record: Journal of the Transportation Research Board, vol. 1965, pp. 3-11, 2006.

[7] J. Lee and B. B. Park, "Lane Flow Distributions on Basic Segments of Freeway Under Different Traffic Condistions," presented at the 89th Annual Meeting of Transportation Research Board Washington DC, 2010.

Model Transferability in the Re-identification of Trucks Based on AVC and WIM Data

Ilyas Ustun, Mecit Cetin

***Abstract*- Origin-destination (OD) flows are important for transportation planning and modeling. Trucks crossing two weigh-in-motion (WIM) sites are anonymously re-identified based on axle spacing and axle weight data. The sites selected for this study are separated by more than 100 miles which are located in Oregon. Two different OD pairs or links are considered in this study. The main objective of this paper is to find the most important parameters in estimating OD flows for trucks based on re-identification algorithms, and whether these model parameters are transferable between datasets belonging to different sites. The means and covariance matrices of the mixture models are examined together with the effect of travel time. Different scenarios are tested and it has been found that means are by far the most important parameter in re-identification and OD estimation.**

***Index Terms*- Vehicle Re-identification, Origin-Destination Flow, Bayes Method, Model Transfer, Gaussian Mixture Model**

I. INTRODUCTION

Origin- Destination (OD) estimation is an important problem that arises in transportation modeling. The purpose of OD estimation is to know accurately the number of vehicles that travel between two locations. Knowing truck flows is critical for freight planning and understanding flow patterns in a transportation network. In this paper, the objective is to determine the most important parameters in estimating truck flows between two weigh-in-motion sites, and then determining whether these parameters are transferable.

Ilyas Ustun is a graduate student at Old Dominion University, Department of Modeling & Simulation, Norfolk, VA 23529, USA (phone: 757-683-4521; fax: 757-683-3200; e-mail: ilyasustun@ gmail.com).

Dr. Mecit Cetin is an instructor and researcher with Old Dominion University, Department of Civil Engineering, Norfolk, VA 23529, USA. (e-mail: mcetin@ odu.edu).

The OD estimation is done after vehicle re-identification. Vehicle re-identification pertains to anonymously re-identifying an entity observed at multiple locations based on attribute data. The attribute data comprises of many attributes among which vehicle length, axle spacings, and axle weights are used in this paper. These attributes are collected at weigh-in-motion (WIM) stations.

The reason for using vehicle re-identification methods is the usefulness of the information it provides and comparably low cost. Although there may be mismatches and some errors in the results, vehicle re-identification makes use of the already installed equipment in the roads such as inductive loops, thus installing new equipment is not necessary. Even though there are other techniques to determine/estimate truck flows, such as license plate recognition (LPR) also named Automatic Vehicle Identification (AVI), and Global Positioning System (GPS), these techniques require additional investment and may not be acceptable due to privacy concerns.

Different vehicle attributes have been used for re-identification of vehicles by different researchers such as vehicle length [1], axle spacing [2] and some characteristics of the signal waveform obtained from the sensor set [3]. Various methods incorporating these attributes have been developed for re-identification of vehicles. As explained in [3] blade sensors have been used to create a signature of waveform for each vehicle which is used for lexicographic optimization to re-identify vehicles. Researchers also use vehicle length, electromagnetic signatures, lane, speed, and time to train a decision tree, which is used to perform vehicle re-identification [4].

The methods mentioned above match a downstream vehicle to the most similar upstream vehicle based on some defined metric such as Euclidian distance, Bayesian method etc. The accuracy of these methods depends on the variation of the attribute data from vehicle to vehicle, number of attributes, the distance between data collection sites, variability of travel time, and type of re-identification algorithm used. In this paper by making use of the algorithm developed in [2] the re-identification is accomplished. Then by applying the screening methods developed in [5] OD flow estimation has been done.

The paper is organized into the following sections: Section II will give information about the dataset we have used for the model development. Section III and IV will be about model development of the re-identification of trucks. OD flow estimation has been done in two phases. In the first phase, as described in detail in Section III, for each individual vehicle observed at a downstream point a matched vehicle is found in the upstream site. In the second phase, methods are presented to screen out vehicles that cross downstream but not the upstream site. By doing this we will estimate the OD flow, which is described briefly in Section IV. Section V is about determining the most important parameters in OD estimation. In section VI we analyze whether these parameters are transferable from one link to another. We conclude the paper in Section VII. Future works are addressed in Section VIII.

II. Dataset Used for Model Development

This research is based on datasets that come from several WIM stations in Oregon. Old Dominion University is conducting a joint project with Portland State University to develop new methods to estimate truck flows across the state of Oregon based on the WIM datasets. This paper is partly based on this ongoing project. In Oregon, trucks carrying special transponders are allowed to bypass the weigh stations. Unique identification numbers are registered when these trucks that are equipped with transponders get in contact with the AVI antennas at the weigh stations. These identification numbers when matched between two WIM sites then provide the means to determine the ground truth OD flows.

The vehicle attribute data gathered from the WIM scales are divided into automatic vehicle classification (AVC) and WIM data. This is done because some sites do not have the capability to provide the weights of the axles. AVC data contains the length and the 1-4 axle spacings of the trucks. WIM spans the set of AVC and includes axle weight data for axles 1-5. These data are used to develop the re-identification models. These models essentially match two vehicles based on the similarity of the attribute data (explained in more detail in the subsequent section). The results obtained from the models are compared with the ground truth data obtained by the AVI tag numbers.

For this paper, we used data from three WIM sites as shown in Fig. 1. The Klamath Falls (station 17) WIM station is referred to as upstream station (US) in

TABLE I
NUMBER OF TRUCKS OBSERVED AT THREE STATIONS, OCT 2007

Trucks	Node 17 (KFP)	Node 14 (LWL)	Node 18 (BND)
With Transponders	8,670	3,811	5,956
Without Transponders	16,969	11,590	17,653
Total	25,639	15,401	23,609
% With Transponders	34%	25%	25%

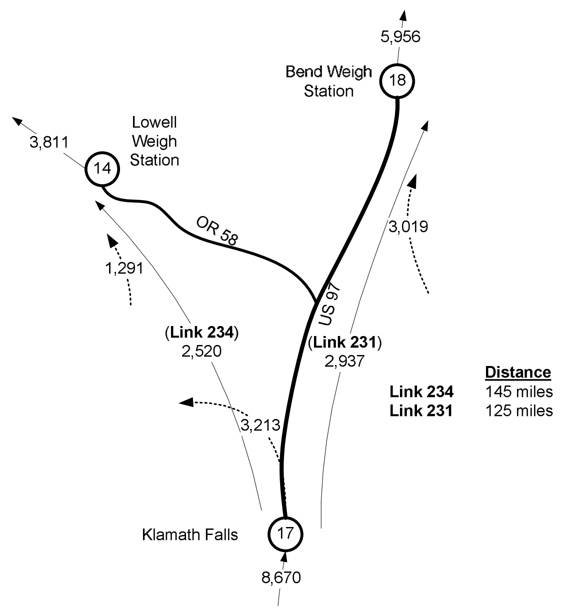


Fig. 1. Number of trucks with transponders for link 234 and 231.

the rest of the paper. There are two downstream sites (DS): Lowell Weigh Station (14) and Bend Weigh Station (18). The link from 17 to 14 is named 234. These two stations are 145 miles apart from each other. The second link is named 231 and is between Klamath Falls (17) and Bend Weigh Station (18) which is 125 miles apart.

The measurements have been taken for a month during October 2007. The system is not “closed” as vehicles can enter and leave the system in between the US and DS, as can be seen in Fig. 1. This makes vehicle re-identification and OD flow estimation tougher compared to a system where all the downstream vehicles also appear in the upstream. Table I shows the total numbers of trucks that pass through each station with and without transponder.

III. The Re-identification Phase

As mentioned in the Introduction, there are two phases in the proposed OD estimation method. First, the re-identification phase is accomplished. Some notations need to be described. Let U and D be two non-empty sets which denote the vehicles at the upstream and downstream WIM stations, and X^U and X^D denote the attribute data collected for upstream and downstream vehicles, X_i^U and X_j^D denote the rows of these matrices for the attributes measured, t_i^U and t_j^D denote the time stamps of each vehicle arrival at respective stations. Then given X^U,X^D, t_i^U and t_j^D vehicle matching problem is to find X_i^U and X_j^D that belong to the same vehicle. Let γ_{ij} be a binary function which equals to 1 if X_i^U and X_j^D belong to the same vehicle and 0 otherwise. The re-identification is going to be accomplished using the Bayesian Method described in [5] where the main objective is to estimate all γ_{ij}'s with minimum error. These are estimated by the Bayesian equation shown in (1).

$$P(\delta_{ij} = 1|x_{ij}) \sim \frac{f(x_{ij}|\gamma_{ij}=1)f(t_{ij})}{f(x_{ij}|\gamma_{ij}=1)f(t_{ij})+\alpha} \qquad (1)$$

Where x_{ij} is a multidimensional vector and represents the difference between X_i^U and X_j^D and α is an arbitrary positive constant. To calculate the probability in (1), $f(x_{ij}|\gamma_{ij} = 1)$ and $f(t_{ij})$ are needed. The first one is a conditional probability density function (pdf) that describes the distribution of x_{ij} when the measurements i and j belong to the same truck. $f(t_{ij})$ is a pdf for travel time information distribution. These pdfs are estimated by statistical mixture models as explained in detail in [5].

The following describes the key steps of the algorithm:

For each vehicle j in D
  Identify a search space $S_j \subset U$
  For each $i \in S_j$
    Calculate $P(\delta_{ij} = 1|x_{ij})$
    $m = \text{argmax}_i\, P(\delta_{ij} = 1|x_{ij})$
    Match vehicle j to m, i.e. $\delta_{ij} = 1$ if $i = m$

For each vehicle in the downstream a match is going to be found in the upstream. To implement the algorithm, first, a search window is identified based on travel time between the two stations. Otherwise, for each DS vehicle the entire dataset for the US need to be searched to find a match. This is both time consuming and may increase the error since the same truck may make more than one trip between the two sites. The time stamp information is available for both upstream and downstream vehicles. Thus, a search space (S_j) for the downstream vehicle j can be determined from the upstream vehicles based on travel times. The minimum travel time can be based on maximum travel speed or on the minimum travel time from the data set. From link 234 data, some trucks have a travel time of 120 minutes. The maximum travel time may have a large variation depending on many situations like weather, rest stops, traffic flow interruptions such as accidents, road maintenance etc. Alternatively, the largest travel time from the data set can be taken. For link 234, there are trucks that have a maximum travelling time of 316 minutes between the stations. Thus, the search space composed of upstream vehicles will have vehicles that have started their journey from 120 to 316 minutes in the past compared to the time stamp of the downstream vehicle. This is illustrated in Fig. 2 and mathematically expressed in (2).

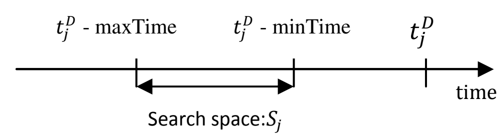


Fig. 2. Search space for each DS vehicle is found by the minimum and maximum travel times.

$$S_j = \{i \in U | t_j^D - maxTime \le t_i^U \le t_j^D - minTime\} \qquad (2)$$

After determining the US vehicles for the search space which are going to be candidates of a possible match for the downstream vehicle, the probability for a match with the downstream vehicle is found by using the Bayesian equation (1). After finding the probabilities for a match, the upstream vehicle that yields the highest probability is said to be the match to the downstream vehicle. This same procedure is followed for each downstream vehicle and a match is found for all of them.

IV. Origin- Destination Estimation

After the first phase is completed, OD estimation can be done. It should be noted that since the system is open meaning that vehicles can enter and leave the system between the upstream and downstream stations, there are three possible results for matching. One for correct matches, one for mismatched vehicles even though the vehicles crossed the upstream, and one for mismatched vehicles in which the vehicles did not appear in the upstream at all. Some screening method should be applied in order to eliminate some of the vehicles that cross the downstream but do not appear in the upstream. Four types of screening methods have been proposed in [5] for the sake of matching vehicles correctly. In these methods different combinations of posterior probabilities are compared with threshold values which are between 0 and 1. This procedure will eliminate the mismatchings according to the screening method and will improve the accuracy. For details, readers are encouraged to read [5].

V. Finding the Most Important Parameters

We have done re-identification and origin destination estimation in our previous works [6]. In this paper we wanted to analyze whether the parameters that are used in the mixture model for one link could be transferred to other links without the need of model training for these new links. Transferability of parameters will result in less effort and time. This will also enable us to re-identify trucks without transponders, because we will not have the chance to train a model for them as the ground truth data does not exist for them.

Total data for link 231 and link 234 have been divided into training and testing sets. The training set is the dataset for the first half of the month of October 2007 and testing set is the second half. The training set consists only of common vehicles which cross both upstream and downstream stations and have transponders. The purpose of training is to get the mixture model components. This is achieved by the EMMIX program developed in [7]. After getting these parameters which are the means, covariance matrix, and proportions for each component we begin our analysis using the testing set which consists of all the trucks with transponders.

We have done our analysis one with 2 components mixture model and one with 3 components mixture model. This is to see whether having more components, or having a higher degree in terms of polynomial fitting, significantly improves the results. Second we have done each analysis one for AVC data and one for WIM data. This is to see whether having more attributes will improve the matching accuracy.

The testing dataset for the downstream of link 231 consists of 3116 vehicles each equipped with a transponder. 1670 of these vehicles cross both WIM station 17 and 18, and 1446 of them cross only downstream WIM station 18. The upstream dataset contains all the trucks regardless of having a transponder or not. Similarly, the testing dataset for the downstream of link 234 consists of 1936 vehicles out of which 1361 cross both upstream and downstream stations, and 575 cross only downstream.

Table II
3 Component Matching Results for AVC and WIM data

	AVC		WIM	
Link	**231**	**234**	**231**	**234**
0	1289	1216	1525	1259
1	381	145	145	102
-1	1446	575	1446	575

Table III
2 Component Matching Results for AVC and WIM data

	AVC		WIM	
Link	**231**	**234**	**231**	**234**
0	1274	1213	1542	1255
1	396	148	128	106
-1	1446	575	1446	575

The results after applying the Bayesian model with each links' own mixture model parameters to the testing dataset are shown in Table II and Table III for 3 component and 2 component mixture models respectively. In Table II we can see that for link 231 AVC data 1289 vehicles have been matched correctly (Group 0), and 381 are mismatched though the vehicles crossed both stations (Group 1), 1446 are mismatched because the vehicles never passed upstream station (Group -1). The other entries for link 231 and 234 can be interpreted similarly.

Then we have created three scenarios in which all of them, link 234 mixture model has been taken as basis, and one parameter is changed at a time of the mixture model. This model is then applied to both link 234 and 231 to see the effect of it. The purpose of these experiments is to find the most important parameter that affects the results, and then using this parameter as a starting point for model transferability between links.

In scenario 1 we have used link 234 means, and link 231 covariance matrix. The travel time (TT) mixture model is from link 234. The new models obtained both for AVC and WIM data are then used in the Bayesian model. The results are shown in Table IV and Table V. The numbers in the parentheses are percentages that show the difference from the true outcomes (group 0) of link 231 and 234's original mixture model application (Table II and Table III). There has been a decrease of 76%

and 64% in correct matching for link 231 AVC and WIM data respectively with 3 component model, but only 2 and 5% for link 234. This shows that mean is more important than other parameters. Interestingly, 2 component model has performed better.

TABLE IV
3 COMPONENTS, LINK 231 COVARIANCE, LINK 234 MEAN AND TT

	AVC		WIM	
Link	**231**	**234**	**231**	**234**
0	308 (-76)	1190(-2)	542 (-64)	1202 (-5)
1	1362	171	1128	159
-1	1446	575	1446	575

TABLE V
2 COMPONENTS, LINK 231 COVARIANCE, LINK 234 MEAN AND TT

	AVC		WIM	
Link	**231**	**234**	**231**	**234**
0	483 (-62)	1098 (-9)	629 (-59)	1200 (-4)
1	1187	263	1041	161
-1	1446	575	1446	575

In scenario 2 we have used link 231 means, and link 234 covariance matrix, and link 234 travel time mixture model. The new models obtained both for AVC and WIM data are then applied and the results are shown in TABLE VI and TABLE VII. From the results we can again say that mean is the most important parameter. The decrease in link 234 correct matching is less than the decrease in link 231 correct matching in scenario 1, implying that travel time mixture model is more important than covariance matrix and less than means.

In scenario 3 we have used link 234 means, link 234 covariance matrix, and link 231 travel time mixture model. The new models obtained both for AVC and WIM data are then applied and the results are shown in TABLE VI and TABLE VII. These results together with previous scenarios confirm that mean is the most important parameter followed by the travel time mixture model. The covariance matrix takes the least significant role.

TABLE VI
3 COMPONENTS, LINK 231 MEAN, LINK 234 COVARIANCE AND TT

	AVC		WIM	
Link	**231**	**234**	**231**	**234**
0	1282 (-1)	742 (-39)	1501 (-2)	955 (-24)
1	388	619	169	406
-1	1446	575	1446	575

TABLE VII
2 COMPONENTS, LINK 231 MEAN, LINK 234 COVARIANCE AND TT

	AVC		WIM	
Link	**231**	**234**	**231**	**234**
0	1253 (-2)	624 (-49)	1498 (-3)	852 (-32)
1	417	737	172	509
-1	1446	575	1446	575

TABLE VIII
3 COMPONENTS, LINK 231 TT, LINK 234 MEAN AND COVARIANCE

	AVC		WIM	
Link	**231**	**234**	**231**	**234**
0	538 (-58)	1004 (-17)	1035 (-32)	1135 (-10)
1	1132	357	635	226
-1	1446	575	1446	575

TABLE IX
2 COMPONENTS, LINK 231 TT, LINK 234 MEAN AND COVARIANCE

	AVC		WIM	
Link	**231**	**234**	**231**	**234**
0	508 (-60)	1011 (-17)	862 (-44)	1131 (-10)
1	1162	350	808	230
-1	1446	575	1446	575

VI. PARAMETER TRANSFERABILITY

Now that we have found the means as the most important parameter in matching vehicles, we are interested in finding whether the means of one link is transferable to the other. By doing this and expanding it to other parameters as well in the future, we will avoid the need of fitting a mixture model to each link. This will save us time and effort. Also, this will make the model applicable to re-identifying vehicles in other places and datasets that may not have transponders at all.

The trick in transferring a model is finding the right method to change the parameters, because no model can be applied directly from one to another. While doing this we have not used the matched vehicle data, as our purpose is to be able to apply the model to non-transponder data as well. Instead we have taken the most common trucks on roads of Oregon which are class 11. By using the same class of trucks at upstream and downstream we want to stay consistent with the measurements. The means have been found for class 11 trucks at the upstream and downstream stations and the difference between downstream and upstream is divided by downstream to get only one value for computations and also to stay consistent with how the mixture model was obtained. Then, these values are used to get new means.

After many experiments we have found one method that gives reasonably low errors. In this method we take the difference of the length, the first axle spacing, and in WIM case also the first axle weight between link 231 and 234. This is due to the fact that there are 3 different signal readers, one for length, one for axle spacing, and one for axle weights. Also, the length, first axle spacing and first axle weight are pretty consistent in class 11 trucks. The differences found above are added to the means of link 234 and the new mixture model is applied to link 231. The results for link 231 are given in TABLE X and TABLE XI. It can be seen that there is very little difference between 2 and 3 component models. The WIM data gives much better results than AVC, showing that using more attributes as inputs yields better results. The accuracy has decreased by 18% for AVC scenarios, and 10 to 12% in WIM scenarios compared to the cases where link 231's own means were used. Thus, it is possible to get reasonable results by adjusting the means and applying them to other links.

TABLE X
3 COMPONENTS, ADJUSTED MEANS APPLIED TO LINK 231

	AVC		WIM	
Link	**231**	**% Change**	**231**	**% Change**
0	1059	-18	1371	-10
1	611		299	
-1	1446		1446	

TABLE XI
2 COMPONENTS, ADJUSTED MEANS APPLIED TO LINK 231

	AVC		WIM	
Link	**231**	**% Change**	**231**	**% Change**
0	1044	-18	1353	-12
1	626		317	
-1	1446		1446	

VII. CONCLUSIONS

This paper demonstrated that trucks crossing two WIM sites can be matched effectively by a Bayesian re-identification method. The results show that there is very little difference between 2 and 3 component models. Thus, having more components is not necessarily better, at least for these two links. All of the results indicate that having more attributes as inputs to the model will increase accuracy significantly. Based on the results of the three scenarios, it is found that having accurate means (for the mixture models) is more critical than having accurate covariance matrix and travel time distribution. The results show that having the correct means and wrong covariance and TT, the correct matching differs by only 1 to 3% for link 231 in scenario 2. This implies that only having the correct means will yield quite reasonable results.

Having determined the most important parameter, we have also found that this parameter can be transferable. The results show that by adjusting only the means, reasonable accuracy (within 10 to 18% of the actual outcome) can be reached. It is no doubt that better results can be achieved if other parameters are adjusted as well. Transferability of parameters implies less model training and effort for other links equipped with or without transponders.

VIII. FUTURE WORK

Current study focused on a small network having one upstream and two downstream stations. Future work will include more stations and thus forming a greater network. This will increase the confidence in the outcomes. Analyzing 4, 5, 6, 7 component models for all the links in the network will provide the optimum number of components to use, and may also lead to some criteria of how many components to use for different type of links. Also, adjusting the travel time mixture model and covariance matrix and making them transferable between the links will certainly increase the accuracy of matching.

IX. ACKNOWLEDGMENT

The authors would like to thank Oregon Department of Transportation Motor Carrier Division for providing the data.

X. REFERENCES

[1] B. Coifman and M. Cassidy, "Vehicle reidentification and travel time measurement on congested freeways," *Transportation Research, Part A (Policy and Practice),* vol. 36A, pp. 899-917, 2002.

[2] M. Cetin and A. P. Nichols, "Improving the accuracy of vehicle re-identification by solving the assignment problem," *Transportation Research Board: Journal of the Transportation Research Board, 2129,* pp. 1-8, 2009.

[3] C. Sun*, et al.*, "Use of vehicle signature analysis and lexicographic optimization for vehicle re-identification on freeways," *Transportation Research Part C (Emerging Technologies),* vol. 7C, pp. 167-85, 1999.

[4] A. Y. Tawfik*, et al.*, "Using decision trees to improve the accuracy of vehicle signature re-identification," 2004, pp. 24-33.

[5] M. Cetin*, et al.*, "Bayesian models for re-identification of trucks over long distances based on axle measurement data," *Journal of Intelligent Transportation Systems,* vol. 15, issue 1, pp. 1-12, 2011.

[6] M. Cetin*, et al.*, "Investigating the key factors affecting the accuracy of re-identification of trucks over long distances based on axle measurement data," *Transportation Research Record,* 2011 submitted.

[7] G. McLachlan and D. Peel, Eds., *Finite Mixture Models*. New York, 2000.

Vehicle Re-identification Using Neural Networks

Tanweer Rashid, Mecit Cetin.

***Abstract* – Vehicle re-identification is an important research topic in transportation. This research effort investigates the use of neural networks for solving the re-identification problem. The neural network used is a multilayer perceptron with 1 hidden layer. The training of the neural network is a supervised learning procedure, and is achieved using the Back-propagation algorithm. The neural network makes use of vehicle travel time, length, axle spacing and axle weight as inputs for classification, and tries to determine whether a pair of vehicle attributes belongs to the same vehicle. The neural network model developed is then compared to a previously published model (a Bayesian model), and it is found that the neural network model's performance is slightly lower than that of the Bayesian model.**

***Index Terms* – Vehicle re-identification, neural networks, multilayer perceptron**

I. INTRODUCTION

VEHICLE re-identification is the process by which vehicle attributes measured at one point on a road network are compared to vehicle attributes measured at another point. A match is made if the two measurements belong to the same vehicle. The attributes being measured can be actual physical characteristics of the vehicle such as length [1] and axle weights and spacing between axles [2]. Other measurements can be the characteristics extracted from the waveforms generated by inductive loop detectors [3].

In general, most vehicle re-identification systems have three parts. The first part is concerned with collection of vehicular data. Usually, researchers have made use of inductive loop detectors (ILDs). These can be single loops or dual loops [3-7]. In other cases, researchers have used different, more advanced detectors like Weigh-In-Motion (WIM) [2, 8-10]. The second part deals with the data itself. ILDs generate a waveform that is unique to vehicles, and researchers employ feature extraction techniques to extract meaningful data from these 'signatures'. In other cases, researchers make use of actual physical attributes of vehicles like vehicle length [1], weight, axle weights and axle spacing [2]. The last part of the vehicle re-identification system deals with the algorithm(s) used for the re-identification purpose. Researchers have introduced lexicographic-based techniques [3], decision trees [6] and Bayesian probability-based methods using mixture models [9, 10] for the re-identification task.

Tanweer Rashid is a graduate student at the Department of Modeling, Simulation and Visualization Engineering, Old Dominion University, Norfolk, VA 23529, USA. (e-mail: trash001@ odu.edu)

Dr. Mecit Cetin is an instructor and researcher with Old Dominion University in the Department of Civil Engineering, Norfolk, VA 23529, USA. (e-mail: mcetin@ odu.edu).

The most obvious benefit of a vehicle re-identification system is that it will allow for estimating origin-destination flows. Other benefits include estimation of travel delays and travel times. Another important aspect to consider is the fact that the proposed method of vehicle re-identification (as well as previous methods established by other researchers) allows for anonymity. This goes a long way towards avoiding privacy-related issues, as well as avoiding the need to install specialized tracking IDs.

This paper is divided into five sections. Section I gives an introduction to vehicle re-identification. Section II discusses the Back-propagation algorithm very briefly, and then describes the neural network used for re-identification. Section III discusses the re-identification process utilized in this research. Section IV discusses the results of the neural network models, and Sections V discusses some limitations and concludes this research paper.

II. NEURAL NETWORKS

The neural network used is the multilayer perceptron with 1 hidden layer (Figure 1). This type of neural network is trained in a supervised manner, i.e. the desired outputs for corresponding inputs are provided to the neural network during training. The neurons (also called units) are interconnected by weighted values (between -1 and +1), and these weights represent knowledge. Learning is achieved by adjusting the values of the weights. In the beginning, all weights are assigned random values between -1 and +1. During the training process, these weights are adjusted.

The training process is performed using the back-propagation algorithm. This algorithm can be mathematically summarized as a generalization of the delta rule given by Equation (1):

$$\Delta w_{jk} = \gamma \mathrm{y_j}(\mathrm{d_k} - \mathrm{y_k}) \tag{1}$$

In the above equation, the j^{th} neuron or unit provides input to the k^{th} unit. d_k is the desired value, and is provided in a supervised manner, w_{jk} is the weight between the j^{th} unit and the k^{th} unit, γ is the learning rate. In the end, the back-propagation algorithm can be expressed by Equations (2) and (3).

$$\delta_o^P = (d_o^P - y_o^P)\, F'(s_k^P) \tag{2}$$

δ_o^P is the error signal from the o^{th} output layer. It is computed using the desired output for the P^{th} pattern d_o^P and the output of the o^{th} output y_o^P.

$$\delta_h^P = F'(s_h^P) \sum_{o=1}^{N_o} \delta_o^P\, w_{ho} \tag{3}$$

In Equation (3), the delta value δ_h^P for the hidden layer h is being computed using the error signal from the hidden layer δ_o^P. P is the P^{th} input pattern to the neural network.

The multilayer perceptron used is shown in Figure 1. The output layer has one neuron that gives a value between 0 and 1. There are 11 neurons in the input layer, and the hidden layer has a variable size. "SPC 1" to "SPC 4" in the figure represents axle spacing 1 to axle spacing 4, respectively. "AXL 1" to "AXL 5" represents the weight of axle 1 to axle 5, respectively. The size of the hidden layer is variable, given by n.

The size of the hidden layer is an important factor, and needs to be determined experimentally. If the hidden layer has too few neurons, then the neural network will not be able to learn and generalize properly. On the other hand, if the hidden layer has too many neurons, then the neural network will suffer from over-learning, and will perform poorly.

The vehicle re-identification problem can be categorized as a 2 class problem. The neural network system will identify vehicles that either match or do not match. Therefore, the output of the neural network can be something as simple as 1 and 0: 1 indicating a match and 0 indicating no match. However, in most cases, it will be seen that the neural network does not give outputs as integers but as continuous values like 0.9523247 or 0.001258745. The easiest and most logical thing to do is to round up the output values to the nearest integer. Output values greater than or equal to 0.5 will be rounded to 1 and output values less than 0.5 will be rounded to 0. This threshold value of 0.5 will be used throughout the experiments to make a separation between Matching and No Matching vehicles.

III. VEHICLE RE-IDENTIFICATION PROCESS

A large part of the dataset used in this research comes from the Green Light project. Green Light is a preclearance/prescreening program started in 1995 and is sponsored by the Oregon Department of Transportation (ODOT). It is used by commercial vehicles to bypass weigh in stations at highway speeds. Through Green Light, Oregon installed twenty two stations that include weigh-in-motion (WIM) devices and automatic vehicle classification (AVI). Figure 2 shows the locations of the stations depicted by the white circle with a number. When Commercial vehicles approach a station, they are weighed and checked for height violations. Vehicles with transponders, or tags, are read by an overhead reader. Observations at the stations consist of the number of axles, axle weights and spacing, as well as speed, time of observation, lane, vehicle length, gross vehicle weight. More details of the Oregon WIM system can be found in [11].

As in typical vehicle re-identification procedures, a vehicle from a downstream data collection point is compared against a set of vehicles from an upstream data collection point. These collection points can be as low as one 1800 ft apart, or as far as 214 mi. A link is defined as the path between two data collection points. The information in Table 1 shows the four links that were used in this research, and

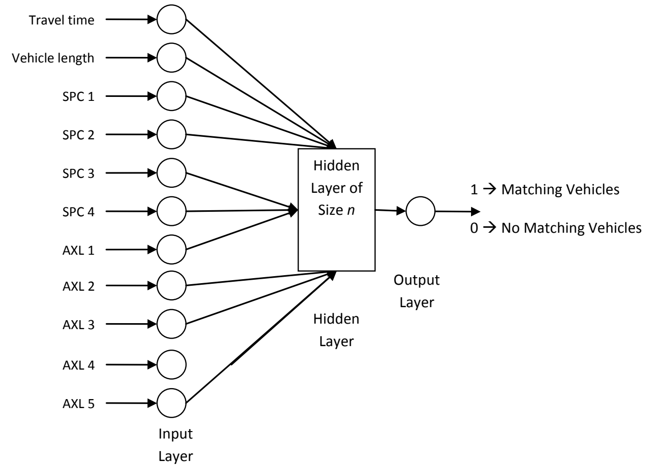


Figure 1: The multilayer perceptron used for re-identification.

their respective upstream and downstream data collection points, distance between data collection points and the links' minimum and maximum travel times.

Since vehicles rarely travel at the same speeds, it would be impractical to assume an average travel time between stations. Some vehicles might travel at a slower speed, while others might be travelling at higher speeds, and still other vehicles might stop at a rest area for some period of time. Therefore, it is more practical to use a time window (with a minimum travel time and a maximum travel time) to find out how many vehicles traverse any particular link. The minimum travel time (*minTime* in Table 1) for any particular link can be easily calculated based on an assumed maximum travel speed. In this research, the maximum travel speed is assumed to be 70 mph. The maximum travel time (*maxTime* in Table 1) can be taken as a multiple of the minimum time, in this case, the multiple is assumed to be 2.5. The maximum travel speed and multiplication factor are not chosen arbitrarily, and these values (previously used in [10]) allow for a large majority of the vehicles to be encompassed in the upstream station time window.

Training and testing data were created from the four links shown in Table 1. The whole dataset consisted of sensor readings from October 2007. Readings from 1st October, 2007 to 15th October, 2007 were used to create training data, and the remaining days of October 2007 were used to create testing data. As mentioned previously, the training and testing data consisted of vehicle travel time, length, axle spacing and axle weights. The mean and variance for each attribute in the testing dataset was computed, and then the testing and training data were normalized using the said mean and variances according to Equation (4).

$$\bar{\bar{X}}_i^{Attribute} = \frac{X_i^{Attribute} - \mu^{Attribute}}{\sigma^{Attribute}} \qquad (4)$$

In Equation (4), the superscript *Attribute* is the vehicle attribute like travel time, or length, or axle spacing or axle weight. X_i represents the data for the i^{th} row in the training or testing dataset. $\mu^{Attribute}$ and $\sigma^{Attribute}$ are the mean and variances of the respective vehicle attributes.

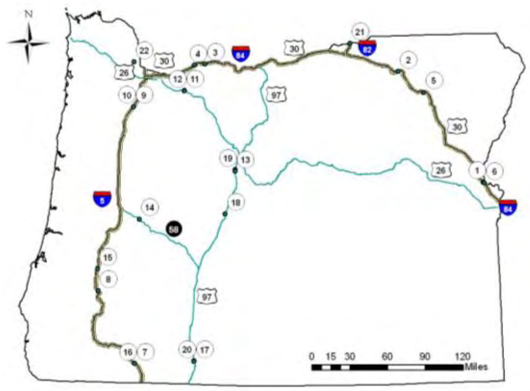


Figure 2: Location of WIM sites in Oregon.

Since some of the vehicles have IDs that are detected and recorded by the data collection stations, it is possible to know beforehand which row of the training and testing dataset is for matching vehicles.

IV. RESULTS AND DISCUSSIONS

An important issue for discussion is the meaning of the results of the neural network. There are four possible output scenarios:

1. Scenario 1: The system indicates a match when, in fact, there is a match.
2. Scenario 2: The system does not indicate a match but, in fact, there is a match.
3. Scenario 3: The system indicates a match but, in fact, there is no match.
4. Scenario 4: The system does not indicate a match when, in fact, there is no match.

The terms used to describe the above four scenarios are borrowed from Signal Detection theory formalized by Tanner, Swets and Green [12, 13]. The term "Hit" is used to describe Scenario 1. The term "Miss" is used to describe Scenario 2. "False Alarm" or (FA) is used to describe Scenario 3. "Correct Rejection" or (CR) is used to describe Scenario 4. These four terms will be used throughout the remainder of this document to describe the results of the neural networks. The objective is to have as many Hits as possible, while at the same time, keep the FAs as low as possible.

As mentioned before, the training and testing datasets consist of vehicle travel time, length, axle spacing and axle weights as input. The neural network used has 1 hidden layer, and the size of the hidden layer was experimentally determined. For the same training and testing dataset from Link 234, numerous neural networks with varying hidden layer sizes were trained and tested, and it was found that a hidden layer size of 25 gave the highest Hits and lowest FAs. Using this hidden layer size, five neural networks were trained separately for Links 229, 231, 234 and 237. The reason for multiple training and testing is that since the weights of neural networks are randomly initialized in the beginning, no two neural networks will give the same exact results. For each link, the training dataset was created from date range 1st October 2007 to 15th October 2007, and for

Table 1: Links and their upstream and downstream stations, separation distance and travel times.

Link	Up-stream Station	Down-stream Station	Distance (miles)	***MinTime*** **(minutes)**	***MaxTime*** **(minutes)**
229	14	9	103	88.29	220.71
231	17	18	125	107.14	267.86
234	17	14	145	124.29	310.71
237	19	12	89.8	76.97	192.43

testing, the remaining days of October 2007 was used.
Table 2 shows the results of testing the five neural networks for the four links. The columns 1 to 5 in the table represent five different neural networks trained for each link. The rightmost column shows the average Hit as a percentage of the total actual matches. The total actual matches can be found by summing the Hit and Miss for each neural network.

The vehicle re-identification process developed in [9, 10] is used as a baseline to compare the results obtained in this research. Figure 3, Figure 4, Figure 5 and Figure 6 show the error between the neural network models, and the Bayesian model, for Links 229, 231, 234 and 237, respectively. The horizontal axis shows the number of matched vehicles as a percentage of total common vehicles. The vertical axis shows the mismatched vehicles (i.e. the FAs) as a percentage of the total vehicles matched. In mathematical terms, the values of the horizontal axis were calculated using Equation (5). The values for the vertical axis were calculated using Equation (6).

$$\frac{Hit+FA}{Total\ Actual\ Matching\ Vehicles} \times 100\% \qquad (5)$$

$$\frac{FA}{Total\ Matches\ indicated\ by\ the\ model\ (Hit+FA)} \times 100\% \qquad (6)$$

In the last part of Section II, it was mentioned that the default threshold value used for screening matches and no matches is 0.5. In this comparison, the threshold value is changed from 0.0 to 1.0, and their corresponding matches and errors are plotted in Figure 3, Figure 4, Figure 5, and Figure 6. As the threshold is increased, the total number of vehicles being matched decreases.

In the figure for Link 229 (Figure 3), when the system is matching 90% of vehicles, the error in the Bayesian model is around 25%. For the neural network models, the error for 90% match on the horizontal axis is around 30%. This approximately 5% difference can be seen for nearly all percentage of matches. A spike can be seen to occur in the upper left corner of the graph, and this is due to the fact that, for example, the system indicates a total of five matches and out of those five only one is a real match with the remaining four being FAs.

In the figure for Link 231 (Figure 4), when the system is matching 90% of vehicles, the error in the Bayesian model is around 6%. For the neural network models, the error for 90%

match on the horizontal axis is somewhere around 12%. At 100% on the horizontal axis, the error for the Bayesian model is about 10%, and 16% for the neural network models.

In the figure for Link 234 (Figure 5), when the system is matching 90% of vehicles, the error in the Bayesian model is around 11%, and for the neural network models, the error is around 16%.

In the figure for Link 237 (Figure 6), the neural network models seem to be outperforming the Bayesian model. This difference in performance is obvious, especially for values above 60% on the horizontal axis. The reason why the performance for Link 237 is better than the other links can be due to data cleanliness: the data in Link 237 is much cleaner than the other links.

V. CONCLUSION

This research tackled the vehicle re-identification problem by using neural networks. The dataset used in research was obtained from WIM stations that were installed by Oregon DOT’s Green Light project. In the majority of cases, training data was created from date range 1st October, 2007 to 15th October, 2007. Testing data was created from the remaining days of October 2007.

An optimal size for the hidden layer of the neural networks was determined experimentally, and this hidden layer size was then used to train and test separate neural networks for Links 229, 231, 234 and 237. The results of the neural networks were compared to the results of a Bayesian model. The comparison shows that the performance of the neural networks lags the Bayesian model’s performance by about 5% to 10%.

One limitation is that the tagged vehicles in the training and testing datasets were all trucks. So the models developed in this research are only valid when used for re-identifying trucks. Personal vehicles such as cars, SUVs and trailers are not considered here.

Another limitation is that the default threshold value for separating Matches and No Matches in the neural networks’ output is 0.5. It would be more appropriate to determine an optimal threshold value, based on the concepts of Signal Detection Theory (SDT).

REFERENCES

[1] B. Coifman and M. Cassidy, "Vehicle reidentification and travel time measurement on congested freeways," Transportation Research, Part A (Policy and Practice), vol. 36A, pp. 899-917, 2002.

[2] A. P. Nichols and M. Cetin, "Numerical characterization of gross vehicle weight distributions from weigh-in-motion data," Transportation Research Record, pp. 148-154, 2007.

[3] C. Sun, S. G. Ritchie, K. Tsai, and R. Jayakrishnan, "Use of vehicle signature analysis and lexicographic optimization for vehicle reidentification on freeways," Transportation Research Part C: Emerging Technologies, vol. 7, pp. 167-185, 1999.

[4] B. Coifman and B. Banerjee, "Vehicle reidentification and travel time measurement on freeways using single loop detectors-from free flow through the onset of congestion," Cambridge, MA, United states, 2002, pp. 40-47.

[5] B. Coifman, "Estimating travel times and vehicle trajectories on freeways using dual loop detectors," Transportation Research Part A: Policy and Practice, vol. 36, pp. 351-364, 2002.

[6] A. Y. Tawfik, B. Abdulhai, A. Peng, and S. M. Tabib, "Using decision trees to improve the accuracy of vehicle signature reidentification," Transportation Research Record, pp. 24-33, 2004.

[7] C. C. Sun, G. S. Arr, R. P. Ramachandran, and S. G. Ritchie, "Vehicle reidentification using multidetector fusion," IEEE Transactions on Intelligent Transportation Systems, vol. 5, pp. 155-164, 2004.

[8] C. Monsere, M. Cetin, and A. P. Nichols, "Exploratory Methods for Truck Re-Identification in a Statewide Network Based on Axle Weight and Axle Spacing Data to Enhance Freight Metrics," To be published by Oregon Transportation Research and Education Consortium (OTREC), Portland, OR.2010.

[9] M. Cetin, A. P. Nichols, and C. M. Monsere, "Bayesian Models for Re-identification of Trucks over Long Distances Based on Axle Measurement Data " Journal of Intelligent Transportation Systems, vol. Submitted, 2010.

[10] M. Cetin, C. M. Monsere, A. P. Nichols, and I. Ustun, "Investigating the key factors affecting the accuracy of re-identification of trucks over long distances based on axle measurement data," Submitted to the 90th Annual TRB meeting for presentation and to Transportation Research Record for publication, 2010.

[11] ODOT. (2011, February 05, 2011). Motor Carrier Transportation Green Light Weigh Station Preclearance. Available: http://www.oregon.gov/ODOT/MCT/GREEN.shtml

[12] D. M. Green and J. A. Swets, Signal detection theory and psychophysics. Huntington, N.Y.: R.E. Krieger Pub. Co., 1974.

[13] W. P. Tanner and J. A. Swets, "A decision-making theory of visual detection," Psychological Review, vol. 6, pp. 401-409, 1954.

Table 2: Neural network results for Link 229, 231, 234 and 237.

Link 229								
Trials	**1**	**2**	**3**	**4**	**5**	**AVG**	**STD DEV**	**% Avg Hit**
Hit	532	521	488	498	541	516	22	66
Miss	251	262	295	285	242	267	22	
FA	308	282	238	249	289	273	29	
CR	391660	391686	391730	391719	391679	391695	29	
Link 231								
Trials	**1**	**2**	**3**	**4**	**5**	**AVG**	**STD DEV**	**% Avg Hit**
Hit	1330	1321	1328	1317	1306	1320	10	86
Miss	200	209	202	213	224	210	10	
FA	404	311	312	283	273	317	52	
CR	134187	134280	134279	134308	134318	134274	52	
Link 234								
Trials	**1**	**2**	**3**	**4**	**5**	**AVG**	**STD DEV**	**% Avg Hit**
Hit	1138	1129	1162	1099	1145	1135	23	83
Miss	223	232	199	262	216	226	23	
FA	280	271	335	242	331	292	40	
CR	89506	89515	89451	89544	89455	89494	40	
Link 237								
Trials	**1**	**2**	**3**	**4**	**5**	**AVG**	**STD DEV**	**% Avg Hit**
Hit	678	648	623	605	639	639	27	74
Miss	183	213	238	256	222	222	27	
FA	113	94	104	88	83	96	12	
CR	49552	49571	49561	49577	49582	49569	12	

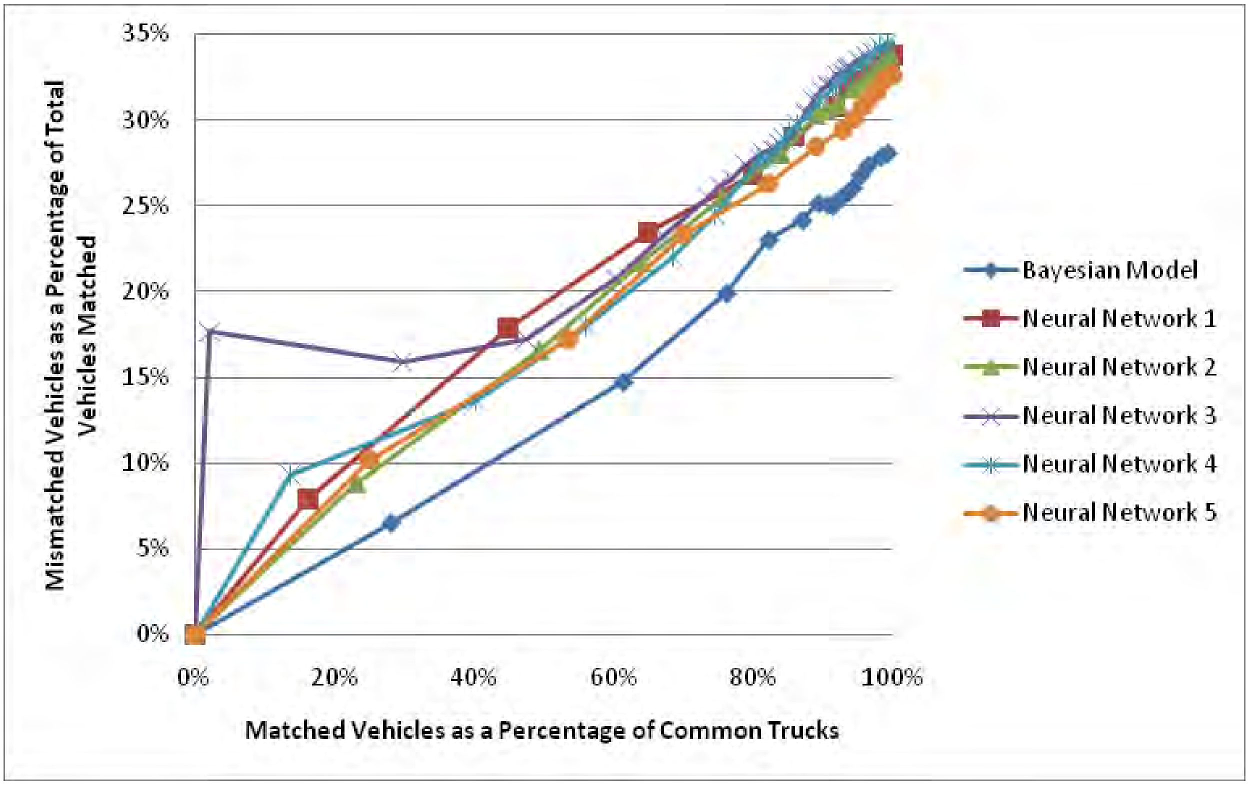


Figure 3: Comparison between Bayesian model and neural network models for Link 229 for WIM dataset.

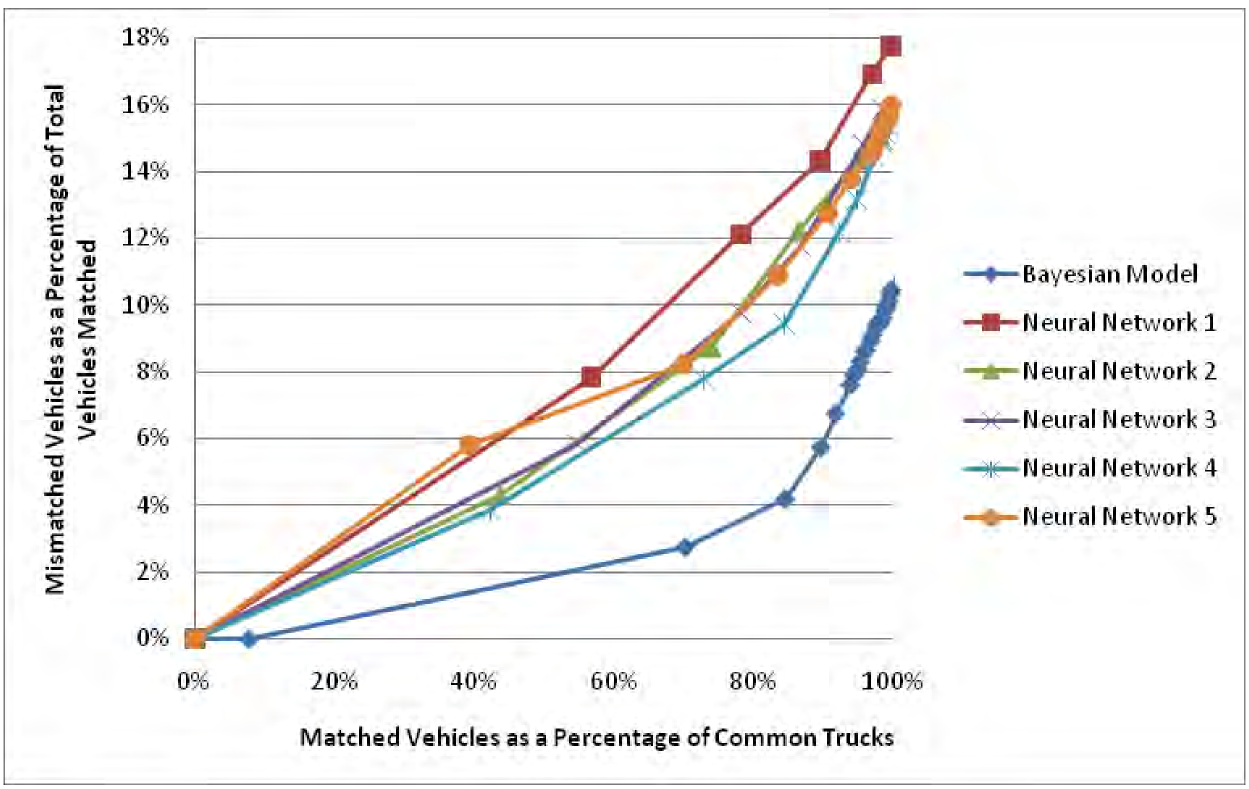


Figure 4: Comparison between Bayesian model and neural network models for Link 231 for WIM dataset.

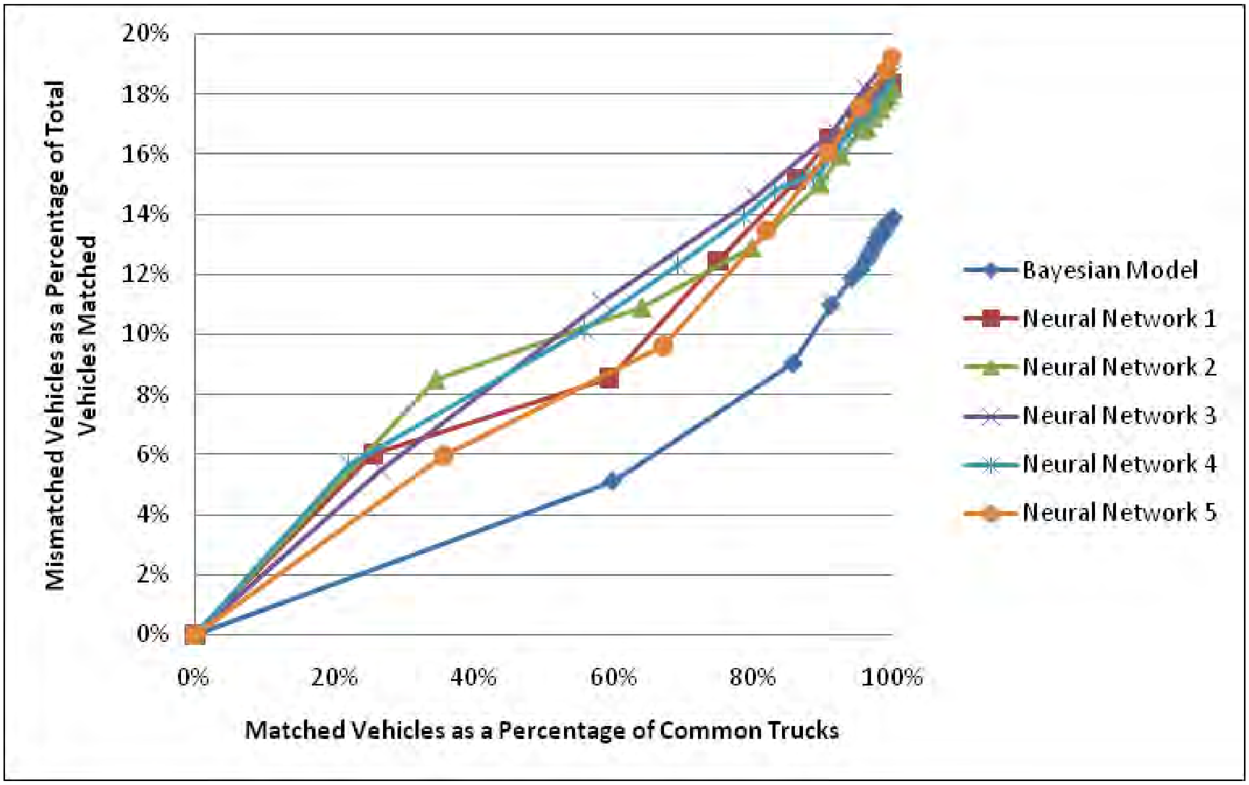


Figure 5: Comparison between Bayesian model and neural network models for Link 234 for WIM dataset.

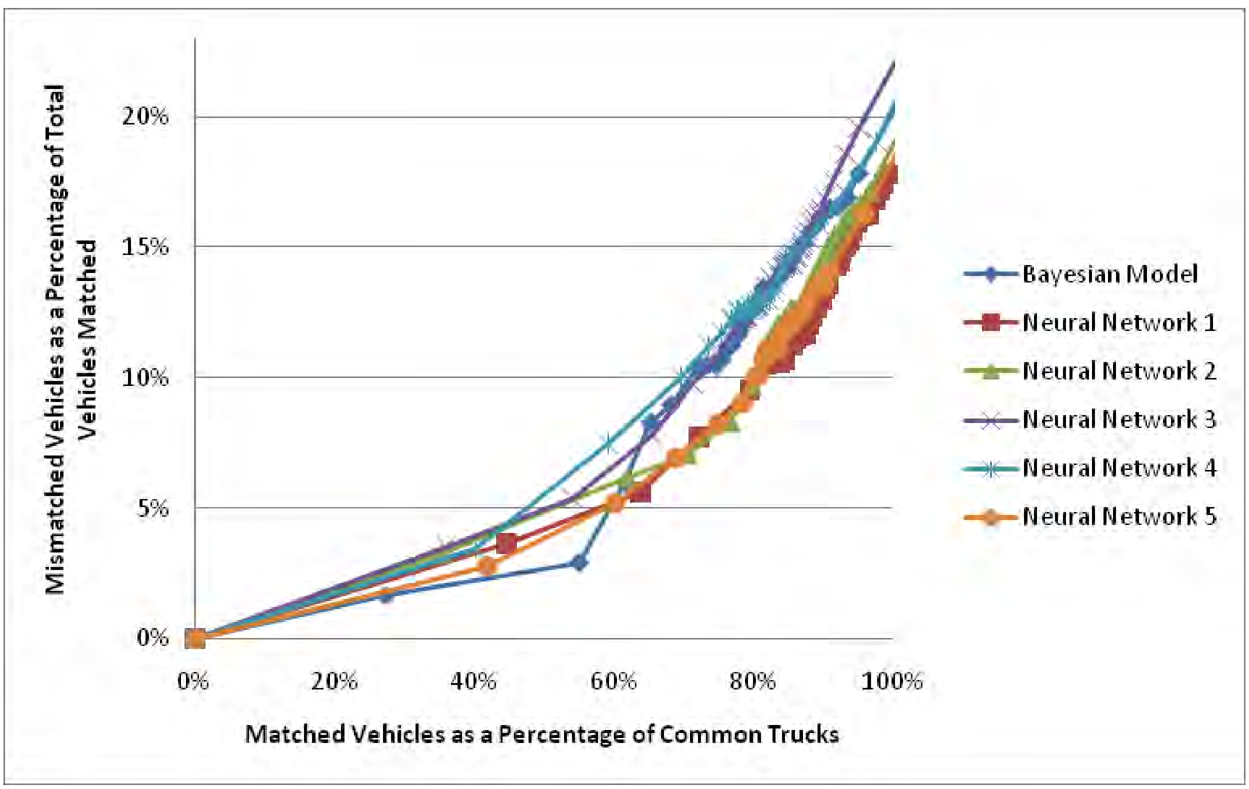


Figure 6: Comparison between Bayesian model and neural network models for Link 237 for WIM dataset.

Homeland Security / Military

Track Lead: Dr. Barry Ezell, VMASC

Judges: Mr. Richard Flannery, Hampton Roads Planning District Commission, Mr. Bill Allen, ALATEC, Inc, Mrs. Terrie Suit, Assistant to the Governor for Commonwealth Preparedness

An Analysis of Enterprise Architectures Using General Systems Theory

Author(s): Christine Hoyland and Andreas Tolk

Abstract Only

3rd Place Gene Newman Award Winner, Homeland Security/Military Track

Exploring the use of computer simulations in unraveling research and development governance issues

Author(s): Mariusz Balaban and Patrick Hester

2nd Place Gene Newman Award Winner, Homeland Security/Military Track

Dismounted Situational Awareness

Author(s): Robert Kewley, and Cadets from West Point

A Schema for Transitioning from Rectilinear Coordinate Spaces to Pseudo-Hexagonal Spaces

Author(s): Alex Hoover

1st Place Gene Newman Award Winner, Homeland Security/Military Track

An Analysis of Enterprise Architectures Using General Systems Theory

Author(s): Christine Hoyland and Andreas Tolk

Keywords: Enterprise Architecture, Systems Theory, Hierarchy, Emergence, Holism

Abstract: Despite numerous attempts to improve methods for developing the enterprise architectures (EA) used during complex DoD acquisition projects, there is general consensus from project managers that the architecture modeling process is costly and returns less that cost-effective benefits. A recent proposal from within the DoD to incorporate Semantic Web techniques into the process promises to resolve identified EA issues. The purpose of this study is to examine the proposed improvements using the Zachman Framework, a luminary for referring to viewpoints of the EA structure, and systems principles as a qualitative measure of effectiveness. General system theory is concerned with the wholeness of systems order not understandable by investigation of their respective parts in isolation. Conclusions of this analysis show that proposed changes should yield some improvement in the area of EA reusability, however, the new methods fall short of incorporating major system principle aspects of hierarchy, emergence and holism.

Dismounted Situational Awareness

Describe the Problem

Shoot, move, communicate. These are the critical actions a dismounted combat unit must be able to perform in order to be effective on the battlefield. Of these three actions, communication is the most important because it precedes and directs how a dismounted unit shoots and moves. Communication is also key in coordinating combat and logistics support. Unfortunately, the current communications systems in use are much more effective for mounted units than dismounted units. Our mission is to provide analysis that will inform and direct the decisions of Army leadership as it provides new communication systems to all dismounted units which modernize and improve military communications. Some ideas that are being considered include tablet PC or smart phones. Some issues with these commercial devices are that the army would have to coordinate with the companies themselves and secure the devices.

How is the Problem Solved Today?

Currently in theater they use a SINCGARS (ASIP) radio which has extremely limited capabilities. It currently has the ability to talk over different radio nets, but at the same time is limited to voice. The stakeholders we talked to seemed to be content with the way the system worked and said that in a firefight they would "forget about the fancy technology and resort to the way they had always done things". The way they had always done things would refer to using a radio to communicate and carry out the mission. The SINCGARS radio is effective, but the question is, "How much more effective an alternative system would be in comparison to what is currently being used in theater?"

Stakeholder Analysis

Army SFC: He has spent over 15 years enlisted in the infantry. He was adamant about the army in general not knowing how to use the technology they already had and how adding new technology would not be helpful. He said that only the PSG/PL and up would need the new technology we are recommending as anything below this would take far too long to train and it would not be helpful in any manner. He liked all of the operational activities that we have and would be interested to see a piece of technology that could use them relatively easily. He said the cost of anything we could come up with would make it irrelevant as compared to what they currently have, to say the least he is doubtful. In addition, he was doubtful on the idea of us translating between languages with our technology, particularly when working with the Afghanistan Police.

Army Captain- deployed to Iraq 2006-2007 with the 101st Airborne Division (Air Assault). He picked up a platoon mid tour and was immediately thrown into action. He was only deployed for five months, but was wounded three times. These consisted of two IEDs and one ambush. He wished he had some sort of technology that tracked his unit's actions and enemy actions in the area because he believed the enemy had operated in the area previously.

Army 1LT: We asked him if he or any of his Soldiers had used the Nett Warrior system before. He had not personally used Nett Warrior, but his platoon sergeant had in 2008 at Ft. Benning during a training exercise. The feedback he gave was that it was too heavy for the benefits it gave his unit. He continued saying that he would not bring it outside the wire with him.

Army SFC: He has served in the Ranger Regiment and participated in several deployments. His main focus was ensuring that whatever technology was being considered as a candidate solution should be easy to train soldiers to use. He said one of the largest issues he had noticed is that soldiers often do not know how to properly operate the communications equipment they currently have. If soldiers do not know how to properly use the technology, it is essentially useless. He used communications technology in the Ranger Regiment similar to some of the systems being examined by our capstone group. This equipment provided video feed, GPS, and text capabilities. He warned that these systems are often very expensive, and may not always provide enough value for that cost. During operations, he said most of the rangers equipped with these devices did not use them. He advised that we find something durable, that is easy to use/train to use, and provides an array of capabilities such as GPS and text communications that would benefit a soldier only at times when not engaged with the enemy. He also advised that no one below the level of squad leader should receive a system.

Army SFC: He has served 12 years in the Army through multiple deployments. He said the systems would be useful in dismounted command and control, but in order for our system to be useful to the soldier it would need to be durable. He commented that if it broke every time a soldier dropped it, the soldier would not carry it. He also noted that the batteries used in this system would need to last at least 12 hours. The batteries would need to be small and light in order to lessen the amount of weight carried by a soldier on a mission. Furthermore, this system would need to have various layers of security to protect the information on it if it were lost or stolen. Otherwise, this system would be too much of a security risk to take on missions. He stated that this system should be able to support voice and text communications, it should be light, using it should be easy, and it should facilitate the request and coordination of combat support assets.

Architecture (CV-1, CV-2, CV-4, OV-1, OV-5a)

The Capability View 1 (CV-1) is part of defining the capabilities we want for our system. This view is the highest level view and shows the overall vision of the system. Our vision for this system is to provide dismounted command and control to different unit types giving the units better situational awareness and communication capabilities. Also in this view we display the desired effects or the organizational objectives. The desired effects are more specific than the vision; however they feed into the vision of the overall system. After talking to stakeholders, we found that we have four capabilities that define our system: communicate with higher and adjacent units, understand friendly situation, understand enemy situation, and simplify and expedite reporting/requests.

The Capability View 2 (CV-2) is part of the system architecture that defines each capability. The capabilities of this system are communicate with higher and adjacent units, understand friendly situation, understand enemy situation, and simplify and expedite reporting/requests. The most important fact about each of these capabilities is that they are all looked at the platoon level.

The Capabilities View 4 (CV-4) is part of the system architecture that defines how the capabilities interact with each other. The capabilities of this system fall under two joint capability areas that are defined by J7 Joint Force Development and Integration Division (JFDID). They are joint capability area command and control and joint capability area battle space awareness.

The Operational View 1 (OV-1) is a visual representation of the capabilities we have found to be essential through stakeholder analysis. Not all capabilities or mission threads are shown in the OV-1, however, the concept of the capabilities are captured. The benefit of the OV-1 is that it shows what capabilities a candidate solution technology may possess but also shows the capabilities necessary to a dismounted Army mission. A prominent take-away from the OV-1 is the level of connectivity that the dismounted platoon or soldier will experience when using the candidate solution technology. They will be connected to UAV, aircraft, other dismounted soldiers, vehicles, mortars/artillery, GPS, and higher headquarters.

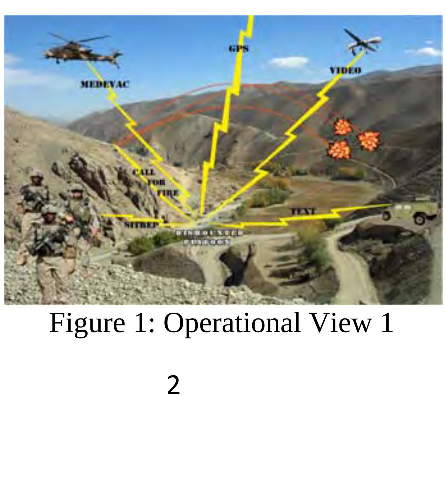


Figure 1: Operational View 1

The Operational View 5a (OV-5a) covers our system capabilities, operational activities, and system functions. It is using the same system capabilities as shown in the CV-2 and CV-4 above. The biggest part to hit with the OV-5a are the Operational Activities, Communicate with Allied Forces, Acquire Video Feed, Send SITREP, Request MEDEVAC, Request EOD, Track Allied Forces, Request for Fire, and Conduct Movement during the Mission. From these activities we decided on functions we would want our system to perform and these will further be broken down in our Value Hierarchy.

Operational Activities (Mission Threads)

We created a mission thread for each of our Operational Activities mentioned in Architecture OV-5a. These Mission Threads go through each of our operational activities step by step to show the different steps that our system is going to be needed for. We need to make sure that the software has the capability to perform all of our desired effects. Most of the threads make use of an RTO (Radio Telephone Operator) which is currently how reports are done on the SINGARs radio. If the system we choose at the end of this project is simple and quick enough, the PL or PSG may be able to perform the role of the RTO and save time instead of translating to the RTO to send messages. Another point of these is to show things that you could not do with the current system, such as requesting a video feed from a UAV. With these mission threads we can show what our system is going to need to do and give us a framework for where to go with our software and hardware requirements.

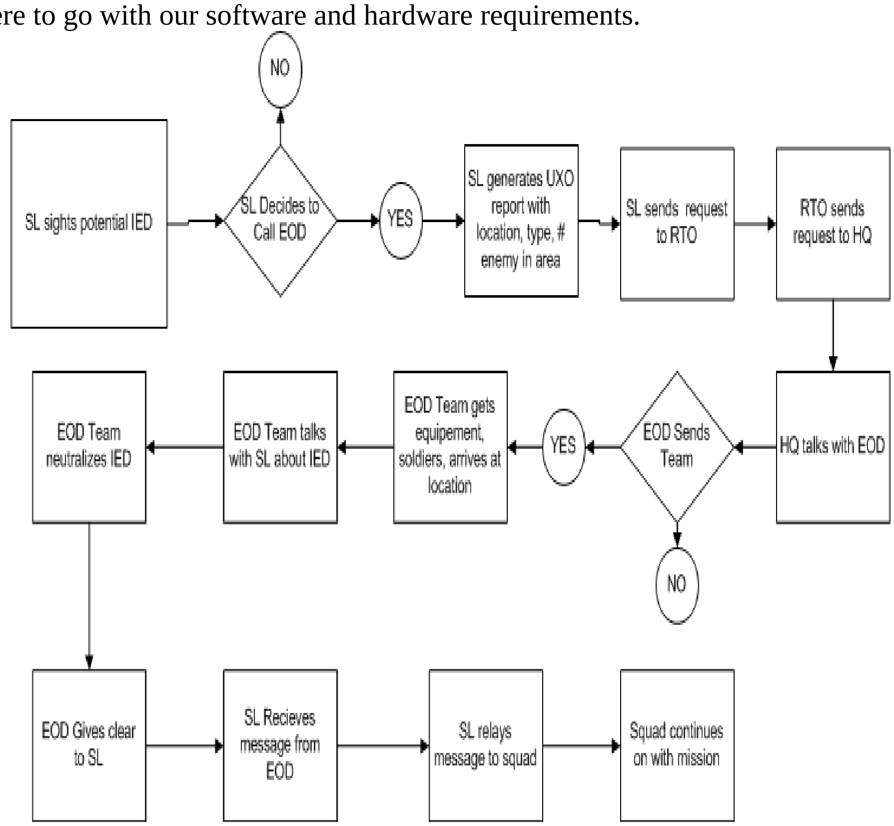


Figure 2: Mission Thread (IED Mission Thread)

Value Hierarchy

The Value Hierarchy is a device that gives us insight to the operational activities, system functions, and system objectives. After conducting a stakeholder analysis, the war fighter, we learned that stakeholders would like the system to perform the following activities: communication with allied forces, acquire video feed, send SITREP, request MEDEVAC, request EOD to clear IED, track allied forces, request call for fire, and conduct movement during the mission. These eight operational activities feed into our six system functions. Our six system functions are: provide text, provide voice capabilities, report enemy/friendly location, provide visual of enemy/friendly, deliver templates to desired unit (EOD, SITREP, call for fire, MEDEVAC), and provide comfort to friendly forces for duration of the mission. Each of our system functions has 2-3 objectives. The objectives are looking either to maximize or minimize an objective.

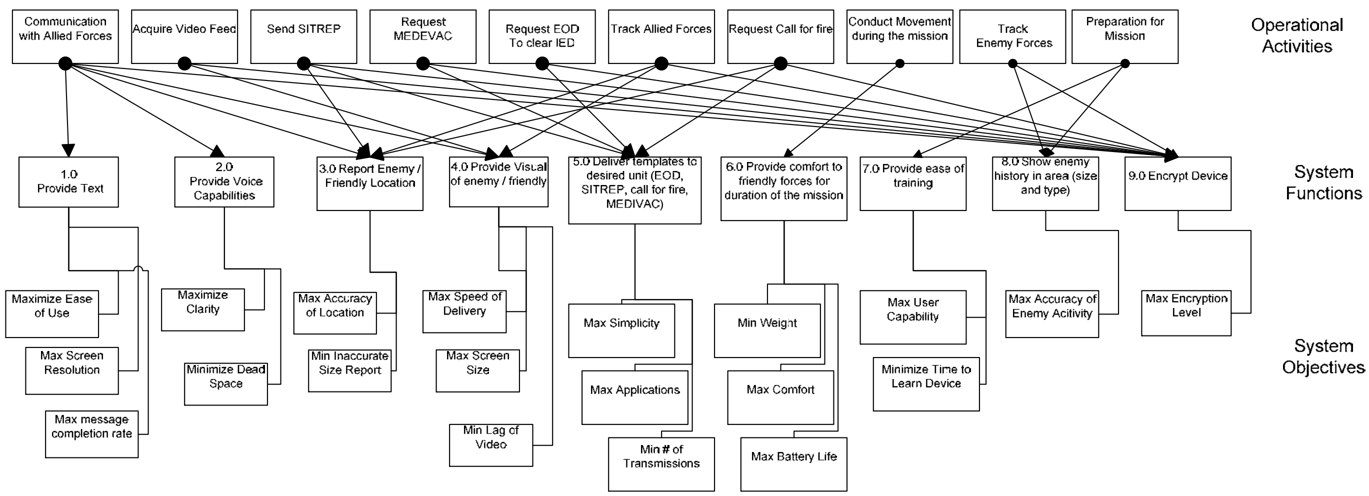


Figure 3: Value Hierarchy

System Objectives

Our system functions fall into the bottom of our value hierarchy. Each system function has a value measure attached to it. The value measures are the criteria on how we are going to score our alternatives. Most of our value measures have an objective measure to them. For example, weight has a measurement of pounds. When looking at alternatives we will research how many pounds it is and score it according to our value model. However, some of our value measures are subjective because they have star ratings. By “star ranking,” we refer to a constructed scale where a user will pick up our device and rate it from 1-5 stars based on how that individual feels the system should be scored.

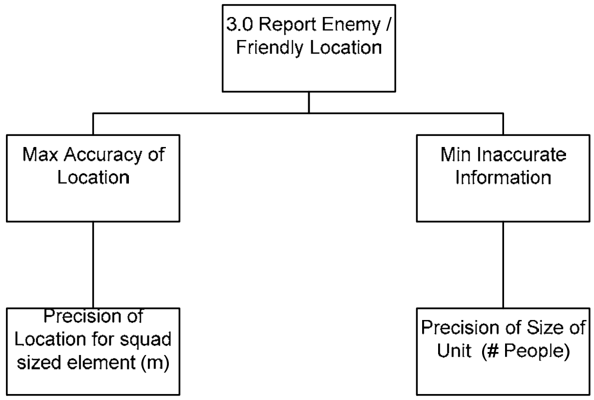


Figure 4: System Objectives

Alternatives

Our current alternatives are Nett Warrior, tablet PCs, and smart phones. Our current baseline and what is being used across the Army, the SINCGARS (ASIP) radio. The Nett Warrior is a device worn by the war fighter and gives the user more capabilities than the ASIP radio. In addition to voice, it provides video, tracking of friendly and enemy forces, and can show the unit's location to higher. It is relatively heavy in comparison to a tablet PC or smart phone. A tablet PC or smart phone are also much less expensive and therefore can be more easily replaced. One problem with the tablet PC and smart phone is that they are unsecure, while Nett Warrior and the ASIP radio have a secret clearance from NSA. A tablet PC and smart phone are extremely similar. The tablet has a larger screen than the smart phone, but does not have voice capability. One big disadvantage of these two options is that they are less durable than the Nett Warrior or the ASIP radio. Currently, Nett Warrior

Value Models

An example Value Models that we created for our System Objectives is shown in Figure 5.These give a numeric value to a specification of the system. With these we can compare the value of our base case to the value of the different options we are weighing against each other. This will allow us to get a total number that we will be able to translate into a score of 1 to 100 which will be talked about in the Swing Weight Matrix portion of this report. We currently have 17 value models for our 17 System Objectives, each with a shape depending on the value associated at different levels. We determined the shapes of the value models based on our personal experiences with technology and talking with our stakeholders and what they thought was important to take this piece of equipment on dismounted missions. For example, weight is an extremely important consideration to the war fighter. The war fighter will already be carrying 50-70 + pounds and sometimes more, on top of being in 110° F weather. After realizing how important weight would be we went back to the war fighter and asked how much additional weight they would be willing to carry to make this command and control unit desirable. We set our ranges of 0 lbs as ideal and 10 lbs as so heavy that they would not take it on the mission.

Weight (lbs)	Value
0	100
3	70
6	45
8	15
10	0

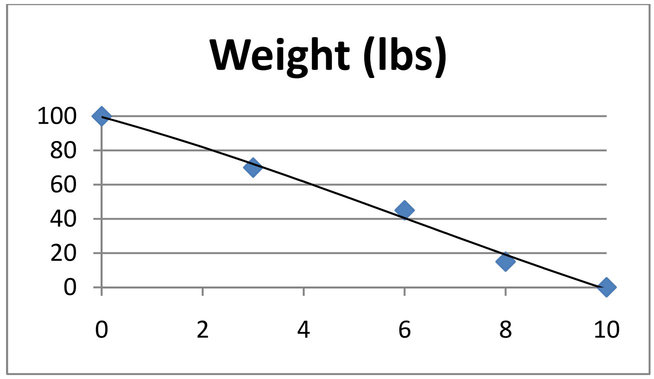


Figure 5: Value Models

Swing Weight Matrix

After talking to stakeholders, we gathered that they have many important value measures. However, we learned that not all the value measures are as important to them. For example, the number one value that stakeholders said is that if it is too heavy they will not use it at all. One of the value measures is precision of location for squad size element. This value measure means how accurate the device will report a friendly/enemy unit. Stakeholders stated that this would not be as important as weight, battery life, or simplicity. The value measures that stakeholders said are most important are in the upper left corner of the matrix and the least important are the value measures in the bottom right corner of the matrix.

		Level of Importance of the Value Measure					
		High		Medium		Low	
variation	High	Weight	100	Fragmented Video	90	Applications	20
		Battery Life	100	Frame Rate	80		
		Perceived Simplicity	100	Accuracy	80		
	Medium	Simplicty	90	Send/Receive Time	80		
		Screen Clarity	85	Voice Clarity	70		
		Completion %	85	Transmissions	60		
		Dead Space	85	Storage Space	55		
	Low	Comfort	70	Time to Learn	50	Percision of Location	15
						Percision of Size of Unit	15

Figure 6: Swing Weight Matrix

<u>Simulations: Modeling Environment</u>

To apply the systems architecture process as described above, we have to get raw data for all of our systems objectives which are shown above in figure 3. We can get data for some of the objectives from the specifications of our different candidate solutions, but some must be tested. The objectives we have to test for are Completion Percent, Propagation, Precision of Location, Precision of Size of Unit, Fragmented Video, Transmissions, and Accuracy. We use a combination of OneSAF and COMPOSER to test our candidate solutions in a scenario to collect raw data for the systems objectives.

OneSAF

OneSAF is a mission simulator that we are using to model and collect data on a scenario that we have designed. OneSAF enables us to input a scenario into a tactical environment using real terrain data.

This allows us to assess the complexities of terrain and communications as part of our simulation and, at once, validates the viability of our scenario as a representation of current military operations. Using a data library of unit types and entities, OneSAF overlays units and tactical graphics onto terrain data and provides a framework in which you can create visual representations of mission execution. Currently, all scenario entities have been input into OneSAF and the scenario execution coding is being completed.

MATREX

MATREX allows several types of communications models to be integrated into a tactical model. This integration allows one to model the situational awareness of the soldier rather than simply a percentage of complete radio messages or tactical movements. It essentially displays the "so-what" of communications and tactical simulations. It shows how these variables affect the soldier. The most important aspect of MATREX is that it supports decision making over the entire acquisition cycle.

COMPOSER

COMPOSER is a Communications Planner for Operational and Simulation Effects with Realism. Essentially what this means is that scenarios or troop movements can be input into COMPOSER along with terrain data from basically anywhere in the world. In addition, COMPOSER will map the connectivity of up to 2000 radio nodes in six hundred times faster than real time. All of these elements combined allow COMPOSER to model real world movements on real world terrain with the characteristics of real world radios. This helps account for factors that cannot be accounted for in other simulations such as dead space due to mountainous terrain. COMPOSER also reports over 7 types of communications traffic flaws in one simulation.

The benefit is that COMPOSER is capable of identifying potential network problems before they become a problem in reality. It can also help screen out candidate solution technologies before any money is invested in them.

Our plan is to input our tactical scenario into COMPOSER and be able to model several different types of candidate solutions through that scenario. The terrain data used in our COMPOSER simulation will be that of McKenna MOUT Site at Fort Benning, Georgia. The movements of entities will be based on the operations order built for our tactical scenario. These movements will be in sync with the movements modeled in our OneSAF simulation as well. The communications traffic will be designed to reflect that of a real unit participating in a mission such as the one we have created in our tactical scenario. Examples of the output provided are messages transmitted vs. time, messages requested vs. time, messages corrupted vs. time, and propagation. COMPOSER will allow us to compare candidate solutions based on several of our value measures. Additionally, it will provide us several other quantitative measures by which to compare the candidate solutions.

COMPOSER is an incredibly valuable tool for our capstone group as well as the Army as a whole. It saves time and money, and allows for accurate modeling of new technologies on real terrain.

Scenario

The scenario we designed captures a mission for the non-infantry Army unit. In the scenario, we have a platoon-sized element with mortar attachments moving to an Afghani village in order to meet with local village leaders. This unit moves along a road approaching the village from the northeast when they spot a series of potential improvised explosive devices (IEDs) buried in the road. Thus, they dismount their vehicles, call in an explosive ordnance disposal unit to deal with the potential IEDs and continue with the mortar section setting up to support the dismounted maneuver platoon. As the dismounted unit nears the village they are attacked by a squad sized element of enemy forces. The dismounted unit returns fire, gains fire superiority and maneuvers to destroy the enemy element. During this engagement, a call for fire is initiated and various situational reports are generated. After the engagement, the dismounted platoon continues to the village, sets up security, conducts the meeting with local leaders and withdraws to the northeast. When they return to the vicinity of the mortar site, they mount their vehicles and return to the forward operating base in the north.

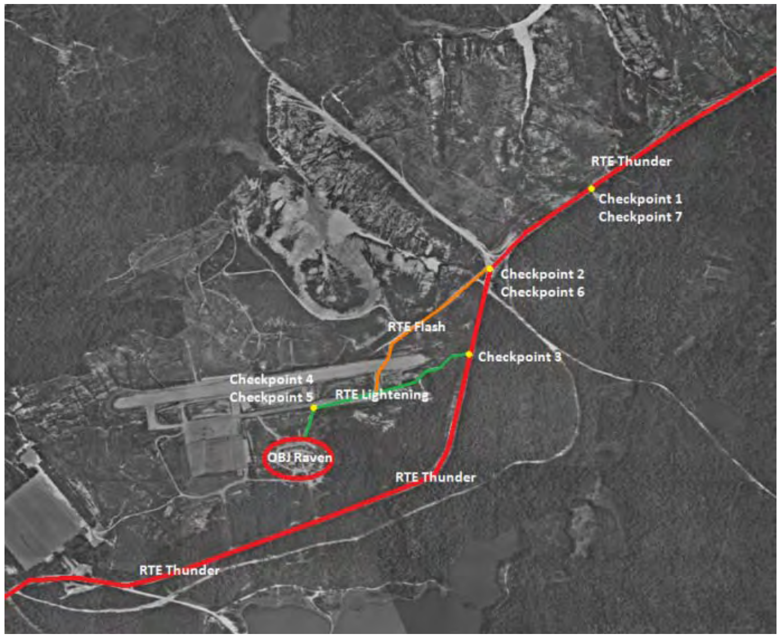


Figure 7: Scenario Graphic

Encryption

Encryption is a baseline requirement for any possible solution. The device needs to be secured so that if it were to fall into the hands of the enemy, they could not use it to find the location of our troops and plans. It also needs to be secured so we can send information without it being intercepted. A potential issue with using applications is that they are easy to intercept, easy to exploit, and the architecture is very open. In addition, a built in GPS would give away the units position if it fell into the wrong hands or was intercepted. The only way to shut off the GPS function is to take the battery out of the device. Even if the enemy could not crack the encryption they still could potentially see that there was activity on the net.

Our Proposals for a New Solution Across the Army

We are currently working on coming up with a solution that can be implemented across the army. We are going to look at various options and weight then to get a score for each item. With this we will be able to measure the value of an option against its cost to provide a recommendation to our stakeholders of which option to go with. In the cost we will need to consider maintenance cost, cost to fix broken units, and training cost. Our next step is going to be getting the numeric values on the options we are looking at. With these we will be able to put them into our architecture and swing weight matrix to get a value for each option.

A Schema for Transitioning from Rectilinear
Coordinate Spaces to Pseudo-Hexagonal Spaces:
Help Me Rhombus, Help, Help Me Rhombus

There are two main approaches to managing wargame units interactions with their environments (movement, line of sight, area of effect, etc.): freespace and gridded. In the freespace approach, the units exist as entities in a continuous volume of (usually) Cartesian 3D space. They move in any direction (based on interaction with "terrain" that occupies the same space) and interact with each other based on references and displacements from their position in that space. In the gridded approach, space is broken up into (usually regular) shaped pieces. Units are considered to occupy the entire volume of one of these pieces; movement, line of sight, and other interactions are based on the relationships amoung the spaces rather than the absolute positions of the units themselves. Both approaches have advantages and drawbacks.

The freespace approach seems to be more "natural" or "realistic" from an intuitive sense. Units can occupy a number of positions and orientations that is only limited by the degree of effort you are willing to put into measurement. It is easy to make analogies between the physical world and the wargame battlespace.

The challenges with freespace are mainly in complexity and level of effort in implementation. Calculating relationships amoung points and volumes in freespace can be difficult, especially when dealing with paths through that space that are non-linear (they "go around" things in the space). There are a lot of physical techniques that can alleviate these challenges, mostly due to the close correspondence to the physical world. These techniques, however, usually increase the level of effort challenges inherent in the freespace approach. The archetypical level of effort challenge puts the player in the position that since they have so many degrees of freedom in action, that to gain best advantage in those actions, they have to put large amounts of effort into examining each action in a high level of detail. For example, an individual combatant who is walking a "S" path in freespace can have a significantly different advantage of position by making minute changes to the curvature and inflection points in the path.

The gridded approach is generally seen to overcome the challenges of freespace. By dividing the space up into reasonably sized discrete *cells*, movement, line of fire, cover, and other interactions can be implemented with relatively simple systems based on whole number counting. Additionally, since there are no different positions within the cells, there are fewer degrees of freedom in actions, and thus the required or desired level of effort ("Waitaminute! If I moved my guy a half a millimeter to the right, would that let me ...") in considering options is lessened. This decreased tactical focus frees up players' cognitive resources for other, generally considered "deeper," considerations such as "look-ahead," strategy, and consideration of the meta-game.

These advantages do come at a cost. The reduction in degrees of freedom equates to a reduction in verisimilitude. For example, units can occupy a certain cell or another adjacent cell, but cannot occupy a space "in between" the cells that might logically equate to a tactical advantage over either, or be the outcome of a more naturalistic decision process such as the desire to stop movement directly "behind" a pillar that acts as an obstruction to fire rather than "to the right" or "to the left" of it. It is frequently the case that in three dimensional gridded systems one of the dimensions (usually "altitude") is treated differently than the other two (forming the "ground"). This usually introduces complexity (in the form of additional rules) and again detracts from the verisimilitude.

There are other, more basic problems with the "feel" of gridded systems from the constraints imposed by the cell structure. In square grids moving four units *north* and then three units *east* results in taking seven unit steps, but the measured distance between the centers of the starting and ending cells is only five units. Breaking this movement up into a sequence of smaller movements does not alleviate the problem.

Hexagonal grids can mitigate this problem with square grids, as the correspondence between the minimum number of steps between cells and the measured distance is much closer. Additionally, hexagonal grids provide six degrees of freedom in movement vice the four of square grids at no increase in rule complexity. Even though there are more options, they tend to seem less "natural," possibly because people are more familiar with rectilinear grids and have more patterns of mind based on "forward, backward, right, left" as opposed to "every sixty degrees." The biology of bilateral symmetry may even have an effect on this, which would be a difficult part of the mindset to overcome.

An additional issue with hexagonal grids is that they are not closed under composition, that is, you cannot take a regular hexagon and break it up into any whole number of smaller hexagons that completely cover the same shape. Square grids, however, are easily decomposed into smaller square grids, or aggregated into larger ones.

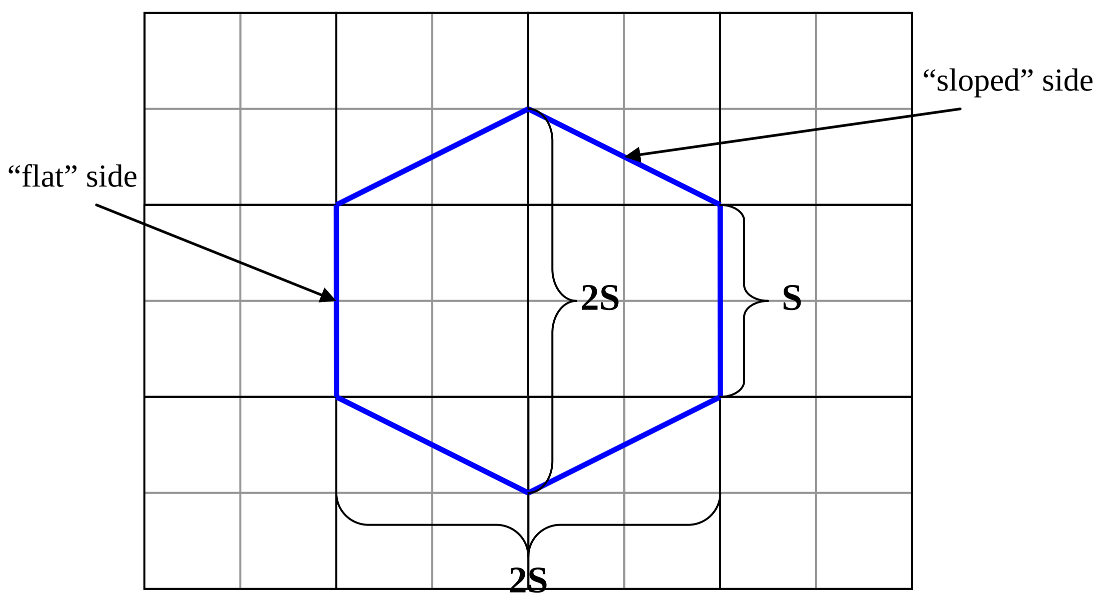
The general issue that this discussion has addressed is that there is no "perfect" approach to implementing a wargaming battlespace. Each of them (and this extends to others not discussed) has different sets of advantages and disadvantages. Nothing will change that basic nature of the various approaches, nor would it be desirable to do so. Along with the advantages, the challenges define the feel of the game and focus the thinking of the players on certain aspects and away from others.

Directing players' thinking and behavioral patterns is a very important factor in the enjoyment, catharsis (immersion), and experiential learning value of wargames. While the approach to implementing the battlespace is not the only driver of the players' experiences, it is a significant one since it affects most decisions and other thinking about the game. The goal of the proposed schema is to provide an approach to integrating square and hexagonal grids to make a hybrid battlespace that will allow the game to provide players access to the feel of either system, in portions where appropriate, with a minimum of difficulties in transition between the two.

So what are the general feels of the square and hexagonal systems and what is a reasonable use case where you would want both combined?

The square gridded environment provides a *best fit* for man made environments, especially the interiors of structures and built up urban areas. The rectilinear structure of streets and hallways can usually be lined up with the grid itself, creating a natural harmony between the referent and the representation. In general, the more densely built up a man made environment is, the less opportunity there is for “diagonal” movement, so the less important the challenges with moving large distances at angles not aligned with the grid are. This system still has significant challenges representing large open areas and structures with multiple grids, offset angularly from each other.

As a sort of compliment, the hexagonal grid is most popular with outdoor environments. It provides the additional degrees of freedom of motion expected in open areas and has much less separation between steps and measured distances over larger portions of the environment. Likewise, the hexagonal system has significant challenges representing rectilinear shapes. In general, they have to be constrained to unnatural, offset positions and the hexes that contain the edges of the rectilinear shape usually have some *loss* – large portions of area that are inaccessible to the units in the game, but not actually occupied by the represented terrain itself.



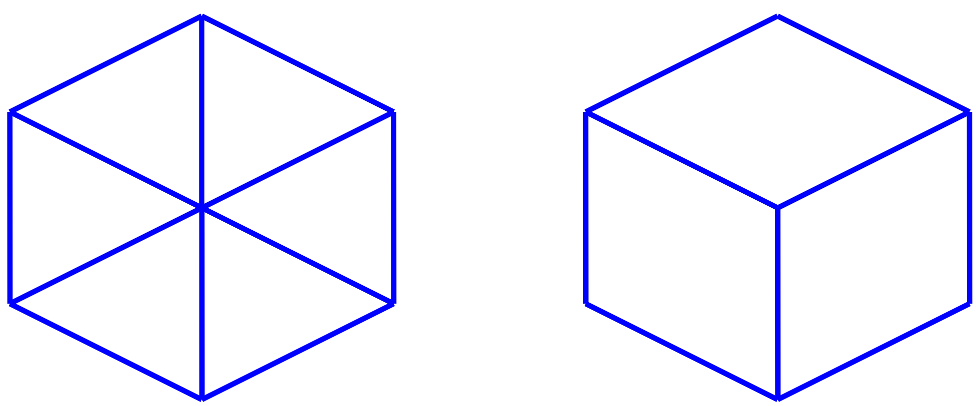
Our stretched hexagon

One of the most common methods to bring these two approaches together is to not bring them together. Or, more specifically, to conduct different parts of an overall scenario on different maps that have entirely one or the other type of grid, as appropriate. The second most popular method is to severely limit one or the other type of terrain on the map and allow it to be represented with the approach appropriate for the dominant part of the terrain. The entry point to the city might have a small strip of outdoor terrain at one edge of the map or the number houses and outbuildings in a rural area might be minimized. Both of these methods work well enough, but are certainly compromises to the idea of a hybrid square and hexagonal grid. The easy way, just mashing the two grids together and accepting or artificially putting obstructions in the inconvenient places where the two are not compatible is also somewhat unsatisfying.

The proposed approach to combining square and hexagonal approaches, which we will call the rhombus interface, leverages rhombuses constructed from equilateral triangles into which the hexagon can be decomposed to bridge the gap between the approaches, maintain relative consistency between the two as much as possible, and provide most of the feel of the hexagonal approach.

The approach is based on a chosen square grid, with sides of the cells being length s. The hexagonal grid to be integrated will be made of slightly "stretched" hexagons (not quite regular hexagons), adjusted to mesh with a square grid of side $s/2$. One pair of opposite sides for these hexagons will be of length s and parallel to one pair of sides of the square grid and spaced $2s$ apart from each other, with end points aligned. We will call these the *flat* sides. The point-to-point diagonal of the hexagon between the flat sides will also be of length $2s$, and will be in the middle of the flat sides, centered on their length. The sides of the hexagon that connect the flat sides to this diagonal will be called the *sloped* sides.

Note that by arranging one of the flat sides with one side of the square of the base grid, the points of the point-to-point diagonal that defines the sloped sides align with the half grid. Also note that the area of this hexagon is exactly $3s^2$, or the area of three of the full squares of the base grid.

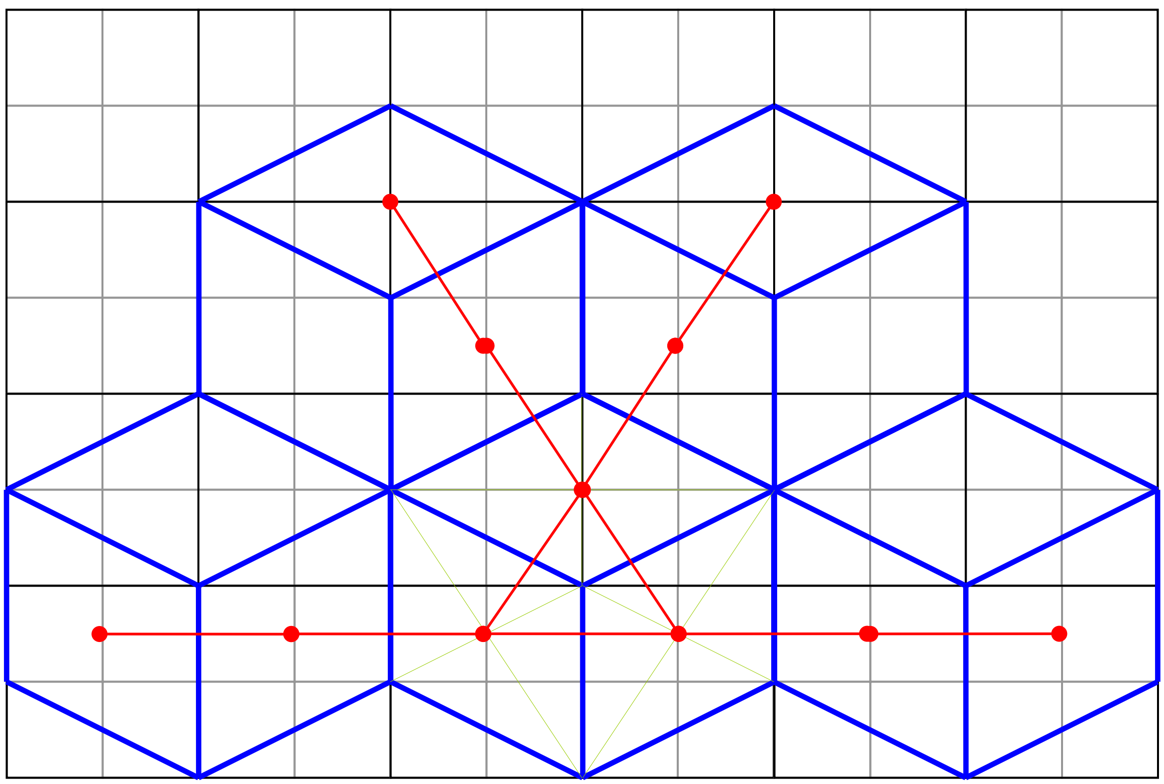


triangles and a top set of rhombi

The vertices of the hexagon align with a whole number ratio of the base grid, with no need to approximate irrationals. The half-unit is a standard subdivision of both English and Metric measures and fairly easy to physically approximate without measuring. This makes calculation, vector construction, graphical representation, hand drawing of the grid, and calculation of an individual hexagon from a reference point without calculation of all the intervening hexagons all relatively easy and computationally simple. These properties will propagate across the grid as the hexagons are repeated.

Next, we will break this hex down into six triangles, then aggregate them to make the rhombi that will be the base cell for the organic part of the grid. Dividing the hexagon is simply connecting the vertices with the center. Creating the rhombi is likewise a simple matter of choosing pairs of triangles that share a side and eliminating that side. The resulting four sides make a rhombus. There are two different ways you can do this for each hexagon. The one illustrated above will be called a *top* set, the opposite will be called a *bottom* set.

Since each of the three rhombi are congruent (each is composed from a pair of congruent triangles), it is easy to see that each has an area one-third of the original hexagon, or s^2, the same area as the squares in the base grid. Also, the lengths of the pairs of sides are s for the flat sides (parallel to the flat hexagon sides) and a little less than 1.2s ($^{\sqrt{5}}/_2\ s$, to be precise) for the sloped slides (again, parallel to the sloped sides of the hexagon).



cell to cell transitions

The rhomboid cell, then is decently close in basic size and shape parameters to the square grid cell and fairly simple to construct from that underlying grid. These are nice descriptive parameters, but the real key performance parameters are how the grid supports interactions amoung units in different cells.

Qualitatively, we can see that the rhomboid grid gives us similar degrees of freedom for cell to cell transition as the square grid – four choices by passing into an adjacent rhombus that shares a flat (in some cases) or a sloped (in all cases) side. In the aggregate, however, the cell to cell transition is similar to the hexagonal system, giving us six "straight line" directions overall.

Quantitatively, the cell to cell distance transitioning through a flat side is exactly *s*. The cell to cell distance transitioning through a sloped side is $\sqrt{(^{13}/_{16})}s$, or a little over .9*s*, for just less than a 10% loss for steps vice geometric distance. When you consider the effect on multiple transitions that change direction, the variances get larger, but also become dependent on path selection to get from one cell to another. A good metric to evaluate the overall distortion over larger distances is to calculate the relative area of a geometric circle of radius *x* and compare it to the sum of the areas of the unique cells that can be reached by *x* cell to cell transitions.

steps	**circle area**	**square grid area**		**hexagonal grid area**		**rhomboid grid area**	
1	3.14	5	159.15%	6.06	192.97%	5	159.15%
2	12.57	13	103.45%	16.45	130.94%	13	103.45%
3	28.27	25	88.42%	32.04	113.33%	27	95.49%
4	50.27	41	81.57%	52.83	105.10%	43	85.55%
5	78.54	61	77.67%	78.81	100.34%	65	82.76%
6	113.10	85	75.16%	109.99	97.25%	93	82.23%
7	153.94	113	73.41%	146.36	95.08%	123	79.90%
8	201.06	145	72.12%	187.93	93.47%	161	80.07%
9	254.47	181	71.13%	234.69	92.23%	199	78.20%
10	314.16	221	70.35%	286.65	91.24%	247	78.62%
11	380.13	265	69.71%	343.81	90.45%	293	77.08%
12	452.39	313	69.19%	406.17	89.78%	351	77.59%
13	530.93	365	68.75%	473.72	89.22%	405	76.28%
14	615.75	421	68.37%	546.46	88.75%	473	76.82%
15	706.86	481	68.05%	624.40	88.34%	535	75.69%
16	804.25	545	67.77%	707.54	87.98%	613	76.22%
17	907.92	613	67.52%	795.88	87.66%	683	75.23%
18	1017.88	685	67.30%	889.41	87.38%	771	75.75%
19	1134.11	761	67.10%	988.13	87.13%	849	74.86%
20	1256.64	841	66.92%	1092.06	86.90%	947	75.36%

Because all these approximations converge on their own "best fit" to a circle, you can make any of them come arbitrarily as close as you want by changing the area of the grid cell. However, the point of the comparison is to evaluate the grids that also have a good fit to the "one unit step" from center cell to center cell. The square and rhomboid used above are the ones discussed. The hexagonal grid used is the one with a flat to flat diameter of *s*, which is the one that gives a center cell to center cell transition of *s*. At long ranges, the rhomboid approach ends up converging on an approximation of a circle somewhat better than the square, but not quite as good as the hexagonal.

Based on the above analysis, the rhomboid grid is *almost* as good as the hexagonal in several ways, so why not just use the hexagonal grid? The answer comes in trying to provide a reasonable interface between the two grids. A regular hexagon will either have at least one side length or one of the diameters that is an irrational value, which means only certain integer multiple numbers of rows and columns of hexagons will make a good fit with a square grid. Regardless of where those convergence points are for any two particular schema, there will still be significant loss where they are incompatible. Additionally, in the interim between the interfaces, groups of real number interpolations will have to be calculated both within one grid or the other and always to cross the interface. The advantage of a gridded system is the ability to quickly calculate transitions using counting and whole number math and systematic, repeating formulas, especially where this enables calculating relationships between non-adjacent cells without having to calculate states for intervening cells.

Because of the alignment of the rhombic dimensions with the square grid, there are a number of ways to interface between the two, creating a hybrid cellular space. A schema is proposed in the diagram below that provides a number of highly desirable characteristics for the composite space, and also illustrates a few of the challenges.

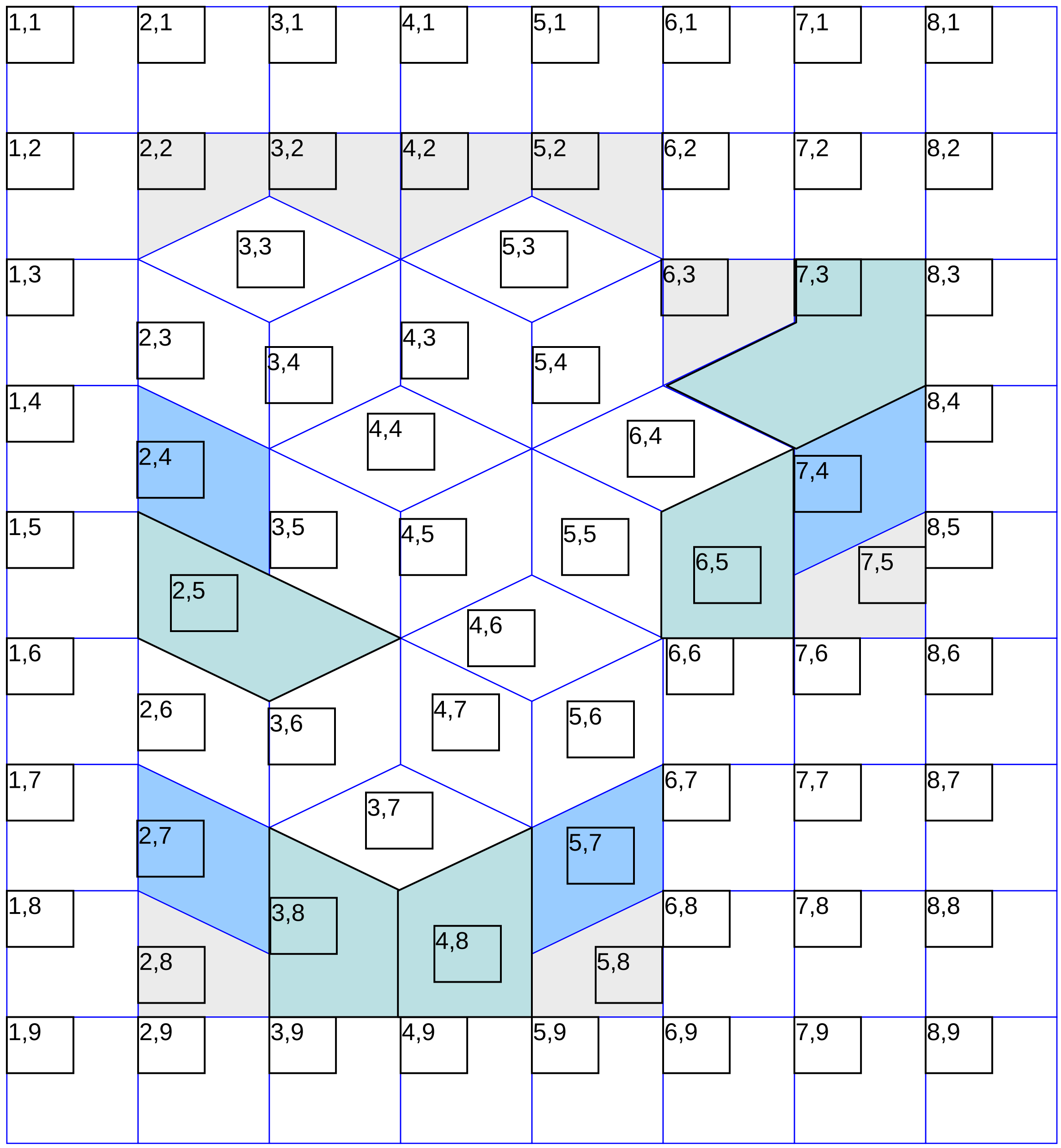
At first glance, the diagram shows two important things: (1) that the interface can be created through the repeated application a small number of simple steps, (2) that the *area* of the square grid can be maintained (that is, among the square cells, rhombic cells, and interface cells, all the cell coordinates that would have been in a homogeneous square grid of the same size are accounted for). The color coding scheme highlights one of the obvious challenges. The scheme shows: *normal* square and rhombic cells in white; *truncated* cells in gray; *augmented* cells in cyan; and *added* cells in blue.

For the truncated cells, a square cell that overlaps with a rhombic cell is considered to be only the part of the square cell that is exclusive of a rhombic cell. Conceptually parallel to the idea of being truncated, some of the truncated cells (the “points down” ones) only share a side with three different cells. The effect on implementation is that change of position in two different directions could lead to the same place.

For the augmented cells, a square cell that overlaps with a rhombic cell is subsumed in the rhombic cell. In a similar conceptual parallel, these cells share a side with more than four other cells, and may share multiple sides with one other cell or have single geometric sides that border two cells. This can be handled in two primary ways: (1) ignore one of the border cells so there is no legal cell to cell transition where there is a physical sense that there should be one, (2) add custom decision and management criteria for transitions from these cells.

The added cells are less problematic. They are simply additional rhombic cells that don’t align with the existing pattern that are created to aid in the interfacing process.

Across all the interface cells, there are challenges with numbering and transitions, but these are relatively easy to deal with through the repetitive application of integer based algorithms. Their most significant challenge, however, is that they cause noticeable problems with the *flow*, of the grid, which is the next topic.



an interfacing schema

As was earlier mentioned, the entire set of expected coordinates from the square grid are mappable to the hybrid grid, with a small number of difficulties and a reasonable number of reasonably easily handled special cases (This is especially true when you consider the typical application stated for this approach – the meshing of relatively large areas of the two systems to create a combined urban-open area that is tactically useful for wargaming. The number of normal cells grows geometrically with expansion of the area, while the number of interface cells grows only linearly, with the problem cells being a relatively small portion of those.) But the piecewise changes from cell to cell are not as important as the overall performance, especially over large distances and across multiple transitions – the flow of the hybrid grid. Using the base square grid as a referent, we will look at the overall horizontal and vertical flows.

The horizontal flow is the easiest to address and the simplest to evaluate. One can transition from the (1,x) to the eighth column cells in seven steps, the same as the square grid. Almost all of them are through a path that stays in the same row and most of them can be achieved with a repeated series of east moves, again the same as the square grid. Since the variations are small in number and size, this seems to be a very good fit. Again, going back to the typical use case, the existence of a few places where you have to move a little to the side to cross an area of rough ground does not seem to be incongruous.

The vertical flow has more challenges. The only paths through the rhombic grid that can maintain the number of crossing steps that the square grid has are ones that start near the vertical edges of the interface area. The minimum transit paths vertically through the center of the rhombic grid only cost about one more transition in ten, which is compatible with the distance loss of the squashed hexagonal grid, and isn't especially incongruous with the typical use case. Additionally, all the minimum vertical paths that don't run along the interface edge require some lateral movement. A large variation is not required, but it does affect a large number of paths.

When you couple the horizontal and vertical flow behaviors, an interesting characteristic emerges – there is an "easy" direction to transit the area, and a "hard" direction, that is, horizontal transition consistently requires less variation from the square grid performance than vertical. This behavior is not incongruous with many open, natural areas which have bias in the sloping and variation of the terrain, but it is not strongly compatible with all types of typical use case terrain (such as a groomed and manicured park in an urban area).

When we note that this is only a function of horizontal and vertical as displayed in the example and not all grids in general, we see yet another interesting emergent behavior. Rotational symmetry of the grid allows us to swap the vertical and horizontal of the sample without changing the performance parameters relative to the original horizontal and vertical – the easy way becomes the hard way, and vice versa. This property creates the opportunity to, within the same overall map, have subsections of rhomboid grid that have different *orientations*, which could be managed to impact that tactical characteristics of the terrain.

One additional emergent behavior of the approach is that swapping out top for bottom rhomboid configurations in the superposed hexagonal cells does not affect the long overall performance of interfaced gridded areas, but can cause additional local variations where the changes occur. Again, this behavior could be managed to support specific terrain effects that impact the tactical characteristics of the terrain.

As a final significant parameter of the hybrid grid, while it is not possible to easily integrate *every* possible combination of hexagonal and square natural areas, the interface is able to be changed in a regulated fashion at the edges in small increments. This means it would be relatively easy to integrate a wide variety of shapes for both types of sub grid.

In summary, the superposition of a rhombic grid over a base square grid provides the opportunity to leverage the advantages of gridded wargame battlespaces and enables optimization for urban and natural terrain in a piecewise fashion with a minimum of compromises.

EXPLORING THE USE OF COMPUTER SIMULATIONS IN UNRAVELING RESEARCH AND DEVELOPMENT GOVERNANCE ISSUES

Mariusz A. Balaban, Old Dominion University

Patrick T. Hester, Ph.D., Old Dominion University

Abstract

Understanding research and development (R&D) enterprise relationships and processes at a governance level is not a simple task, but valuable decision-making insight and evaluation capabilities can be gained from their exploration through computer simulations. This paper discusses current modeling and simulation methods, addressing their applicability to R&D enterprise governance. Specifically, the authors analyze the advantages and disadvantages of the four methodologies used often by modeling and simulation practitioners: System Dynamics (SD), Discrete Event Simulation (DES), Agent Based Modeling (ABM), and formal Analytic Methods (AM) for modeling systems at the governance level. Moreover, the paper describes nesting models using a multi-method approach. Guidance is provided to those seeking to employ modeling techniques in an R&D enterprise for the purposes of understanding enterprise governance. Further, an example is modeled and explored for potential insight. The paper concludes with recommendations regarding opportunities for concentration of future work in modeling and simulating R&D governance relationships and processes.

1. Introduction to Research and Development Governance

Key tasks of governance depends on many considerations, but general possible governance level responsibilities can be classified as follows: strategy and operational framework design, vision on direction of organization propagation, goals setting, funding allocation, risk reduction and uncertainty assessment, effectiveness and performance assessment, decision making in regard to management or lower level structures, policy adjustments, policy adherence verification.

Research and Development (R&D) governance is a very broad term and have different implications for public and private sectors, yet some rules and perspectives are somewhat similar. Goals and objectives are quite different when one looks at of these two groups, yet both of them provide framework for guidance, oversight, accountability, projection, and evolution that encourages desirable research outcome for enhanced mission capabilities (Keating, et al., 2008). Here, it is important to understand major difference on how both sectors see and evaluate their mission capabilities. Private sector governance concentrate on investments, product lifetime cycle, future planning, sustainability but the most important goal is to make more money. Many other variables can come into play but ultimately what really matters is the profit. Exceptions from this summary could be non-profit organizations or corporations where outcome measures can be closer to public sector measures. Public governance focus on different outcomes like skills of leadership, policy soundness, innovation creation, level of security, health enhancement

etc. that divides into more specific outputs from more detailed managed endeavors. Money spent is important factor, but the main mission objectives do not mention making money as goals.
Another context to consider is the level of R&D governance, because it may span through society, country, industry, company, or individual endeavor within organization. Different approaches to R&D governance can be more effective with different sizes and complexities. Another factor to study is different R&D organization structure for example centralized, de-centralized, or mixed; this can have significant impact on performance of the group (Argyres, et al., 2004). Assessment of R&D endeavor is tightly coupled with goals of the organization. Input in form of goal setting will direct R&D process through execution of projects and ultimately feedback to governing cell in form of e.g. profit, increased market share in private sector and e.g. lowered crime rate, cancer related deaths rate decrease, increase of approval rate for given organization in public sector. Formulated governance strategy establishes context for R&D entities and consist a basis for funding request, programs adjustment or even outsourcing to accomplish its goals. It is also enabler for generating of R&D project ideas, which would help to fill the gap between current state, and organization objectives.

2. Needs for modeling in R&D Governance

Decision making without sufficient information can lead to catastrophic results and that itself should be enough to bring modeling and simulation methodologies into the table. Frequently, available data does not translate into valuable information that facilitates understanding of occurrence. When investing significant amount in a research project one should first know that proposed solution has high chances of success, and if it is directly addressing the problem for which is offered. Another problem can be discovered while looking at reports of government agencies where many hundred pages reports do not provide clear information on bigger picture contexts. Measures of “target met” type results for particular research endeavors does not provide sufficient way for validation of decision-making process and its evaluation. On the other hand, authors realized that a reason for this situation might be lack of knowledge on how to describe successfully R&D processes, and its relations at sufficient level of abstraction. Problem can be instantiated into scope of conceptual understanding to see how this system works. If one knows that, he or she can build a model out of this knowledge that could aid in decision-making process. This sound easy, but as a matter of fact it is not. There are many parts of system influencing each other and it is hard to describe these relations in mathematical formalism. Human in the loop is still a big challenge to model and when dealing with innovation it is obvious that one of major factor to innovation is not only qualifications of scientific staff but also their motivation, character, needs, and all that will influence culture of the group working on the project. That is why evaluation of innovation generation can be subjective and governance level decision-maker should be aware of that. Modeling and Simulation practitioners are working diligently trying to fill this gap up, which common analysis techniques could not. Agent Based Modeling is one of the most promising techniques that could generate understanding of this phenomenon and emerging group behaviors. Question often asked is how to asses and measure R&D effectiveness. One way of approaching the problem is to focus on internal and external factors that facilitate effectiveness and not on innovation product itself, especially when assessing a long-term project where it is difficult to wage progress of research. Another way is to find variables directly coupled with outcome from R&D endeavor keeping in mind that some variables can be measured, but some can only be observed as trends. Length and difficulty of the R&D endeavor can have significant impact on accuracy of prediction, and so assessment as well. When considering the length it is apparent that long-term R&D endeavors should be harder to judge and probability that one will conduct erroneous assessment is higher. By employing modeling techniques, one could address some of the problems mentioned here. Modeling innovative endeavor and its supporting activities can be achieved using System Dynamics or Discrete Event System techniques. They can alleviate a problem of R&D effectiveness assessment and help to evaluate trends and conduct what if analysis on different styles of decision-making, different polices of R&D governance that oversees innovation generation, etc. Successful end of R&D innovation generation is not the end of the process. Another responsibility of R&D governance is to incorporate a new product or concept into market or its functional target place. This task can be very complex and traditional methods may not suffice. Modeling and Simulation (M&S) can mimic time related variables and environmental relationships related to new invention to optimize process of integration to maximize outcome, and avoid pitfalls that could possibly occur. Another area where M&S can be invaluable is assessment on future direction of R&D. Conducting analysis of R&D effectiveness using modeling techniques in context of organization mission can reveal serious strategy flaws and help avoiding them. Modeling and simulation (M&S) is relatively new approach, (except of SD) as a choice for aid in decision-making and evaluation of R&D governance. Authors were looking into areas where M&S is already more advanced to analyze options in applying robust knowledge to R&D governance. The most dominant area of M&S is associated with, and sponsored by military organizations. In (Chaturvedi, et al., 2008) authors used ABM to model Political, Military, Economic, Social, Information and Infrastructure (PMESII) entities to mitigate impact of disaster and military operation. Model allows to mimic rebuilding and expansion of critical infrastructure of Afghanistan. Ability to get insight into process of reconstruction and development can be very valuable tool for a decision-maker in enterprise or government structures, used for peaceful scenarios. Because of the multi-dimensional scheme of the model, authors believe that it would be possible to apply this approach to R&D governance. Adding modules that would allow agents to mimic innovation generation and knowledge sharing is foreseen. Connecting this lower level to middle and higher-level agents e.g. department, organization, society etc. Model could allow analyzing robustness of policy choices under different scenarios. This approach can be used for public as well as in private sector. Adding more complexity to model would theoretically allow for modeling human evolution as a function of R&D progress. Because the progress of R&D could be an enabler for actual change of the system and can spur achievements of higher objectives of system of systems, it can be invaluable to be able to model this process. As a futuristic perspective, such a model can provide valuable information on R&D as a stabilizer or destabilizer of the world. One should try to use it to answer some important question:

Is advancement of R&D proportional with civilizational advancement?

Is our direction of R&D advancement taking us where we want to be?

Because of its complexity, currently only large organizations could handle expenses related to investing in complicated models conveying many levels of abstraction, but this can change as field of M&S advances providing for easier model reuse and better interoperability.

3. Modeling Paradigms

In this section, authors would like to provide quick overview of three popular and often used by practitioners modeling and simulation methodologies. Understanding principles of them is an important step in understanding their possible usage as discussed in the next section.

System Dynamics (SD)

System Dynamics method consist of a feedback loops in form of differential equations that provide for building relations between variables. It is useful for studying complex nonlinear systems, especially finding cause and effect relationships. From the practical perspective on how to build model one should know about stocks, flows, auxiliary variables, and constants as the main blocks providing for metaphor of the complex system. The initial phase of building SD model is creating a causal diagram. Causal diagram represents the most important elements of system and relations between them in form of links that end with arrows indicating what influences what. In, addition, you can mark + or - mark on the line that specifies positive or negative relationship. Positive one means that second variable follows the direction of change of first variable, and in negative case, direction is opposite to the direction of change. Links with variables form loops, and they can be positive or negative as well. Negative loop exist only if number of negative relationships is odd. Having at least one negative loop is a minimum requirement to stabilize the system. Models built with this methodology can help framing issues and problem, revealing dynamics related to change imposed on system. Models are typically used to show trends of relationships and not precisely computing specific values. This methodology can be an invaluable tool in assessing big picture problem, testing alternative policies and strategies at the governance or enterprise level.

Agent Based Modeling (ABM)

World is full of interactions between different types of systems and many of them are not well understood. ABM is a methodology that aims at capturing many types on interactions by using computer created entities called “agents”. These agents are assigned attributes, rules of behaviors, mimicking and carrying set of interactions to gain understanding of the real system, and providing for emerging behaviors. ABM can be used to model perception, autonomous and group behaviors, goals setting, static or dynamic environment, ability of inference and adaptation, and social interaction among other possibilities. For that reason, it is popular to use ABM to generate hypothesis about system behavior. From scientific point of view, this is very powerful feature, but there are some practical limitations related to inability to validate logically such models. Consequently, verification and validation (V&V) is the issue that needs to be

resolved to take full advantage of ABM. One way of partially cracking this drawback is by proving hypothesis false.
Complexity and dynamic behavior of systems encourage use of ABM for investigation of adaptive systems, which are often non-linear and chaotic. Furthermore, mechanisms provided lead to model evolution of adapting agents. This capability set ABM methodology apart from other methodologies discussed in this paper. It is worth mentioning that relative high complexity can be achieved by imposing on agents relatively simple interaction rules. There are many platforms for implementing AMB e.g. NetLogo, Swarm, Mason, Repast, and AnyLogic with some of them providing unique features. Choice of one of these platform would greatly depend on complexity of system that one would like to model and problem that built model is supposed to address.

Discrete Event Simulation (DES)

In a discrete model, changes in your virtual system can occur only at separate points in time. These changes are called events, and everything in the model is related to them in one way or another. For example, event could be an arrival of a R&D project for its approval, then decision made would mark next event, following with outcome event from the project e.g. success or failure. DES model consists of entities, which in our sample case could be represented by R&D projects, flowing through designed by modeler pattern called model logic. Events are stored in a calendar, which contain information that allows model to be executed in accordance to its logic. The central idea of DES is that variables of the model will not change between successive events. In addition, important to understand element of DES are queues. They are just like lines in the store, and can be define depending on the system's real queue, which may have limitations for number of elements that can fit into it, and has different rules reflecting priority of leaving it by stored elements e.g. FIFO means first in first out. Essential component of DES represents resources, which may be personnel, equipment, infrastructure etc. They will be used by entities while going through their journey. DES is often used to capture stochastic behavior of the system but can model deterministic events as well.

4. Model choice

One can model R&D governance at different levels of abstraction, complexity, and will require different levels of fidelity depending of purpose of a model. It is important to understand some basic limitations of M&S to go with this paragraph further. One needs to remember that modeling is not cheap, but in turn can save a lot of money and time. While mentioning time, building model often takes significant amount of time, and model validation will add some to it. There hardly ever exists one model answering all questions too, so it is important to wage the cost effectiveness factor when deciding to use modeling and simulation techniques. Here, meaning if one can address problem with simple deterministic calculations there is really no need for M&S. Analytical Methods (AM) are usually quicker and cheaper. AM are often used to get exact answer to a problem, usually where there is a direct and known relation between input and output. Short length term R&D endeavors should be especially good candidates for use of AM in effectiveness assessment at the managerial level. This could be done e.g. by combining R&D into business process. In some cases AM allow for quick estimates for a problem, which then needs to be address more thoroughly with one of M&S methodology as AM will not allow to capture the bigger scope or too complicated problem. Generally, choice of modeling technique depends on what features of a system can be effectively manipulated with different methods. Aggregation level is another factor that will influence choice of modeling technique. It is important to understand that lower level of aggregation does not necessarily translate into lower risk associate with e.g. prediction. At the lower level of abstraction of R&D governance one can focus on e.g. innovation generation, knowledge management, process of innovation control and supervision, and project management scope. These, among some other cases could provide an instrument used for medium and higher level structures, or can be studied on its own as separate models. Modeling emerging behavior using ABM could potentially be a method of choice at the lowest level of abstraction of R&D governance. One could get insight into new order, new patterns, and structures of R&D by learning from interaction of individual agents or groups. One possible way of implementation would consist of distributed agents in form of humans, information storages or groups working together attributed by goals, knowledge and imposing adaptive tension on one another (Kitaygorodskaya, et al., 2005). Outcome from emerging behavior can be greater than sum of its components e.g. knowledge generation; tacit knowledge depends on interaction and direct experience, whereas explicit knowledge could be easily transferred across distance. Modeling spectrum of interactions would allow for better understanding of complex emerging behaviors, along with some quantification (assumed possible) of expected knowledge increase. Different structures at different levels in enterprise or government along with imposed interactions could for example increase chances for innovation generation. As our understanding raise, emerging behavior capabilities of ABM can provide for building new theories, which in turn can enable better structures for example motivation and knowledge sharing that enables innovation generation. As discussed, it is visible that ABM can be mostly suitable for theoretical development, which should be later validated using methods that are more empirical, as ABM is generally difficult to validate. ABM can also be suitable at higher levels of R&D governance modeling, where individual agents mimic organization, society, region, country, religion etc. and agents' independence can provide for different perspectives comparing to one obtained from e.g. SD method. DES seems appropriate to use where looking and the R&D governance from the processing perspective. Every big enterprise or government entity has to manage many R&D projects, evaluate them, and properly allocate resources. DES allows to model R&D endeavors throughout their stages in life cycle. DES can be apply to model, for instance, R&D projects as entities going through their life-cycle starting from its request for approval, which can end up as a rejection, then development phase through stages at which it is possible for project to be rejected from continuation due to e.g. funding limits. This would let resources allocated to this project to be moved to another project, and therefore increasing its chances of being successful. Finally, project's output needs to be incorporated into production then into market, or other type of environment for which it was created. We cannot forget about product maintenance tasks throughout its life cycle. R&D process calls for a method that can easily incorporate its stochastic character, and DES can surely deliver that. Ability to mimic the logic of the organization, and evaluate alternative structural alignments, for example, centralized versus decentralized makes DES especially valuable. In addition, stochastic character of R&D endeavors makes DES a very powerful tool that could be used for assessment of e.g. outsourcing versus "in-house" and prediction on estimated return of investment. DES can go to the very detailed level of abstraction depending on need, and it is flexible and easy when managing different levels of aggregation. In addition, many commercial

DES packages provide techniques that can mimic different methods that one would like to incorporate into model like SD and some even ABM. SD is the method that can be very useful especially at the enterprise modeling level when trying to capture big-picture or global scope of the problem. SD should be used where there exists a feedback loop, and it is hard to explain phenomenon by intuition itself, which sounds very much like issues from R&D governance range. Problems of choosing between some options are the most appropriate to address with SD, as it is often hard to use model as providing exact numerical information for us. This method could be particularly appropriate to a long time-span models. SD structure is based on differential equations, but most of the commercial simulation packages allow incorporating stochastic behavior into the model. Using this method for testing policies and future trends can be invaluable when trying to plan in long term, which is specifically related to R&D governance type decision-making. Modeling competing environment allows for a comprehensive perspective of what is needed to obtain desired outcome. Important advantage of SD is that data used to model it do not have to be ample or even direct. Often one would like to learn the system and restore its behavior without direct knowledge about it, by building relations that are only correlated with it or related to the phenomenon. Another area where SD may be appropriate is to learn about how significant are roles of parts of a complex system on functioning of the whole body, e.g. not only how outsourcing affects single project outcome, but also what are the long time implications.

To manage different levels of abstraction one could potentially increase effectiveness by using nested or sometimes called hybrid models. These models use different methodologies that could be arranged hierarchically depending on type of problem the model is to address. Multilayer of ABM, DES, and SD methods could be constructed allowing communication and so influencing each other. Some structures can be also embedded by another if that reflects system structure. Many M&S practitioners try to combine both ABM and SD and that is often a good idea. These two methods can work well together and provide good combination of e.g. lower level ABM with SD capturing larger view of the system. But alignment of ABM-SD can be multi-layered, for instance, some part of agent behavior (internals of agent) can be described using differential equations from SD, then the environment level described in ABM terms with agents at different level of aggregation, and on top of that SD world-view, capturing global dependencies. Some of commercial as well as free software packages provide for that configuration e.g. ANYLOGIC, and NETLOGO. In the diagram authors proposed example of how configuration of hybrid model could look like.

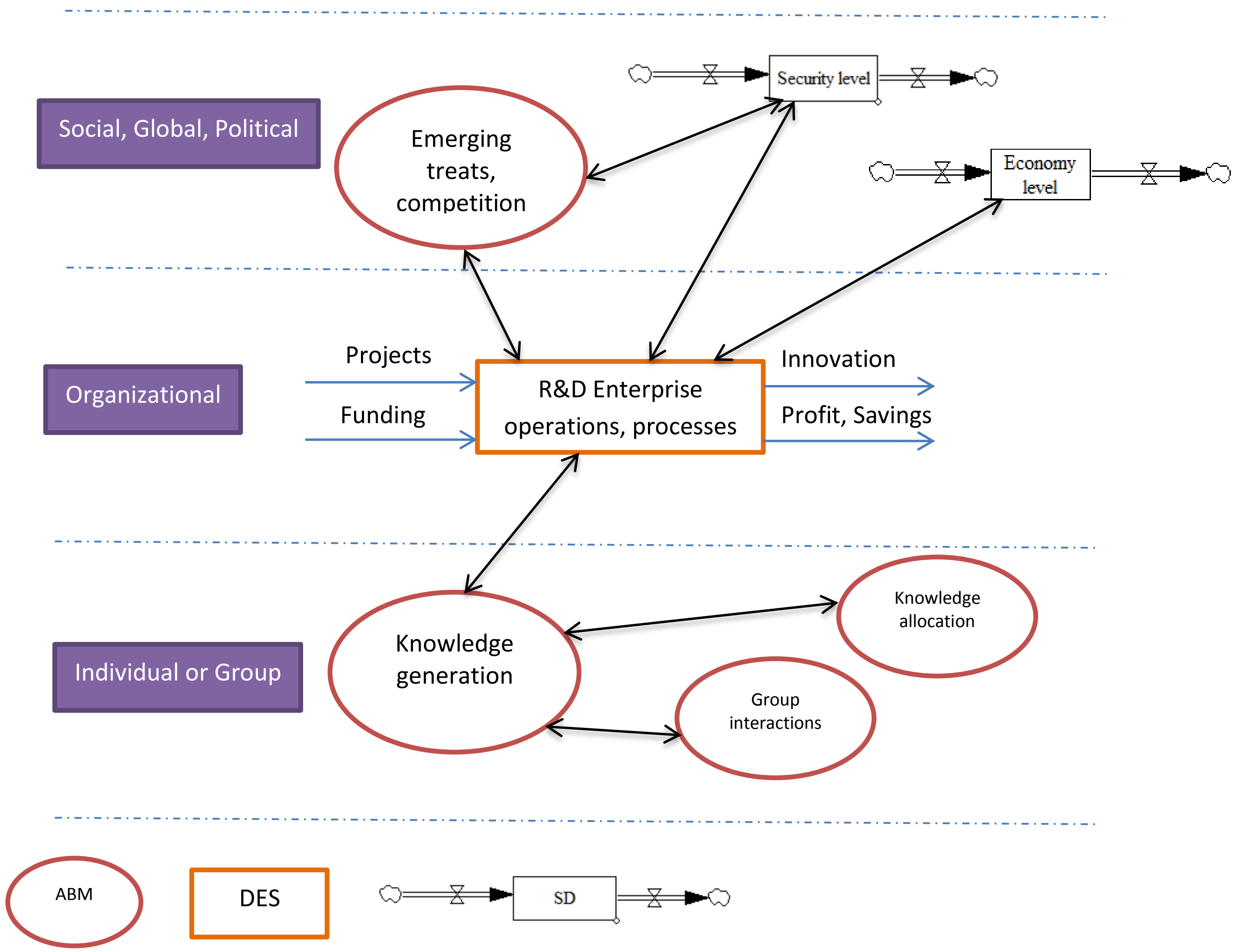


Figure 1. One of possible configurations of hybrid model of R&D governance

5. Model example

Idea to build a model of border security originated while watching news related to illegal cargo smuggling and illegal border crossing issues. Estimation number of illegal immigrants found from official and unofficial sources range from 10 to 23 million people in 2010. This of course can be under or over estimation causing a significant problem in estimations of many statistics and biasing true picture of USA. Proposed by authors conceptual model can be divided into three major segments as indicated in the Figure 2 by colors: red color is used for border security, knowledge diffusion, and illegal border activities; green color depicts money related dependencies, and finally blue color shows different groups of people involved.

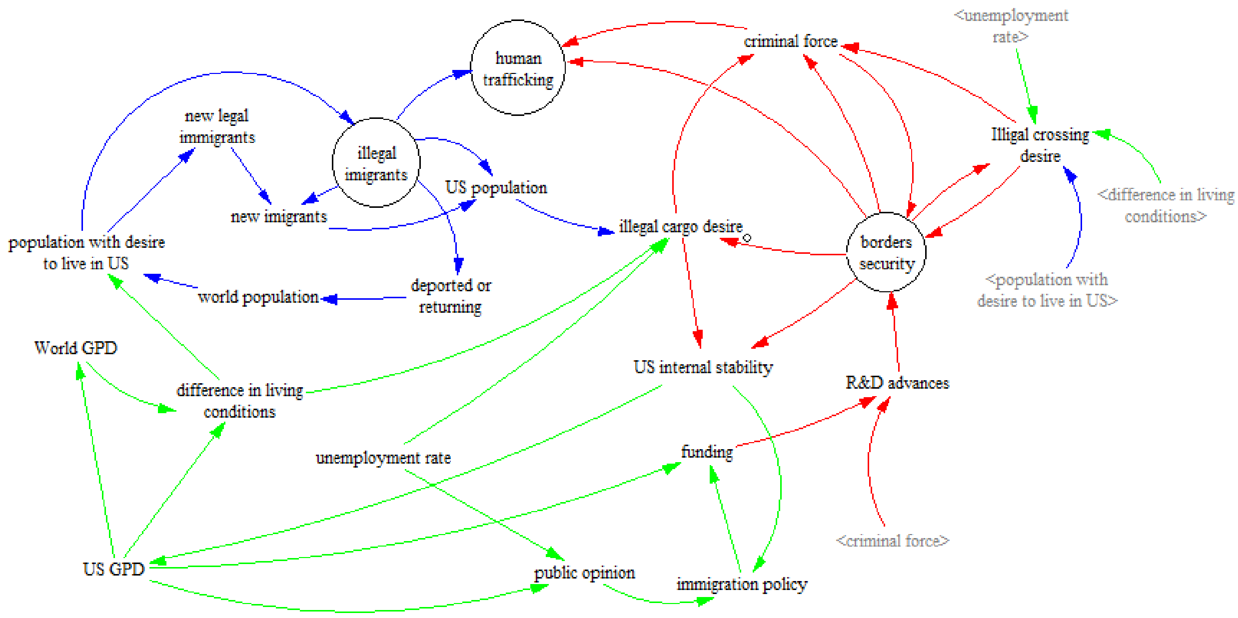


Figure 2: Conceptual model capturing border security and illegal immigration variables

First, initial research of data for population growth rates, legal immigration rates, GPDs, federal R&D funding and particular Department of Homeland Security was conducted. Obtained data used to derive trend lines using MS Excel. These trends could be of course altered to conduct sensitivity analysis of alternative scenarios. Now authors would like to present short description of more interesting parts of this SD model. Red section from conceptual model provides suggestion into assessment on funding for DHS R&D and its effect on output. R&D section is built on concept of knowledge generation (Figure 3) as a process of two competing environments: one being R&D environment where conducted or contracted out research request by DHS is supposed to help eliminate illegal activity around the border, second environment represents knowledge of criminal force to continue with their illegal endeavor. There is an assumption on time delay between these environments with reaction time needed to come up with some new methods. Authors hope that there is an effective way of breaking this pattern, but for a simplification, breaking the pattern is not considered in this model. Future study of that phenomenon using agent-based modeling approach is foreseen, and it would possibly allow assessing how fast criminal force can adjust to changes imposed by border security advancements and vice versa. Insight from future study could increase accuracy of current model.

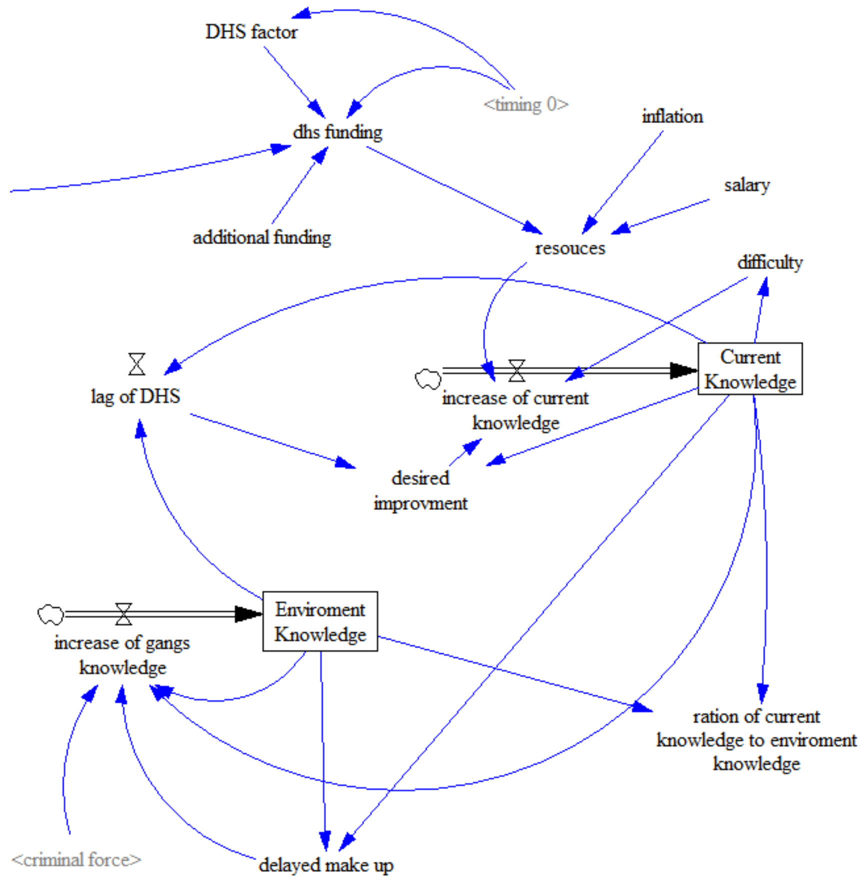


Figure 3: Concept of knowledge generation via competing environments

Another part of "red color" (Figure 4) model is aimed to capture abstract desire/demand for illegal "products" from society as well as illegal crossing desire of foreigners. This will in turn affect border security and propel creating additional criminal force, willing to "fulfill" demand of the people. In contrast, increased border security can decrease effective smuggling and illegal border crossing, which in turn decreases money flow used by criminal force. This could be theoretically a solution, but these individuals will most likely try to overcome that "hardship", and come up with a new strategy.

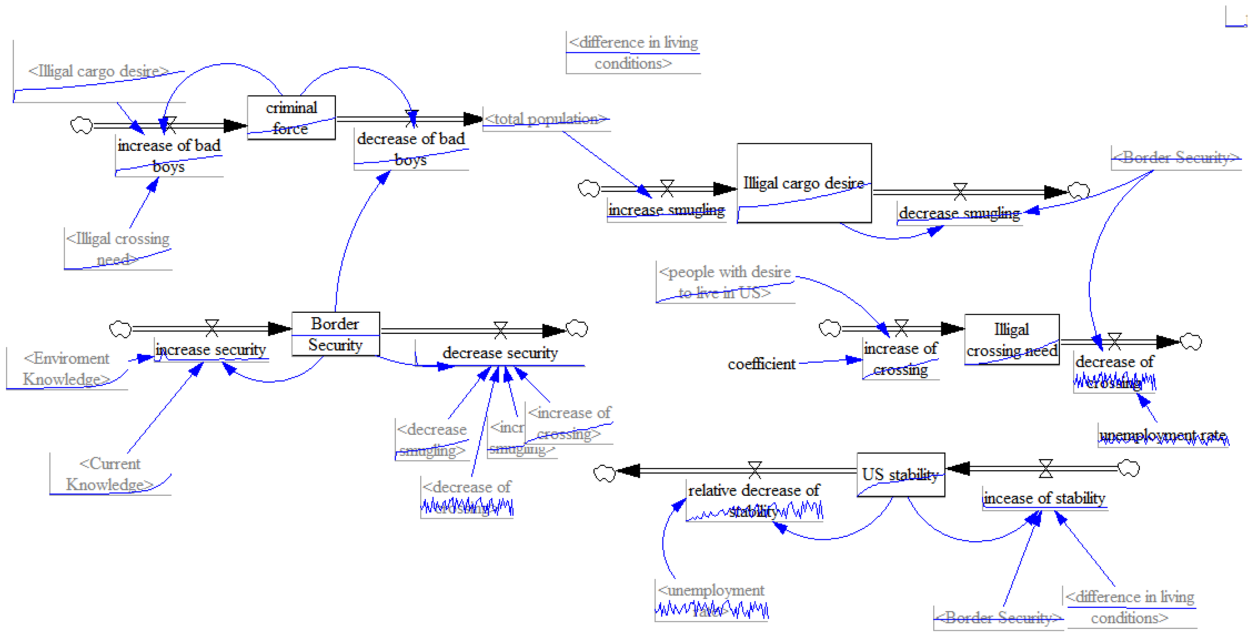


Figure 4: Part of model with trying to capture reasons for illegal cargo and crossing

Last part that authors would like to present here is the "blue color" from conceptual model (Figure 5), which focuses on groups of people. Here, key elements for our analysis are represented by number of illegal immigrants and level of human trafficking.

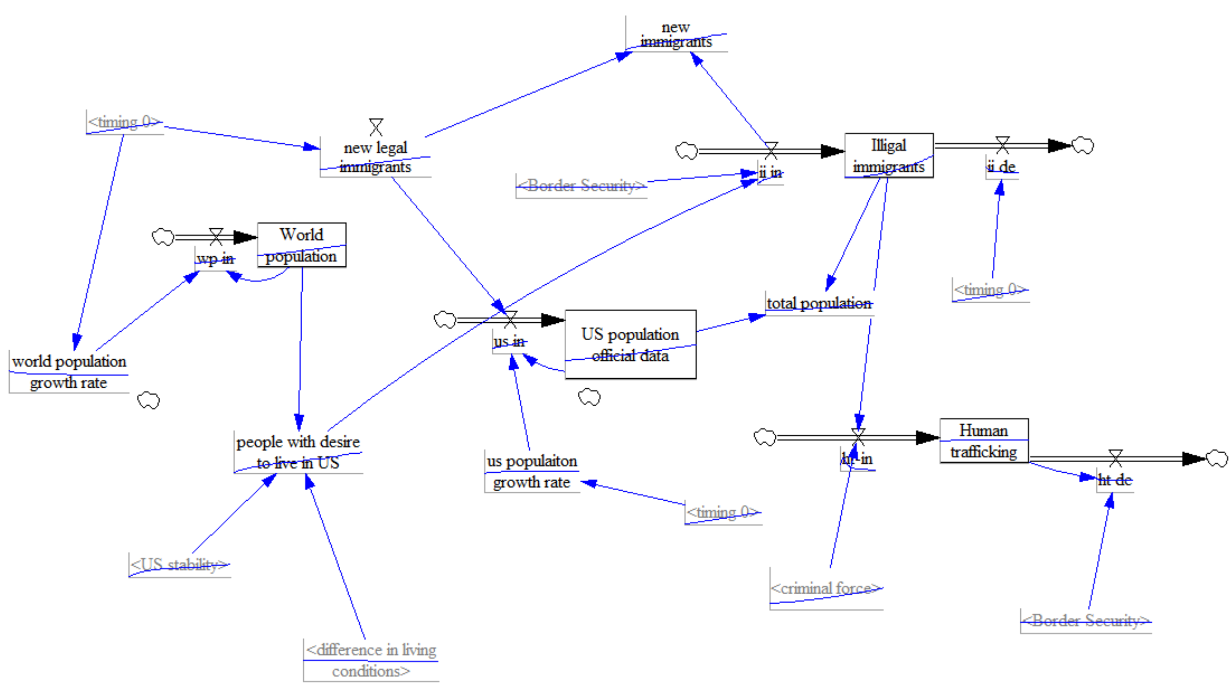


Figure 5: Relations between different groups of people, illegal immigration, and human trafficking

Base for obtaining those is assumption that major of illegal immigrants are those pursuing crossing the borders or overstaying for economic reasons. Hence, authors were looking at estimating differences between gross products per capita (GPD) as a factor of estimating people with desire to live in USA. Additionally, human trafficking is in authors' opinion strongly correlated with illegal immigration. Model is in the late development phase, which mainly consists of tuning and improving look up graphs. The "blue part" of the model still needs some more conceptual level work. Mainly, authors are trying to find additional variables that can increase accuracy of illegal immigrants and human trafficking levels estimation. Initial runs of the model shows potential for gaining valuable insight into border security including estimation of illegal activities and ability to test different approaches to counteract them.

6. Conclusions

This paper presents insights into governance level of R&D. Authors moved from theoretical knowledge of R&D governance, through description of most popular approaches available for Modeling and Simulation practitioners, ending up with discussion on how to model R&D. In addition, authors showed an example of SD model that could be used by DHS. Built model is not at its final stage, but produced meaningful results. Future work is foreseen that would incorporate hybrid modeling to better capture complexity of R&D governance.

REFERNECES

Argyres, S. N. 2004. R&D, Organization structure, and the development of corporate technological knowledge. forthcoming, *Strategic Management Journal.* Special issue on "The Global Acquisition, Protection and Leveraging of Technological Competencies"

Bodner, D. A.,W. B. Rouse, and M. J. Pennock. 2005. Using simulation to analyze R&D value creation. In *Proceedings of the 2005 Winter Simulation Conference,* eds. M.E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 1006-1013.

Bonabeau, E. 2002. Agent-based modeling: Methods and techniques for simulating human systems. In *Proceedings of the National Academy of Sciences of the United states of America* 99: 7280-7287.

Chahal, K., and T. Eldabi. 2008. Applicability of hybrid simulation to different modes of governance in UK healthcare. In *Proceedings of the 2008 Winter Simulation Conference,* eds. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler, 1469-1477.

Chaturvedi , R., S. Mellema, A. Chaturvedi, M. Mulpuri, G. Pinczuk. Continuous Validation Framework: A Case Study of SEAS and Afghanistan. *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC). Volume: 2008 (Conference Theme: Learn. Train. Win!)*

EIRMA. 2003. Assessing R&D effectiveness. *Working Group Reports* WG62 Report: 1-26.

García, J. M. 2006. *Theory and practical exercises of system dynamics*. Barcelona, Spain.

Hester, P. T., and A. Tolk. 2010. Applying methods of the M&S spectrum for complex system engineering. *Emerging Applications of M&S in Industry and Academia (EAIA)*, Society for Modeling and Simulation, San Diego, CA, SCS Simulation Series, 17-24.

Keating, C., P. Hester, T. Meyers, and B. Calida. 2010. R&D system governance: an emerging perspective. In *Proceedings of R&D Management Conference: Information Imagination & Intelligence in R&D Management.* Manchester, London: University of Manchester.

Kelton, W, D., R. P. Sadowski, D. T. Sturrock. 2007. *Simulation with Arena.* New York: McGraw-Hill, Inc.

Kitaygorodskaya, N. 2005. A framework for agent-based modeling of communication and knowledge networks in corporate R&D. In *Computational Cybernetics (3rd International Conference),* 133-138.

Middleton, V. 2010. Simulating small unit military operations with agent-based models of complex adaptive systems. In *Proceedings of the 2010 Winter Simulation Conference,* eds. B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan.

Law, A. M., 2007. *Simulation Modeling and Analysis.* New York: McGraw-Hill, Inc.

Pearce, A. R., and K. L. Sanford Bernhardt. 2010. Sustainability and socio-enviro-technical systems: modeling total cost of ownership in capital facilities. In *Proceedings of the 2010 Winter Simulation Conference,* eds. B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan.

Medical and Healthcare M&S

Track Lead: Dr. Gianluca DeLeo, VMASC

Judges: Rebecca C. Britt MD, EVMS, Dr. Christopher J. Garcia, MYMIC LLC

Preliminary Modeling Solutions for Prevention of Medicare Insolvency for the Baby Boomer Generation
Author(s): Gregory Scott
2nd Place Gene Newman Award Winner, Medical & Healthcare Track

Heart sound cancellation using an adaptive wavelet thresholding method
Author(s): Hoang-Anh T Nguyen, Jiang Li, Frederic Mckenzie, Ender Oguslu

An Improved Virtual Operating Room Training Scenario Model
Author(s): Rebecca Kennedy and Kellie Kennedy
3rd Place Gene Newman Award Winner, Medical & Healthcare Track

Improving nursing infection control practices in a virtual intensive care unit
Author(s): Lydia Wigglesworth-Ballard
1st Place Gene Newman Award Winner, Medical & Healthcare Track

Using Modeling and Simulation to Improve Oral Health Services Delivery in Norfolk Public Health District Dental Clinic, Little Creek
Author(s): Mohammad Alzahrani, Holly Gaff, Deanne Shuman, Rani Kady
Abstract Only

Preliminary Modeling Solutions for Prevention of Medicare Insolvency for the Baby Boomer Generation

Gregory T. Scott, MHA, CMRP, FAHRMM, PhD student, Health Services Research Old Dominion University, College of Health Sciences

Abstract – The purpose of this current study is to examine Medicare insolvency for the baby boomer generation representing years 2010 to 2030 and the mechanisms for prevention using modeling and simulation (M&S). The study examines cost savings initiatives as well as age increases for eligibility and tax increases as measures to prevent future insolvency. These measures are used as individual and group measures to identify prevention techniques enabling the Medicare Trust Fund from impending insolvency.

INTRODUCTION

Medicare is health insurance for people 65 years and older and under the age of 65 with certain disabilities and those with end stage renal disease. Eligibility also requires that beneficiaries must have entered the United States (US) legally and have lived within the U.S for 5 years [1]. Medicare has two problems facing the trust fund. The US has recently been in an economic recession that has created more demand for public health while at the same time revenues from taxes are down. Legislators are looking for ways to sustain Medicare by reducing health care spending in general through interventions that may include growth of pay-for-performance programs, cuts in payments to Medicare Advantage programs and promotion of electronic health records, sophisticated chronic disease management, preventive health care and the medical home [3]. The cost of healthcare continues to create insolvency issues as the inflation rate for healthcare is higher than non-healthcare inflation [2]. Medicare insolvency has become a growing issue over the past decade as beneficiary payments will start to exceed funding received from payroll taxes as the baby boomer generation becomes eligible for Medicare [2]. The trust fund can still "prop up" the Medicare fund possibly for another decade or two before the trust fund becomes insolvent. There are many ideas that have been mentioned through the years to resolve the funding issue and ensure that Medicare will remain available to all people who are 65 years of age and older, those at the end stage of renal disease and social security disability recipients [1]. Measures will have to be applied as extending the age of eligibility, fraud reductions, tax increases. These measures will be painful for the American people but are necessary to prevent loss of health care for Medicare recipients. Americans could also be burdened by health care cost for their parents who are 65 years or older or dependents receiving social security disability or at the end stages of renal disease if Medicare fails to provide the care necessary due to budget funding cuts to prevent insolvency if new measures are not introduced in the near future. The research will establish a preliminary review of mechanisms that could slow or prevent the Medicare Trust Fund from insolvency over the next twenty years represented by the heavy influx of baby boomers receiving Medicare benefits.

The purpose

The purpose of this study is to examine Medicare insolvency prevention initiatives using modeling and simulation. Examination will occur using individual variables as age increases, tax increases and fraud reductions. The variables will be used independently and in combinations to determine if Medicare solvency can be obtained.

Modeling and Simulation in Healthcare

"Healthcare reform is an important issue for the American public, the healthcare industry, policy analysts, and politicians. However, it is difficult for policymakers in either the public or private sector to reliably assess the intended and unintended consequences of such reforms. Until now, it has been practically impossible to have a constructive, objective dialogue about the relative merits of different potential solutions [3]." Modeling and simulation for this study is meant to influence the American public and give potential options to politicians regarding what some argue is one of the largest financial issues facing the federal budget. Models can be constructed and simulated to help identify paths of resolve.

Modeling and Simulation Application
There will be five Monte Carlo simulations computed for this current study illustrating revenue changes related to age and tax increases and fraud reductions. Monte Carlo methods are useful for modeling phenomena with uncertainty as future life expectancy in this current study for the baby boomer generation. Monte Carlo simulations will illustrate the future risk to the Medicare Trust Fund and identify solutions to prevent insolvency through modeling and simulation. Monte Carlo simulations can be used to determine an accurate prediction of a risk in economics and finance rather than pure human intuition [8].

Statement of Problem
Medicare is a critical component of American healthcare insurance. People think of Medicare as insurance only for the elderly, but in reality Medicare insures people who are receiving Social Security disability under the age of 65 and those who are deemed at the end stages of renal disease [3]. The funding issue of Medicare has occurred because there are less people paying into Medicare every year than those receiving benefits [2]. The lack of even payments and benefits paid create an environment of certain future insolvency.

Significance of the Study
This study will use modeling and simulation to illustrate prevention steps in regards to impending Medicare insolvency. Anyone who pays taxes or receives Medicare or has a relative who receives Medicare should be extremely concerned. The models will attempt to identify the path of least resistance to prevent hardships on the American tax payer.

The prevention steps are recommendations to slow or stop catastrophic financial burdens. Without these measures the federal government will have to borrow money, reduce defense, reduce education spending and possibly reduce Social Security. This could be the largest issue in American finance if the Medicare Trust Fund depletes. The Medicare Trust Fund as of 2010 resulted in a $2 billion dollar shortfall with revenues of $507 billion and cost of $509 billion.

Research Questions
Will tax increases resolve the Medicare Trust Fund from insolvency by year 2030 for the baby boomer generation? Will age increases resolve the Medicare Trust Fund from becoming insolvent by year 2030 for the baby boomer generation? Will less fraud prevent the Medicare Trust from insolvency by year 2030 for the baby boomer generation?

Directional Hypotheses
Medicare age increases for eligibility will slow or prevent insolvency. Medicare tax increases will slow or prevent insolvency. Reducing fraud within Medicare will slow or prevent insolvency. All measures may prevent the Medicare Trust Fund from insolvency by year 2030 for the baby boomer generation.

REVIEW OF THE LITERATURE
Increasing taxes is one measure than can be addressed that could be used to prevent Medicare insolvency in the future. The current amount of payroll tax is 1.45% for the employee and 1.45% for the employer with a combined 2.9% of total payroll taxes collected for Medicare [2]. A measure as increasing the payroll tax by 2% can be applied to illustrate additional funding to Medicare. This measure creates a positive impact by increasing taxes for the employee and employer by 2% for the Medicare Trust Fund, but the trust remains insolvent in 2030 with a $1.371 trillion deficit. The tax increase model is illustrated as measure 1.

Medicare Tax Increase Impact for the Medicare Trust Fund (measure # 1)

Start Age 67 Year	Medicare Annual Cost (Billions)	Medicare Annual Cost after Post Inter-vention + Growth (Billions)	Medicare Revenue (Billions)	Average Starting Beneficiary Cost $10,400 Beneficiaries	Medicare Revenue Growth	Medicare Average Annual Beneficiary Growth	Beneficiary Death Expectancy	Average Beneficiary Growth Less Death	Health-care Inflation Average	Tax Increase Revenue Gain 2% (Billions)	Medicare Trust Fund (Billions)	Current Cash Flow
2010	$ 509		$507	46,300,000	5.50%				6.90%		$ 507	
2011	$ 544	$ 491	$535	49,214,975	5.50%	3,500,000	585,025	2,914,975	6.90%	$ 83	$ 550	$ 43
2012	$ 582	$ 527	$564	52,089,350	5.50%	3,500,000	625,625	2,874,375	6.90%	$ 85	$ 588	$ 37
2013	$ 622	$ 565	$595	54,918,645	5.50%	3,500,000	670,705	2,829,295	6.90%	$ 86	$ 618	$ 30
2014	$ 665	$ 606	$628	57,702,860	5.50%	3,500,000	715,785	2,784,215	6.90%	$ 88	$ 641	$ 23
2015	$ 711	$ 649	$663	60,436,955	5.50%	3,500,000	765,905	2,734,095	6.90%	$ 90	$ 654	$ 13
2016	$ 760	$ 696	$699	63,117,220	5.50%	3,500,000	819,735	2,680,265	6.90%	$ 92	$ 658	$ 3
2017	$ 812	$ 746	$738	65,739,455	5.50%	3,500,000	877,765	2,622,235	6.90%	$ 93	$ 649	$ (8)
2018	$ 868	$ 799	$778	68,299,250	5.50%	3,500,000	940,205	2,559,795	6.90%	$ 95	$ 628	$ (21)
2019	$ 928	$ 857	$821	70,792,160	5.50%	3,500,000	1,007,090	2,492,910	6.90%	$ 97	$ 592	$ (36)
2020	$ 992	$ 918	$866	73,213,950	5.50%	3,500,000	1,078,210	2,421,790	6.90%	$ 99	$ 540	$ (52)
2021	$ 1,060	$ 984	$914	75,560,280	5.50%	3,500,000	1,153,670	2,346,330	6.90%	$ 101	$ 470	$ (70)
2022	$ 1,134	$ 1,054	$964	77,823,310	5.50%	3,500,000	1,236,970	2,263,030	6.90%	$ 103	$ 380	$ (90)
2023	$ 1,212	$ 1,129	$1,017	80,004,265	5.50%	3,500,000	1,319,045	2,180,955	6.90%	$ 105	$ 268	$ (112)
2024	$ 1,295	$ 1,210	$1,073	82,095,060	5.50%	3,500,000	1,409,205	2,090,795	6.90%	$ 107	$ 131	$ (137)
2025	$ 1,385	$ 1,296	$1,132	84,090,935	5.50%	3,500,000	1,504,125	1,995,875	6.90%	$ 110	$ (33)	$ (164)
2026	$ 1,480	$ 1,388	$1,194	85,987,270	5.50%	3,500,000	1,603,665	1,896,335	6.90%	$ 112	$ (227)	$ (194)
2027	$ 1,582	$ 1,487	$1,260	87,779,340	5.50%	3,500,000	1,707,930	1,792,070	6.90%	$ 114	$ (455)	$ (227)
2028	$ 1,692	$ 1,593	$1,329	89,462,490	5.50%	3,500,000	1,816,850	1,683,150	6.90%	$ 116	$ (718)	$ (264)
2029	$ 1,808	$ 1,706	$1,402	91,032,205	5.50%	3,500,000	1,930,285	1,569,715	6.90%	$ 119	$ (1,022)	$ (304)
2030	$ 1,933	$ 1,827	$1,479	92,484,950	5.50%	3,500,000	2,047,255	1,452,745	6.90%	$ 121	$ (1,371)	$ (348)

Figure 1: Table Source: Gregory Scott

Medicare solvency could be gained by elevating the ages of beneficiaries similar to what Social Security has already done. Life expectancy has increased for the baby boomer generation since Medicare was passed by Congress [4]. Congress passed Medicare based on life expectancies in the 1960's with the intention that most people would only receive Medicare for five to seven years as their life expectancy did not exceed their early 70's [4]. Life expectancy today (2011) for most Americans is reaching closer to 80 years of age creating an additional burden on Medicare as cost is being incurred that Congress did not expect when the Medicare bill passed in the 1960's [5]. The solution could simply be to increase the age to 67 years of age to receive Medicare instead of the current age of 65. This measure could put the Medicare age closer to the original life expectancy that did not exceed 5 to 9 years as Congress intended. This approach reduces the Medicare Trust Fund deficit, but the fund is left with a $2.503 trillion dollar short fall for year 2030. The age increase models are illustrated as measure 2 and 3.

Cost reductions are needed for Medicare to slow funding issues associated with the trust fund. Medicare over head could be reduced by being more aggressive in regards to fraud recoupment to slow or prevent Medicare insolvency with no change to taxes or age of eligibility [6]. These measures would be the most pain free action to take as the public would not be directly impacted. Cost associated to fraud could be reduced saving billions of dollars annually for Medicare [6]. Measures to reduce cost are the most popular with the public and politicians as they satisfy the public and do not risk votes needed for re-election. While this is a very popular approach to Medicare insolvency prevention, the reality is that these steps would likely slow insolvency but would not necessarily prevent insolvency from occurring in the near future by themselves. The fraud reduction measure reduces the Medicare Trust Fund deficit, but the fund is left with a $1.929 trillion deficit for year 2030. The model is illustrated as measure 4.

Life Expectancy (measure 2)

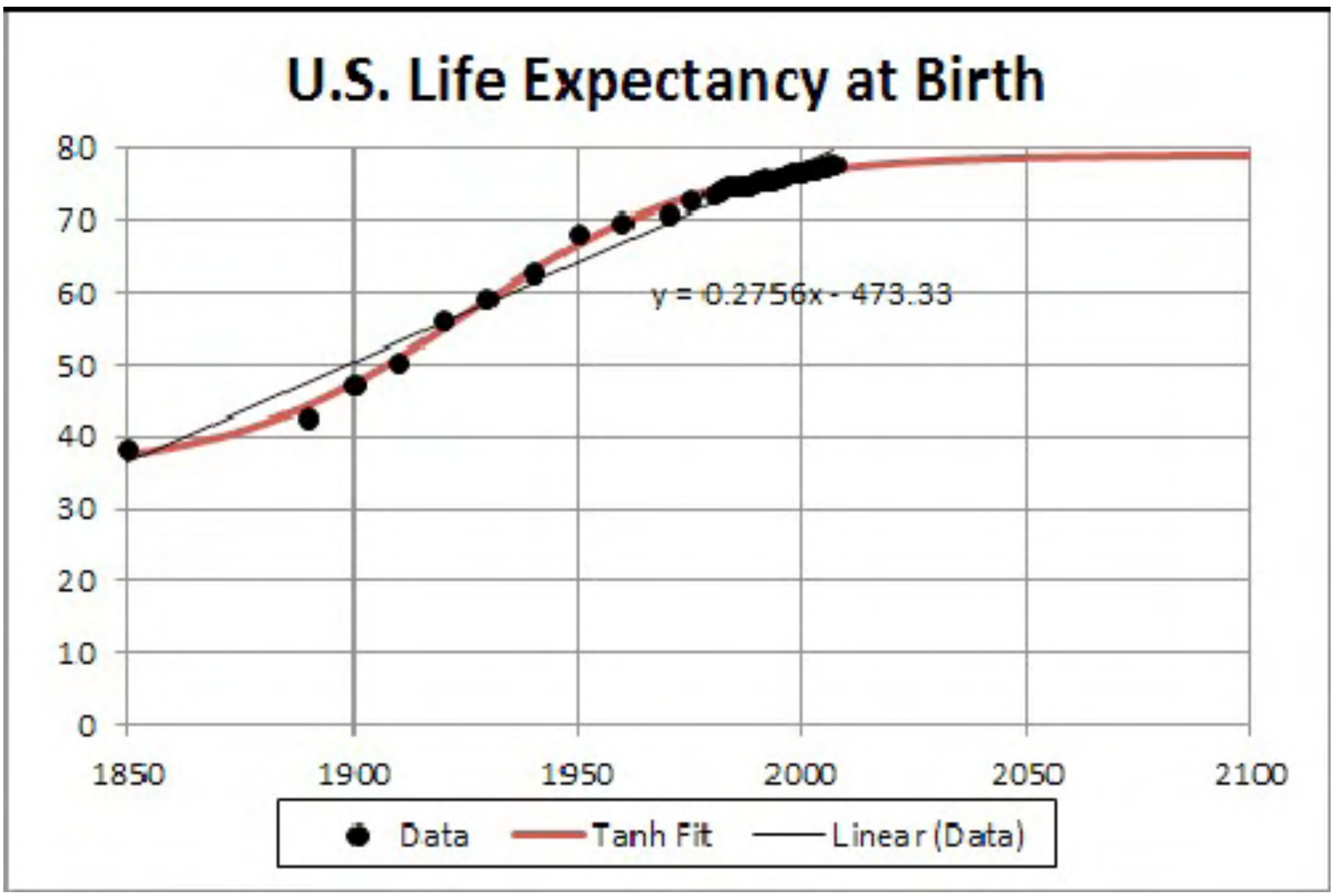


Figure 2: Life Expectancy at Birth. Source: United States Census Bureau, 2009

Age Increase Financial Impact for the Medicare Trust Fund (measure 3)

Start Age 67 Year	Medicare Annual Cost (Billions)	Medicare Annual Cost after Post Inter-vention + Growth (Billions)	Medicare Revenue (Billions)	Average Starting Beneficiary Cost $10,400 Beneficiaries	Medicare Revenue Growth	Medicare Average Annual Beneficiary Growth	Beneficiary Death Expectancy	Average Beneficiary Growth Less Death	Health-care Inflation Average	Age Increase to 67 Savings (Billions)	Medicare Trust Fund (Billions)	Current Cash Flow
2010	$ 509		$507	46,300,000	5.50%				6.90%		$ 507	
2011	$ 544	$ 538	$535	49,214,975	5.50%	3,500,000	585,025	2,914,975	6.90%	$36	$ 504	$ (3)
2012	$ 582	$ 574	$564	52,089,350	5.50%	3,500,000	625,625	2,874,375	6.90%	$37	$ 494	$ (10)
2013	$ 622	$ 613	$595	54,918,645	5.50%	3,500,000	670,705	2,829,295	6.90%	$38	$ 476	$ (18)
2014	$ 665	$ 655	$628	57,702,860	5.50%	3,500,000	715,785	2,784,215	6.90%	$39	$ 449	$ (27)
2015	$ 711	$ 700	$663	60,436,955	5.50%	3,500,000	765,905	2,734,095	6.90%	$39	$ 412	$ (37)
2016	$ 760	$ 747	$699	63,117,220	5.50%	3,500,000	819,735	2,680,265	6.90%	$40	$ 364	$ (48)
2017	$ 812	$ 798	$738	65,739,455	5.50%	3,500,000	877,765	2,622,235	6.90%	$41	$ 303	$ (61)
2018	$ 868	$ 853	$778	68,299,250	5.50%	3,500,000	940,205	2,559,795	6.90%	$42	$ 228	$ (75)
2019	$ 928	$ 911	$821	70,792,160	5.50%	3,500,000	1,007,090	2,492,910	6.90%	$43	$ 138	$ (90)
2020	$ 992	$ 974	$866	73,213,950	5.50%	3,500,000	1,078,210	2,421,790	6.90%	$44	$ 30	$ (108)
2021	$ 1,060	$ 1,040	$914	75,560,280	5.50%	3,500,000	1,153,670	2,346,330	6.90%	$44	$ (97)	$ (127)
2022	$ 1,134	$ 1,112	$964	77,823,310	5.50%	3,500,000	1,236,970	2,263,030	6.90%	$45	$ (245)	$ (148)
2023	$ 1,212	$ 1,188	$1,017	80,004,265	5.50%	3,500,000	1,319,045	2,180,955	6.90%	$46	$ (416)	$ (171)
2024	$ 1,295	$ 1,270	$1,073	82,095,060	5.50%	3,500,000	1,409,205	2,090,795	6.90%	$47	$ (613)	$ (197)
2025	$ 1,385	$ 1,358	$1,132	84,090,935	5.50%	3,500,000	1,504,125	1,995,875	6.90%	$48	$ (839)	$ (226)
2026	$ 1,480	$ 1,451	$1,194	85,987,270	5.50%	3,500,000	1,603,665	1,896,335	6.90%	$49	$ (1,096)	$ (257)
2027	$ 1,582	$ 1,551	$1,260	87,779,340	5.50%	3,500,000	1,707,930	1,792,070	6.90%	$50	$ (1,387)	$ (291)
2028	$ 1,692	$ 1,658	$1,329	89,462,490	5.50%	3,500,000	1,816,850	1,683,150	6.90%	$51	$ (1,716)	$ (329)
2029	$ 1,808	$ 1,773	$1,402	91,032,205	5.50%	3,500,000	1,930,285	1,569,715	6.90%	$52	$ (2,087)	$ (371)
2030	$ 1,933	$ 1,895	$1,479	92,484,950	5.50%	3,500,000	2,047,255	1,452,745	6.90%	$53	$ (2,503)	$ (416)

Figure 3: Table Source: Gregory Scott

Fraud Reduction Impact for the Medicare Trust Fund (measure 4)

Start Age 67 Year	Medicare Annual Cost (Billions)	Medicare Annual Cost after Post Inter-vention + Growth (Billions)	Medicare Revenue (Billions)	Average Starting Beneficiary Cost $10,400 Beneficiaries	Medicare Revenue Growth	Medicare Average Annual Beneficiary Growth	Beneficiary Death Expectancy	Average Beneficiary Growth Less Death	Health-care Inflation Average	Fraud Reduc-tions (Billions)	Medicare Trust Fund (Billions)	Current Cash Flow
2010	$ 509		$507	46,300,000	5.50%				6.90%		$ 507	
2011	$ 544	$ 514	$535	49,214,975	5.50%	3,500,000	585,025	2,914,975	6.90%	$60	$ 527	$ 20
2012	$ 582	$ 550	$564	52,089,350	5.50%	3,500,000	625,625	2,874,375	6.90%	$61	$ 541	$ 14
2013	$ 622	$ 589	$595	54,918,645	5.50%	3,500,000	670,705	2,829,295	6.90%	$62	$ 548	$ 7
2014	$ 665	$ 630	$628	57,702,860	5.50%	3,500,000	715,785	2,784,215	6.90%	$64	$ 546	$ (2)
2015	$ 711	$ 674	$663	60,436,955	5.50%	3,500,000	765,905	2,734,095	6.90%	$65	$ 535	$ (11)
2016	$ 760	$ 721	$699	63,117,220	5.50%	3,500,000	819,735	2,680,265	6.90%	$66	$ 512	$ (22)
2017	$ 812	$ 772	$738	65,739,455	5.50%	3,500,000	877,765	2,622,235	6.90%	$68	$ 478	$ (34)
2018	$ 868	$ 826	$778	68,299,250	5.50%	3,500,000	940,205	2,559,795	6.90%	$69	$ 431	$ (48)
2019	$ 928	$ 884	$821	70,792,160	5.50%	3,500,000	1,007,090	2,492,910	6.90%	$70	$ 368	$ (63)
2020	$ 992	$ 945	$866	73,213,950	5.50%	3,500,000	1,078,210	2,421,790	6.90%	$72	$ 289	$ (79)
2021	$ 1,060	$ 1,012	$914	75,560,280	5.50%	3,500,000	1,153,670	2,346,330	6.90%	$73	$ 190	$ (98)
2022	$ 1,134	$ 1,083	$964	77,823,310	5.50%	3,500,000	1,236,970	2,263,030	6.90%	$75	$ 72	$ (119)
2023	$ 1,212	$ 1,158	$1,017	80,004,265	5.50%	3,500,000	1,319,045	2,180,955	6.90%	$76	$ (70)	$ (141)
2024	$ 1,295	$ 1,240	$1,073	82,095,060	5.50%	3,500,000	1,409,205	2,090,795	6.90%	$78	$ (236)	$ (167)
2025	$ 1,385	$ 1,326	$1,132	84,090,935	5.50%	3,500,000	1,504,125	1,995,875	6.90%	$79	$ (431)	$ (195)
2026	$ 1,480	$ 1,419	$1,194	85,987,270	5.50%	3,500,000	1,603,665	1,896,335	6.90%	$81	$ (656)	$ (225)
2027	$ 1,582	$ 1,519	$1,260	87,779,340	5.50%	3,500,000	1,707,930	1,792,070	6.90%	$82	$ (915)	$ (259)
2028	$ 1,692	$ 1,625	$1,329	89,462,490	5.50%	3,500,000	1,816,850	1,683,150	6.90%	$84	$ (1,211)	$ (296)
2029	$ 1,808	$ 1,739	$1,402	91,032,205	5.50%	3,500,000	1,930,285	1,569,715	6.90%	$86	$ (1,548)	$ (337)
2030	$ 1,933	$ 1,861	$1,479	92,484,950	5.50%	3,500,000	2,047,255	1,452,745	6.90%	$87	$ (1,929)	$ (382)

Figure 4: Table Source: Gregory Scott

The last measure is to combine measures 1, 3 and 4 to prevent the Medicare Trust Fund from insolvency for the baby boomer generation by year 2030. These measures which include tax and age increases in conjunction with fraud reductions increase Medicare revenue to sustain payments for the baby boomer generation. All measures create a surplus of $972 billion in year 2030. The model is illustrated as measure 5. Medicare Trust Fund insolvency can be prevented if taxes and age are increased and fraud is reduced. This intervention creates a surplus of $972 billion in year 2030 opposed to a deficit of $3.387 trillion dollars without any intervention as illustrated by measure 6.

Tax Increase, Age Increase and Fraud Reductions Financial Impact for Medicare Trust Fund (measure 5)

Start Age 67 Year	Medicare Annual Cost (Billions)	Medicare Annual Cost after Post Inter-vention + Growth (Billions)	Medicare Revenue (Billions)	Average Starting Beneficiary Cost $10,400 Beneficiaries	Medicare Revenue Growth	Medicare Average Annual Beneficiary Growth	Beneficiary Life Expectancy	Average Beneficiary Growth Less Death	Health-care Inflation Average	Age Increase to 67 Savings (Billions)	Tax Increase Revenue Gain 2% (Billions)	Fraud Reduc-tions (Billions)	Medicare Trust Fund (Billions)	Current Cash Flow
2010	$ 509		$507	46,300,000	5.50%				6.90%				$ 507	
2011	$ 544	$ 395	$535	49,214,975	5.50%	3,500,000	585,025	2,914,975	6.90%	$36	$ 83	$60	$ 647	$ 140
2012	$ 582	$ 429	$564	52,089,350	5.50%	3,500,000	625,625	2,874,375	6.90%	$37	$ 85	$61	$ 783	$ 136
2013	$ 622	$ 465	$595	54,918,645	5.50%	3,500,000	670,705	2,829,295	6.90%	$38	$ 86	$62	$ 913	$ 131
2014	$ 665	$ 503	$628	57,702,860	5.50%	3,500,000	715,785	2,784,215	6.90%	$39	$ 88	$64	$ 1,038	$ 125
2015	$ 711	$ 545	$663	60,436,955	5.50%	3,500,000	765,905	2,734,095	6.90%	$39	$ 90	$65	$ 1,156	$ 118
2016	$ 760	$ 589	$699	63,117,220	5.50%	3,500,000	819,735	2,680,265	6.90%	$40	$ 92	$66	$ 1,266	$ 110
2017	$ 812	$ 637	$738	65,739,455	5.50%	3,500,000	877,765	2,622,235	6.90%	$41	$ 93	$68	$ 1,366	$ 100
2018	$ 868	$ 689	$778	68,299,250	5.50%	3,500,000	940,205	2,559,795	6.90%	$42	$ 95	$69	$ 1,455	$ 89
2019	$ 928	$ 744	$821	70,792,160	5.50%	3,500,000	1,007,090	2,492,910	6.90%	$43	$ 97	$70	$ 1,533	$ 77
2020	$ 992	$ 803	$866	73,213,950	5.50%	3,500,000	1,078,210	2,421,790	6.90%	$44	$ 99	$72	$ 1,596	$ 63
2021	$ 1,060	$ 866	$914	75,560,280	5.50%	3,500,000	1,153,670	2,346,330	6.90%	$44	$ 101	$73	$ 1,643	$ 48
2022	$ 1,134	$ 934	$964	77,823,310	5.50%	3,500,000	1,236,970	2,263,030	6.90%	$45	$ 103	$75	$ 1,673	$ 30
2023	$ 1,212	$ 1,007	$1,017	80,004,265	5.50%	3,500,000	1,319,045	2,180,955	6.90%	$46	$ 105	$76	$ 1,683	$ 10
2024	$ 1,295	$ 1,085	$1,073	82,095,060	5.50%	3,500,000	1,409,205	2,090,795	6.90%	$47	$ 107	$78	$ 1,671	$ (12)
2025	$ 1,385	$ 1,169	$1,132	84,090,935	5.50%	3,500,000	1,504,125	1,995,875	6.90%	$48	$ 110	$79	$ 1,634	$ (37)
2026	$ 1,480	$ 1,259	$1,194	85,987,270	5.50%	3,500,000	1,603,665	1,896,335	6.90%	$49	$ 112	$81	$ 1,570	$ (64)
2027	$ 1,582	$ 1,355	$1,260	87,779,340	5.50%	3,500,000	1,707,930	1,792,070	6.90%	$50	$ 114	$82	$ 1,475	$ (95)
2028	$ 1,692	$ 1,458	$1,329	89,462,490	5.50%	3,500,000	1,816,850	1,683,150	6.90%	$51	$ 116	$84	$ 1,346	$ (129)
2029	$ 1,808	$ 1,568	$1,402	91,032,205	5.50%	3,500,000	1,930,285	1,569,715	6.90%	$52	$ 119	$86	$ 1,179	$ (166)
2030	$ 1,933	$ 1,687	$1,479	92,484,950	5.50%	3,500,000	2,047,255	1,452,745	6.90%	$53	$ 121	$87	$ 972	$ (208)

Figure 5: Table Source: Gregory Scott

The Financial Impact for the Medicare Trust Fund without Intervention (measure 6)

Start Age 67 Year	Medicare Annual Cost (Billions)	Medicare Annual Cost after Post Inter-vention + Growth (Billions)	Medicare Revenue (Billions)	Average Starting Beneficiary Cost $10,400 Beneficiaries	Medicare Revenue Growth	Medicare Average Annual Beneficiary Growth	Beneficiary Death Expectancy	Average Beneficiary Growth Less Death	Health-care Inflation Average	Medicare Trust Fund (Billions)	Current Cash Flow
2010	$ 509		$507	46,300,000	5.50%				6.90%	$ 507	
2011	$ 544	$ 574	$535	49,214,975	5.50%	3,500,000	585,025	2,914,975	6.90%	$ 467	$ (40)
2012	$ 582	$ 612	$564	52,089,350	5.50%	3,500,000	625,625	2,874,375	6.90%	$ 420	$ (47)
2013	$ 622	$ 651	$595	54,918,645	5.50%	3,500,000	670,705	2,829,295	6.90%	$ 364	$ (56)
2014	$ 665	$ 694	$628	57,702,860	5.50%	3,500,000	715,785	2,784,215	6.90%	$ 299	$ (66)
2015	$ 711	$ 739	$663	60,436,955	5.50%	3,500,000	765,905	2,734,095	6.90%	$ 222	$ (76)
2016	$ 760	$ 787	$699	63,117,220	5.50%	3,500,000	819,735	2,680,265	6.90%	$ 134	$ (88)
2017	$ 812	$ 839	$738	65,739,455	5.50%	3,500,000	877,765	2,622,235	6.90%	$ 32	$ (102)
2018	$ 868	$ 895	$778	68,299,250	5.50%	3,500,000	940,205	2,559,795	6.90%	$ (84)	$ (117)
2019	$ 928	$ 954	$821	70,792,160	5.50%	3,500,000	1,007,090	2,492,910	6.90%	$ (217)	$ (133)
2020	$ 992	$ 1,017	$866	73,213,950	5.50%	3,500,000	1,078,210	2,421,790	6.90%	$ (368)	$ (151)
2021	$ 1,060	$ 1,085	$914	75,560,280	5.50%	3,500,000	1,153,670	2,346,330	6.90%	$ (540)	$ (171)
2022	$ 1,134	$ 1,157	$964	77,823,310	5.50%	3,500,000	1,236,970	2,263,030	6.90%	$ (733)	$ (193)
2023	$ 1,212	$ 1,234	$1,017	80,004,265	5.50%	3,500,000	1,319,045	2,180,955	6.90%	$ (950)	$ (218)
2024	$ 1,295	$ 1,317	$1,073	82,095,060	5.50%	3,500,000	1,409,205	2,090,795	6.90%	$ (1,195)	$ (244)
2025	$ 1,385	$ 1,406	$1,132	84,090,935	5.50%	3,500,000	1,504,125	1,995,875	6.90%	$ (1,468)	$ (274)
2026	$ 1,480	$ 1,500	$1,194	85,987,270	5.50%	3,500,000	1,603,665	1,896,335	6.90%	$ (1,774)	$ (306)
2027	$ 1,582	$ 1,601	$1,260	87,779,340	5.50%	3,500,000	1,707,930	1,792,070	6.90%	$ (2,116)	$ (341)
2028	$ 1,692	$ 1,709	$1,329	89,462,490	5.50%	3,500,000	1,816,850	1,683,150	6.90%	$ (2,496)	$ (380)
2029	$ 1,808	$ 1,825	$1,402	91,032,205	5.50%	3,500,000	1,930,285	1,569,715	6.90%	$ (2,918)	$ (423)
2030	$ 1,933	$ 1,948	$1,479	92,484,950	5.50%	3,500,000	2,047,255	1,452,745	6.90%	$ (3,387)	$ (469)

Figure 6: Table Source: Gregory Scott

The interventions of tax and age increases and fraud reduction create a solvency probability of 100% for the Medicare Trust Fund. Using only one intervention creates a lower probability for solvency for the Medicare Trust Fund as illustrated by measure 7. The lowest probability for Medicare Trust Fund solvency is no intervention.

Medicare Trust Fund Solvency Probability (measure 7)

MEDICAL TRUST FUND (Billions) After Intervention						
Year	Age, Tax, & Fraud	Age	Tax	Fraud	None	
2011	P	P	P	P	P	
2012	P	P	P	P	P	
2013	P	P	P	P	P	
2014	P	P	P	P	P	
2015	P	P	P	P	P	
2016	P	P	P	P	P	
2017	P	P	P	P	P	
2018	P	P	P	P	N	
2019	P	P	P	P	N	
2020	P	P	P	P	N	
2021	P	N	P	P	N	
2022	P	N	P	P	N	
2023	P	N	P	N	N	
2024	P	N	P	N	N	
2025	P	N	N	N	N	
2026	P	N	N	N	N	
2027	P	N	N	N	N	
2028	P	N	N	N	N	
2029	P	N	N	N	N	
2030	P	N	N	N	N	
TOTAL POSITIVE OUTCOMES	20	10	14	12	7	63.0
TOTAL POSSIBLE OUTCOMES						80.0
PROBABILITY						0.8

N = Negative Balance
P= Positive Balance

Figure 7: Table Source: Gregory Scott

Sample Description and Selection
Sampling for this study will be collected primarily from Medicare.gov., the United States Census Bureau and the Congressional Budget Office. Inflation rates for healthcare and non- healthcare will be used to ascertain cost estimates for future Medicare payments. Models will simulate Medicare insolvency to Medicare solvency using years 2010 to 2030. Modeling and simulation via programs as Monte Carlo will be incorporated to visualize Medicare solvency results. The models will include and preclude variables to demonstrate the impact on solvency for Medicare. Data will include all Americans 65 years of age and older.

Research Design
The design will be an experimental pretest/post test design. Medicare insolvency will have a pretest that evaluates no intervention for each year from 2010 to 2030 and a post test with intervention from 2010 to 2030.

The variables will be age, fraud cost, and tax increases to establish a relationship regarding solvency. This will be the foundation for examining the potential solvency aspects the Medicare Trust Fund in the future.

Data Collection Instrument
Data for this current research will be collected from The Congressional Budget Office and the Centers for Medicare and Medicaid Services and primary sources. The data will be collected via literature review to validate age, associated cost to Medicare, inflation, projected Medicare cost and life expectancy.

CONCLUSION

Measures 1, 3, 4 over a twenty year period represent a positive impact for the Medicare Trust Fund as a whole. Increasing the Medicare pay role tax by 2 percent for employees and employers creates $1.66 trillion in additional revenue. Increasing the age to 67 years increases Medicare revenue by $728 billion. Last, decreasing fraud for Medicare decreases cost by $1.2 trillion over 20 years. The main question for this research is can Medicare insolvency be prevented? Measures 1, 3 and 4 prevent Medicare from insolvency during the baby boomer generation with a $972 billion dollar surplus remaining in the Medicare Trust Fund. The models demonstrate that Medicare insolvency prevention can take place if the measures mentioned are enacted sooner than later. The Medicare Trust Fund will become insolvent starting with year 2018 with an $84 billion deficit and by year 2030 the fund will have over a $3 trillion deficit if intervention does not take place to return the Medicare Trust Fund to solvency.

REFERENCES

[1]. American Academy of Family Physicians. (2008). Medicare Unsustainable Growth Rate. Retrieved January 20, 2010, fro, http://www.afp.org/fpm/2008/0500/p9.html

[2] Medicare.gov. (2010). Medicare Eligibility Tool. Retrieved January 24, 2010, from http//www.medicare.gov/Medicare Eligibility/Home.?asp?dest

[3] Social Security Administration. (2010). Fact Sheet, 35 - 36. Retrieved January 20, from http://www.ssa.gov/pubs/100003.html

[4] Argonee. (2009). Modeling, Simulation, and Analysis:Compare. Retrieved January 5, 2009 from http://www.dis.anl.gov/pubs/61122.pdf

[5] United States Census Bureau. (2010). Population. Retrieved January 22, 2010, www.census.gov/population/www/ pop- profile/natproj.html.

[6] Congressional Budget Office. (2010). Frequently Asked Questions/Cost Estimates. Retrieved January 24, 2010, from http;//www.cbo.gov/aboutcbo/faqs.shtml

[7] Central Intelligence Agency. (2010). The World Fact Book, 1. Retrieved January 22, 2010, from http://www.cia.gov/library/publications/ the-world-factbook/rankorder/2102 rank.html

[8] Creal, Drew (2009). A survey of sequential Monte Carlo methods for economics and Finance. University of Chicago, Both School of Business, 1 – 12.

Heart sound cancellation using an adaptive wavelet thresholding method

Hoang-Anh T. Nguyen[1], Jiang Li[2], Ender Oguslu[2] and Frederic D. Mckenzie[1],

[1]Modeling, Simulation and Visualization Engineering Department, Old Dominion University

[2]Electrical and Computer Engineering Department, Old Dominion University

***Abstract*— In this study, we implemented a wavelet thresholding method using a time-scale adaptive soft-like thresholding function based on the Stein's unbiased risk estimate (SURE) to cancel heart sounds from contaminated lung sounds. In this method, the signal is first decomposed by wavelet transform. Then, coefficients at a number of frequency sub-bands contaminated by noised heart sounds, are corrected by thresholding. The threshold value is determined via an adaptive procedure based on the properties of wavelet sub-band coefficients. The preliminary results show that the method is efficient and promising to this type of problem.**

***Index Terms* – Lung sounds, heart sounds, wavelet thresholding, SURE risk.**

I. INTRODUCTION

Lung sounds, which can be recorded over chest wall using a stethoscope is known to be useful for respiratory disease diagnosis. Produced by vortical and turbulent flow within lung airways during inspiration and expiration of air [1], the frequency spectrum of lung sounds is found relatively broad, normally in the range of 20-1600 Hz [1] and even up to 2,000 Hz in some cases [2]. However, lung sound signals are often contaminated by various noise sources, such as ambient noise, muscle activities, skin or hair friction, connection of sensor to chest wall and heart sounds [9 - 10]. Mixed together with lung sound recordings, those types of noises make auscultation become more arduous.

Heart sounds, produced by the flow of blood though the heart and by the movement of structures involved in the control of this flow [1], are unavoidable and contribute largely to lung sound contamination. Heart sound noises overlap lung sounds in the low frequency range, less than 150 Hz [2, 4] or from 20 to 300 Hz [5]. Heart sound cancellation from recorded lung sounds without suppressing pulmonary information, is equivalent to pulmonary disease diagnostic improvement using auscultation of lung sounds has been attracting much interest [1, 3-8].

A number of techniques have been proposed to reduce effect of heart sounds from lung sounds. Simply using high-pass filter is not a suitable solution to this sort of issue because by doing this, overlapping spectrum of breath sounds is removed and thus, useful respiratory information is lost as well. An adaptive filtering technique was proposed for the same purpose of heart sound reduction [3]. In order to denoise contaminated signals, the method is implemented in two phases 1) an adaptive filter is obtained after a training procedure and 2) contaminated signals are denoised using such trained filter. The authors mentioned that after correcting, the heart sounds are reduced by 50-80 percent [3]. However, other researchers are not convinced this method performance because the method requires a noise-reference channel, which is not always available [11-12].
An alternative method which is capable of removing heart sounds using a reduced-order Kalman filter (ROKF) was already proposed [4]. This method tries to adapt an autoregressive model using LS-free HS segments based on three assumptions: (1) heart and respiratory sounds are mutually uncorrelated; (2) they have additive interaction and (3) the prior and subsequent HS are linearly related to the HS corrupted by the respiratory signal. All the three assumptions are not convincible in terms of practical implementations [5].

Independent component analysis (ICA), specifically fast ICA, is also utilized to separate HS and LS [13]. In this method, HS and LS are collected from both chest sides using two separate recording channels. Those signals are then decomposed into independent components, which corresponds to noise-free HS and LS. A combined technique using multiscale products and linear prediction are developed for the same purpose of heart sound removal. This method decomposes the original signal first into a set of wavelet coefficients at multiscale frequencies. Once the HS segments are identified, a set of linear prediction filters are utilized to estimate the noise-free wavelet coefficients [14].

Wavelet thresholding, a multiresolution wavelet-based method for signal analysis, was also proposed as another solution to LS/HS problem [6, 7, 8]. Specifically, after decomposed by wavelet transform, the coefficients of noisy signals at different scales are then corrected with some type of soft, hard or adaptive thresholding function. Previous works [6, 7, 8] show that wavelet thresholding is capable of reducing heart sound significantly, meanwhile the useful respiratory information is well-preserved.

The objective of this study is to investigate the application and feasibility of a wavelet thresholding method using a time-scale adaptive soft-like thresholding function based on the Stein's unbiased risk estimate (SURE) [18] is utilized for heart noise cancellation. The method is implemented as follows. Lung sound signals are first decomposed by wavelet transform. The coefficients at low frequency sub-bands where heart sounds appear are corrected using a thresholding procedure where a data-oriented threshold is adaptively obtained. Those thresholded coefficients are then utilized for wavelet reconstruction to reconstruct clean lung sounds. Such study has not been done previously.

The rest of this paper is organized as follows: Section II presents the methodology. Section III shows and provides discussions for experimental results. Section IV concludes the paper and discusses future work.

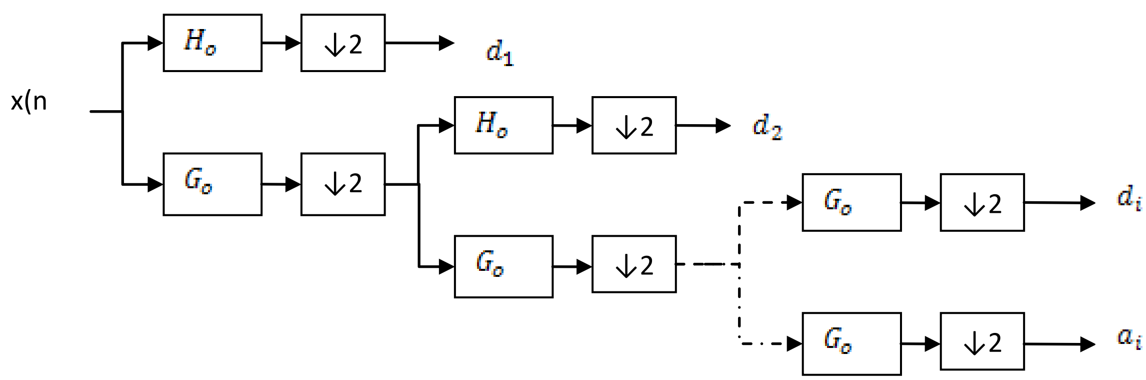
II. METHODOLOGY

Wavelet transform [15,16] is a signal processing approach, which is capable of analyzing signals in both time and frequency domains. The transform is realized by multiscale-corelating signals with a basis function, known as wavelet. The wavelet basis can be constructed from a single function $\psi(t)$, named mother wavelet or analyzing wavelet, by means of translation and dilation.

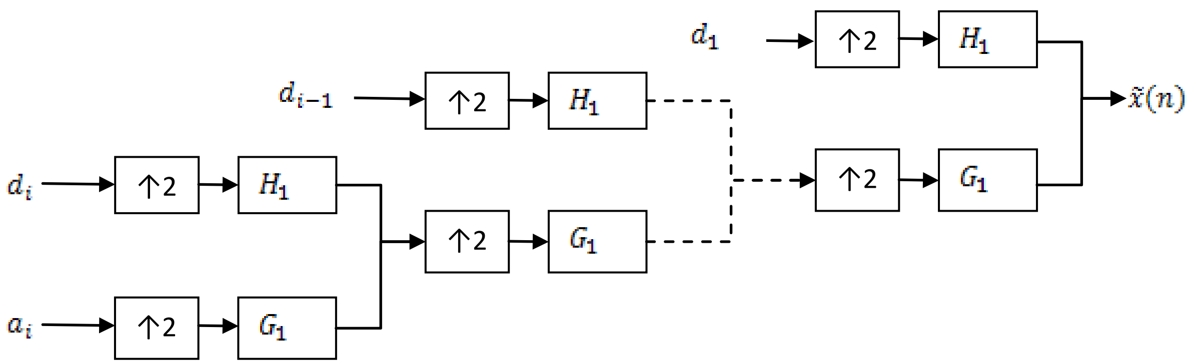
Continuous wavelet transform (CWT) of a signal x(t), defined as the correlation between the wavelet and the signal itself, can be implemented by the following formula:

$$W(a,\tau) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\psi^{*}_{a,\tau}(t)dt \qquad (1)$$

where $\psi^{*}(t)$ denotes the complex conjugate of $\psi(t)$. The above equation idicates that the wavelet is passed through the analyzed signal and yields the set of coefficients representing image of analyzed signal at different scales over time and frequency. The highest scale corresponds to the highest frequencies represented in the signal (less or equal to half of the sampling rate), and the bandwidth of this scale ranges from a half to a quarter of the sampling rate. While the bandwith is reduced by two, the number of coefficients at lower resolution decreases aproximately by a factor of two compared to that of the higher resolution next to it.



a)



b)

Figure 1. Wavelet transform (a) wavelet decomposition and (b) wavelet reconstruction

Wavelet transform and wavelet reconstruction can be illustrated at the figures 1 (a-b). Wavelet decomposition is implemented in the following process: the original signal is first passed through high pass (H_o) and low pass (G_o) filters, that are designed based on the properties of wavelet basis function, and then downsampled by two. After that, detail, or high frequency coefficient series and approximation, or low frequency coefficient series are obtained. Approximation terms are then continously used as the inputs for next levels of wavelet transform. Finally, there would be one approximation (a_i) and details ($d_k, k = 1, .., i$) representing the signal information at different scales. On the other hand, wavelet reconstruction is the reverse process of wavelet decomposition. Approximation and detail at each level are upsampled by two and passed through low pass (G_1) and high pass (H_1) synthetic filters, which are sastified perfect construction condition [16] and then summed up. This process is continued *i* times, equal to the number of levels of decomposition. The result of this process could be the original or corrected signal. In the later case, detail and approximation coefficients at some levels are affected in desirable ways.

Wavelet thresholding is a wavelet-based technique normally rendered in the issues involved signal denoising. For this type of method, the contaminated signals are decomposed by wavelet transform with a proper mother wavelet. Then, large-valued coefficients are generated in the areas corresponding to heart sound noise at low-frequency sub-bands and are considered as an estimate of heart sound signals. Thus, shrinking the amplitude range of these coefficients by nonlinear thresholding functions would be equivalent to noise cancellation. For the purpose of heart sound removal, a wavelet thresholding method was implemented as follows,

- Apply the wavelet transform to the contaminated lung sound signal
- Utilize a thresholding function to automatically correct high-valued coefficients in low-frequency sub-bands
- Reconstruct the lung sound signal based on corrected coefficients dilation.

The threshold function is selected such that it can suppress the noise and keep the useful information maximally making the threshold function selection become a non-trivial step. Being found useful in a previous work related to EEG artifact removal [17], in this study, a time-scale adaptive algorithm based on the SURE risk estimate along with a soft-like thresholding function [18] is investigated for a similar denoising issue, heart sound cancellation.

Threshold function

The following threshold fucntion is utilized to correct the wavelet coefficients of lung sound signals in low frequency sub-bands,

$$w_k^c(w,t) = \begin{cases} w + t - \frac{t}{2k+1}, w < t \\ \frac{1}{(2k+1)t^{2k}} \; w^{2k+1} \quad , |w| \le t \\ w + t + \frac{t}{2k+1}, w > t \end{cases} \qquad (2)$$

where w_k^c represents the corrected version of a wavelet coefficients w using the thresholding value *t*. The optimal value of *t* can be adjusted based on the SURE risk using the following adaptive steps

$$t(i+1) = t(i) - \nabla t(i) \qquad (3)$$

where the adjustment of threshold at step *i* is defined by

$$\nabla t(i) = \alpha(i) \, . \frac{\partial R_s(t)}{\partial t} \qquad (4)$$

where

$$\frac{\partial R_s(t)}{\partial t} = 2 \sum_{i=0}^{N-1} g_i \, . \frac{\partial g_i}{\partial t} + 2 \sum_{i=0}^{N-1} \frac{\partial^2 g_i}{\partial w_i \partial t} \qquad (5)$$

where

$$g_i = w_k^c(w_i,t) - w_i \qquad (6)$$

The SURE risk threshold function is applied to correct the coefficients at approximation and several lowest frequency sub-bands. These corrected coefficients are then used for signal reconstruction.

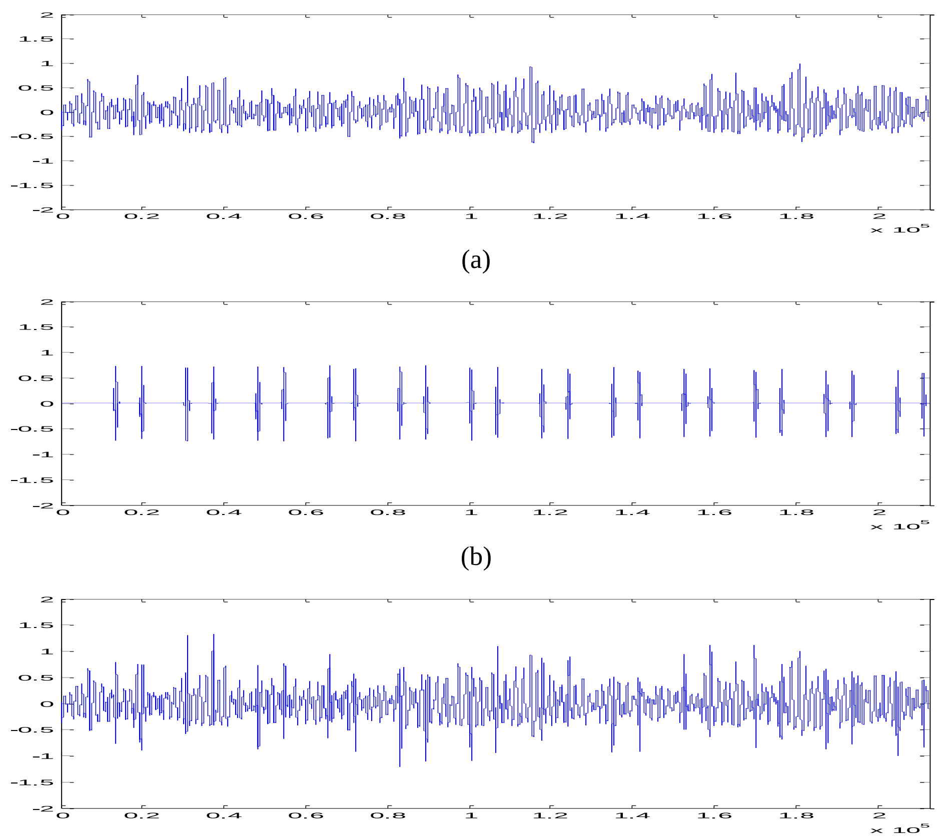
III. Results and discussions

We tested the method in a synthesized contaminated lung sound segment of 2.1269e+005 samples (aproximatedly 9.6459 seconds) sampled by sampling rate of 22500 Hz. That data segment was created by mixing two pure heart and lung sound data segments together. This synthesized lung signal share similar time-frequency properties with real lung sounds. Thus, testing the method in synthesized data offers a significant advantage, knowing exactly locations and amounts of heart sound noise facilitates both quantitative and qualitative result validation.

Figure 2 shows the graphical illustration of pure lung sounds, heart sounds and synthesized contaminated lung sounds. Lung sounds seem to be ‘active’ almost all the time meanwhile, heart sounds only occur at some certain periods. Besides, the occurrence of heart sounds, which depends on the heart beating mechanism and as shown in Figure 2, is not completely periodic.

In order to realize wavelet thresholding technique, a very first task is to select fitted wavelets according to properties of decomposed signals. This is a trial and error procedure, in which the selected wavelet is determined whenever heart sounds are suppresed largely meanwhile noncontaminated lung sounds are well-recovered. During the course of experimental implementation, a number of wavelets and various levels of wavelet decomposition were tested. Among them, the best-fitted ones are Symlet and Coiflet. The number of wavelet stages, which gets on well with the chosen wavelets, is 10. In our experiment we apply the thresholding function to coefficients from approximation and details at level 9,8,7 and 6. Figure 3 shows results of wavelet thresholding correction using coiflet and symlet basis functions with 10 stages of wavelet decomposition. The corrected lung sounds are very well-recovered compared to the original signals. Specifically, by visual inspection, the peaks, which represent for heart sounds, are reduced significantly. Meanwhile, background information, which represents for lung sounds, are well-preserved. Those results are also testified by using auditory inspection method, in which the lung sounds after correction are very much heart-sound-free.

Power spectrum density (PSD), estimated using Welch's method, gives better understanding on those results in frequency domain. As shown in Figure 4a), heart sounds spectra dominant in two ranges of frequency (corresponding to first and second heart beatings) overlap active spectra of lung sounds. Figure 4b) shows lung sounds before and after synthesis. Besides two significant peaks, there are slight changes across synthesized lung sounds' spectra compared to original pure lung sounds. In this case, using some simple regression methods, either in time or frequency domain, is equivalent to useful information lost. Wavelet methods are very efficient for this type of issue and actually, experimental results in this research confirms their capability. However, the method could not remove heart sounds completely. Slight 'clicks' corresponding to heart sounds are still there. The spectra of those 'clicks' correspond to peaks as can be seen in Figure 4c).

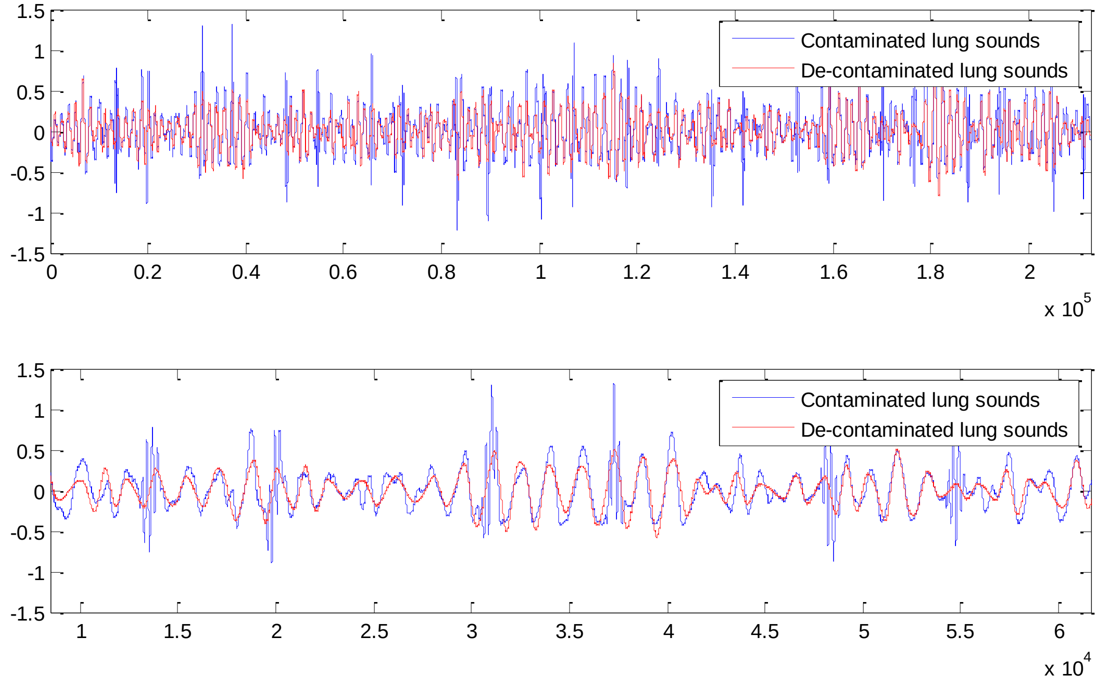


(a)

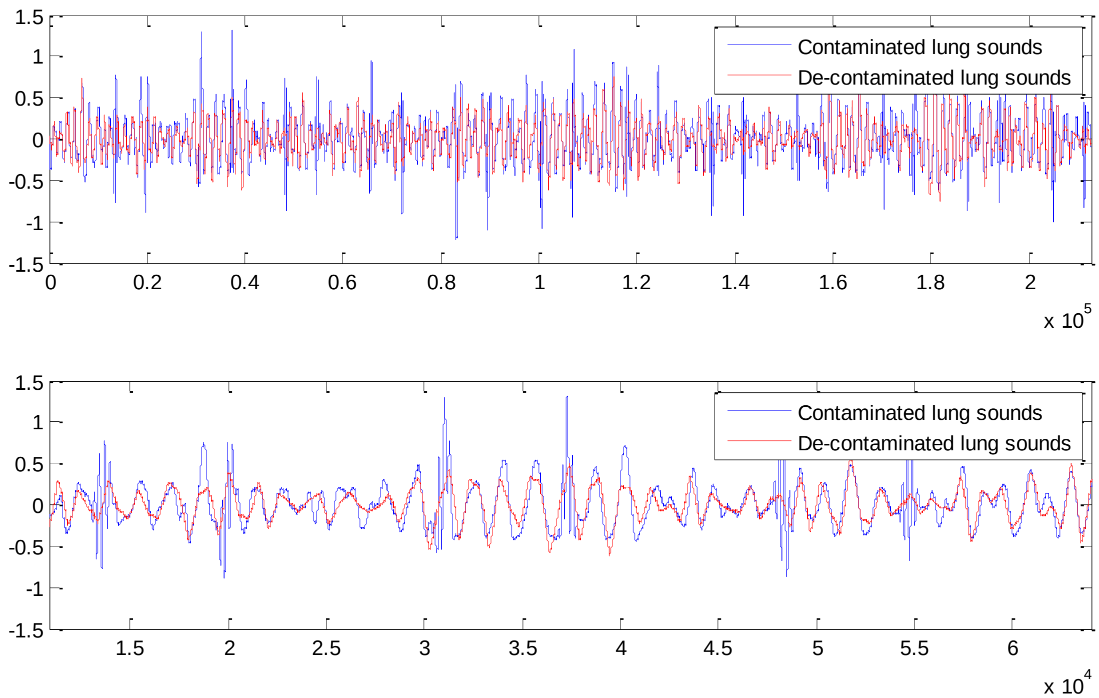
(b)

(c)

Figure 2. Lung sounds, heart sounds and lung sounds contaminated by heart sounds

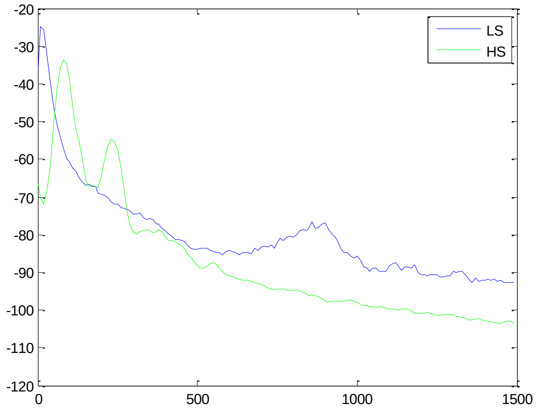


wname = ‘coif4’; lev = 10;

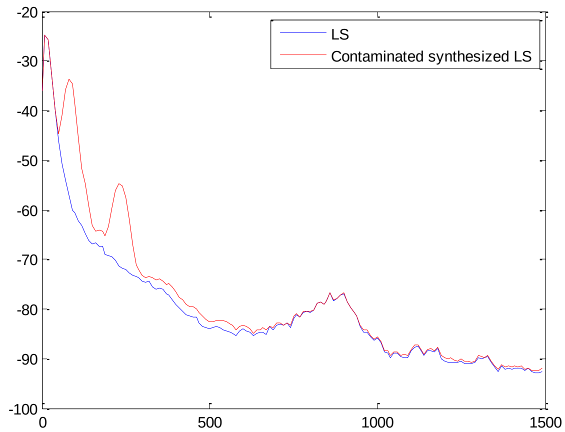


wname = ‘sym4’; lev = 10;

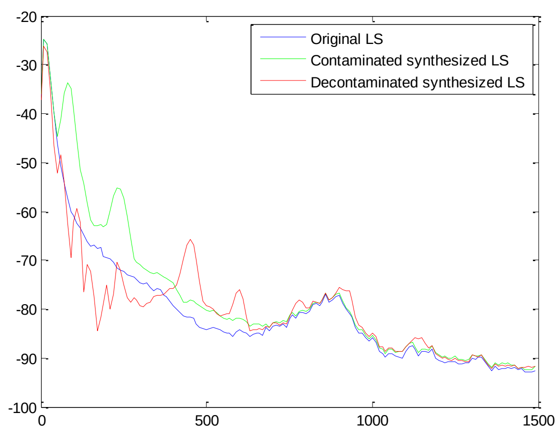
Figure 3. Lung and heart sounds before and after wavelet thresholding correction



(a)



(b)



(c)

Figure 4. Power spectrum density of (a) original LS and HS, (b) original LS and LS and contaminated synthesized LS and (c) original LS, contaminated synthesized LS and decontaminated synthesized LS.

IV. CONCLUSIONS AND FUTURE SCOPE

We implemented a wavelet thresholding method using an adaptive threshold function for lung and heart sound separation. The preliminary results show that the method is capable of removing the unwanted heart sounds. Meanwhile, the useful information is well-preserved.

Our future work includes testing the method on various lung sound signals, further processing the results obtained with wavelet thresholding and implementing other heart sound cancellation methods for comparison.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Jiang Li and Dr. Frederic D. McKenzie for their encouragements and supports during the course of this project implementation.

REFERENCES

[1] *J. Gnitecki and Z. M. K. Moussavi, "Separating Heart Sounds from Lung Sounds - Accurate Diagnosis of Respiratory Disease Depends on Understanding Noises," Engineering in Medicine and Biology Magazine, IEEE , vol.26, no.1, pp.20-29, Jan.-Feb. 2007.*

[2] *V. Gross, A. Dittmar, T. Penzel, F. Schuttler, and P. von Wichert, "The relationship between normal lung sounds, age, and gender," Am. J. Respir. Crit. Care Med., Vol. 162(3), pp. 905-909, Sep. 2000.*

[3] *V. K .Iyer, P. A. Ramamoorthy, H. Fan, and Y. Ploysongsang, "Reduction of heart sounds from lung sounds by adaptive filtering," IEEE Trans. on. Bio. En., Vol. BME-33, No. 12, Dec. 1986.*

[4] *S. Charleston and M. R. Azimi-Sadjadi, "Reduced order Kalman filtering for the enhancement of respiratory sounds," IEEE Trans. on Bio. Engineering, Vol. 43, No. 4, April 1996.*

[5] *M. T. Pourazad, Z. Moussavi and G. Thomas, "Heart sound cancellation from lung sound recordings using time-frequency filtering," Med. Biol. Eng. Comput., Vol. 44, pp. 216-225, 2006.*

[6] *L. F. Hadjileontiadis, S. M. Panas, "A wavelet-based reduction of heart sound noise from lung sounds," International Journal of Medical Informatics, Vol. 52, pp. 183-190, 1998.*

[7] *I. Hossain and Z. Moussavi, "An overview of heart-noise reduction of lung sound using wavelet transform based filter," in Proc. of the 25th Ann. In. Con. Of the IEEE EMBS, Cancun, Mexica, Sep. 17-21, 2003.*

[8] L. J. Hadjileontiadis, and S. M. Panas, "Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter," IEEE Trans. on Bio. Eng., Vol. 44, No. 12, Dec. 1997.

[9] R. Loudon and R. L. H. Murphy, Jr., "Lung sounds," Amer. Rev. Respir. Dis., vol. 130, pp.663 – 673, 1984.

[10] N. Gavriely, M. Nissan, A.-H. E. Rubin, D. W. Cugell, "Spectral characteristics of chest wall breath sounds in normal subjects,"

[11] S. Charleston, M. R. Azimi-Sadjadi, "Interference cancellation in respiratory sounds via a multiresolution joint time-delay and signal-estimation scheme," IEEE Trans. Biomed. Eng., vol. 43, no. 4, pp. 421-424, Apr. 1996.

[12] L. Yip and Y. T. Zhang, "Reduction of heart sounds from lung sound recordings by automated gain control and adaptive filtering techniques," in Proc. 23rd Ann. Int. Conf. IEEE Eng. Medicine Biology Soc., EMBC'01, pp. 2154-2156, 2001.

[13] J. C. Chien, M.-C. Huang, Y.-D. Lin, and F.-C. Chang, "A study of heart sound and lung sound separation by independent component analysis technique," in Proc. EMBS Intl. Conf., pp. 5708 – 5711, Sep.2006.

[14] D. Flores-Tapia, Z. Moussavi, and G. Thomas,"Heart sound cancellation based on multiscale products and linear prediction," IEEE Trans. on Biomedical Eng., vol. 54, no. 2, pp. 234 – 243, Feb. 2007.

[15] Daubechies Ingrid, Ten lectures on wavelets, Philadelphia: Soc. Ind. Applied Math., 1992.

[16] O. Rioul and M. Vetterli, "Wavelet and signal processing," IEEE Signal Processing Magazine, Vol. 8, No. 4, pp. 14 – 38, 2002.

[17] V. Krishnaveni, S. Jayaraman, L. Anitha and K. Ramadoss, "Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients," Journal of Neural Engineering, Vol. 3, pp. 338-346, 2006.

[18] X. P. Zhang and M. D. Desai "Adaptive denoising based on SURE risk," IEEE Signal Processing Lett., Vol. 10, No. 5, pp. 265 – 267, Oct. 1998.

An Improved Virtual Operating Room Training Scenario Model

Rebecca Kennedy & Kellie Kennedy
Old Dominion University

Abstract—**The Virtual Operating Room (VOR) is a fully immersive training environment suitable for surgical trainees to practice procedural and team training skills in a realistic but safe context. Users are able to interact with virtual avatars representing other surgical team members via voice recognition and text-to-speech technology. The VOR is useful for implementing specific training scenarios according to training needs. However, a limitation of the system is that the state machine lacks the flexibility necessary for a smooth and adaptive training experience. The two main problems with the current model are that queries are organized linearly so that questions must be asked in a specified order to be recognized by the system, and a scenario cannot successfully be completed until the user has completed all of the queries. Here, a new model for the state machine is proposed to address the limitations of the current model. In the new model, the state machine is organized using an array that allows queries to be asked in any order. Scenario success is ultimately determined by the user asking the terminal question, which can occur at any point in the simulation with or without having exhausted all the other query options. The new model will make scenarios more flexible and realistic, and the improved organization will make it easier for trainers (non-programmers) to modify scenarios according to specific needs.**

I. Introduction

Virtual environments used for training in the medical domain allow trainees to practice skills repeatedly without risking patient safety. High-fidelity, interactive virtual environments provide a lifelike training experience that creates context for procedural skills. This context is what separates virtual environments from individual simulators targeting specific skills. The Virtual Operating Room (VOR), created by Old Dominion University and Eastern Virginia Medical School, was intended to augment procedural simulator training by providing an immersive virtual environment representative of actual operating room conditions [1, 2]. Specific training scenarios can be selected to test surgical team members' skills under different conditions and with different possible complications.

A potential advantage of the VOR system is the ability to modify scenarios for specific training needs. However, a limitation of the current VOR system is that the rigid structure of the underlying scenario models does not allow for easy modification.

Therefore, the goal of the present research is to make improvements to the VOR to increase the overall flexibility of the simulation through scenario restructuring. The improved scenario model will increase the realism of the simulation, and therefore be more applicable to users of varying experience levels. Additionally, a single basic framework that can be applied across multiple scenarios will allow trainers to easily generate custom scenarios, rendering the system more useful to a wider variety of clients.

II. VOR Training Scenario: Previous Work and Development

A. *Benefits for Team Training*

The VOR is a training system capable of supporting surgical team training, which is particularly advantageous in a medical setting where communication is critical for patient safety. Team training is traditionally performed with multiple individuals; however, the VOR allows for interactive virtual agents to take the place of human team members. Therefore, a single trainee can benefit from team training without requiring actual team members. Specifically, junior surgeons can use the VOR to learn how to appropriately interact with the attending surgeon and other members of the surgical team.

B. *Surgical Procedure*

The VOR scenario within the current research requires a trainee to identify and properly treat the symptoms of a serious medical emergency that occurs during a common procedure. The procedure selected for development in the VOR is a laparoscopic cholecystectomy (gall bladder removal). A basic script, based on fundamental procedural steps, was written for the simulation. In a specific scenario, which will be the scenario discussed in this paper, a complication of the patient experiencing a serious medical emergency (malignant hyperthermia) was also identified and a script was written for this complication.

C. *Agent Development*

Virtual agents in the simulation are each independent, autonomous state machines presented as an avatar. The avatars in the VOR can interact with the user through voice recognition, environmental sensors, and equipment simulators. Each agent is capable of outputting animations and verbal responses that are triggered by either pre-specified inputs, via the user's speech, or a predefined amount of time elapsing without recognizing an input. The outputs generated by a virtual agent are defined by a knowledge base derived from experienced surgical team members. Further, the nature of the outputs may vary based on personality types assigned to avatars, such as an agent who is markedly impatient or highly supportive.

The knowledge base for each virtual agent was created using cognitive and hierarchical task analyses of practicing surgeons, nurses, and anesthetists. The procedure is divided into three phases: anesthesia induction, surgery, and emergence from anesthesia. Within each phase are sequences of tasks for each surgical team member. The initial task analyses were reviewed and revised by subject matter experts, and then used to create a framework for the cholecystectomy procedure under ideal conditions as well as for primary complications.

The personality types for the agents were created using the Big Five model [3], which describes personality according to a person's inherent tendencies defined by five characteristics: openness, conscientiousness, extraversion, agreeableness, and emotional stability. The personalities are manifested though agent speech, physical movement, and posture. At present, there are four personalities that have been developed for VOR agents: helpful, arrogant, impatient, and reserved. For example, as a prompt, a helpful agent may say, "I think you may want to ask about the patient's allergies," whereas an arrogant agent may say, "I assume you've asked about the patient's allergies. Okay, let me help you – ask about the patient's allergies." In addition to varying the actual words spoken, speech can be altered in terms of pitch and speed to convey different moods of the agent.

D. *User/Agent Interactions*

Interactions between the user and virtual agents are guided by an underlying script, which consists of a sequence of steps organized according to how the trainee is expected to behave. The script provides the sequence of steps, interactions, and branches that could potentially occur in the scenario. Pre-scripted agent responses are stored for each training scenario and activated using a computer generated text-to-speech engine when the appropriate input is recognized. Speech output is generated according to the specific training scenario, the role of the agent, and agent personality type.

The model underlying the scenario scripts is a finite state machine model. Each entity (e.g., avatar, patient) has its own state machine [4]. The simulation is agent based, and as described, interactions between the trainee and the simulated agents follow the scenario script and expected interactions. Although an easy to use graphical interface has been developed to allow non-programmers to define training scenarios, the interface is limited to the selection of pre-existing simulation components. The rigidity of each scenario does not allow additional scenarios to be easily applied.

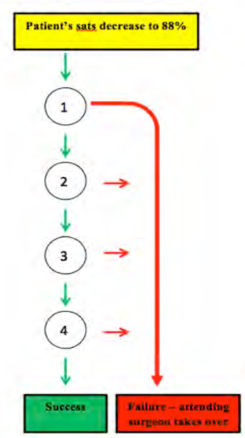
In summary, the VOR is a finite state machine through which different training scenarios are created and implemented. A unique capability of the simulator is to accommodate team training using agents in an immersive virtual environment, which we can demonstrate using the existing scenarios. However, a goal for the system is to sell it commercially and allow consumers to create their own scenarios according to specific training needs. With this consideration in mind, the current model design is difficult to modify and provides users with very little flexibility. The goal for this project is to create and implement a new model to address these limitations.

III. Current Model Limitations

For the purposes of this paper, we will focus on one training scenario for the laparoscopic cholecystectomy in which the patient suffers from malignant hyperthermia during the procedure. The need for a new simulation stems from the current VOR scenario requiring the user to interact with the simulation in an ordered series of steps. This strict order is a problem for experienced users who are capable of recognizing, diagnosing, and treating the emergency without following the predefined series of steps.

Figure 1 shows the VOR training scenario using the linear decision tree structure in the current model. As the green arrows indicate, the steps are expected to be performed in the order set forth by the model despite the fact that in normal surgical procedures these tasks may be performed in any order. The red arrows indicate the trainee has failed to perform the procedure correctly based on the linear sequence of steps. Because only queries appropriate to the next step are recognized as correct, failure can be due to an incorrect query, a correct query unrecognized by the speech recognition software, or a correct query given out of sequence.

Figure 1. Current model showing the steps in a linear order.

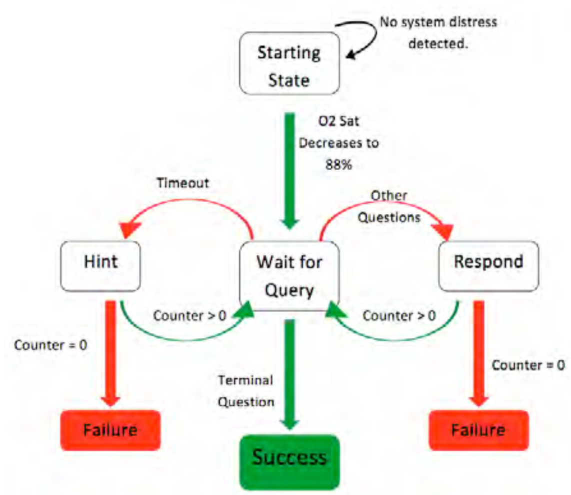


In the malignant hyperthermia scenario, it is expected that users ask the surgical team specific questions in order before learning that the patient has an elevated temperature. The elevated temperature is the final decision point designed as a cue for the halting of the procedure and administration of the appropriate reversal agents. However, in initial testing, experienced subject matter experts in this training scenario made the decision to stop the case prior to exhausting all questions. Although users gave the correct solution, the system could not acknowledge the successful completion of the scenario because the users did not exhaust all steps in the sequence.

IV. RESTRUCTURED MODEL

Figure 2 shows the new scenario model restructured to allow a more flexible and general framework that can be modified to fit the training needs of the users. This restructuring allows questions to be asked in any order or remain unasked and the user can still arrive at the correct solution. The training scenario will be considered successful as long as the correct outcome is achieved, which can now occur at any time during the simulation.

FIGURE 2. RESTRUCTURED MODEL WITH QUERY FLEXIBILITY.



In the new model, the system state changes when the patient's arterial blood gas, referred to as "Sats" (percentage of oxygen 'saturation' present in the blood) decreases to 88%. Once this prompt occurs, the system waits for a query from the user. This query could be any of the recognized questions or the terminal question, which ends the scenario. When a query is recognized (e.g., "What is the abdominal pressure?"), the avatars provide the appropriate answer (e.g., "15mmHG.") and the system again waits for a query. If the user speaks but asks an unrecognized question, the avatars will respond by saying, "I did not understand you."

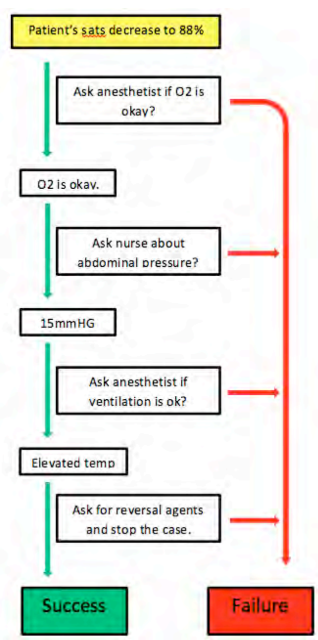
If, however, the user does not provide any queries within a certain amount of time (i.e., a timeout occurs), an avatar will provide a prompt or hint to encourage the user to query the system. The system determines which queries the user has already asked, if any, and provides a prompt for the next query in the sequence. After the hint is given, the system continues to wait for a query. Hints will only be given a prespecified number of times as represented by the counter being greater than zero. Each time a hint is given the counter will decrease by one. At the time the counter reaches zero the progress of the trainee is considered to be unacceptable and the trainee will fail the simulation.

The simulation comes to an end when the user asks the terminal question, which signifies ultimate success in the training scenario. In the malignant hyperthermia scenario, the terminal point occurs when the user directs the avatars to stop the procedure and administer the appropriate reversal agent, Dantrolene. When this direction is recognized by the system, the training scenario is considered correctly completed. In the old model, this terminal question could only be recognized if the user had exhausted the series of pre-specified questions. In the new model, the terminal question can be recognized at any time and is separately considered from other types of questions. Therefore, whenever the terminal question is asked, the simulation is successfully ended. Failure occurs in the simulation if the user exhausts the allotted number of prompts without arriving at the terminal question. In this scenario, failure occurs when the attending surgeon avatar deems the user inadequate to continue and takes over the procedure.

A. *Array Table*

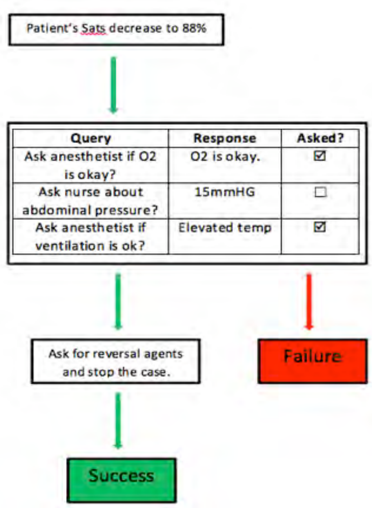
The original model for the state machine, shown in Figure 3, required that questions be asked in a specific order for the system to recognize inputs. At each state, only one specific question could be asked to move forward within the scenario. During pilot testing with subject matter experts, this linear design was found to be unnatural and rigid in structure, unlike in a real-world operating room. Therefore, programming efforts will be made to restructure the state machine using an array table rather than the linear decision tree format.

FIGURE 3. EXISTING STATE MODEL WITH LINEAR QUESTION ORDER.



The array table design, by contrast, will allow state selection to exist as a list of possible queries with the associated system responses or prompts. The information is organized as an array table instead of as individual states for each query. Figure 4 shows an example of the proposed array table. Whenever a reference is made to a particular query in the array table, the corresponding system output is provided for that state and the value in the array is checked. This check allows the system to indicate which questions have been asked and which questions can still be prompted for the user, ensuring no overlap. The prompt system activates if the participant does not respond within a set period of time (called a timeout). If a timeout occurs, the system will present the prompt slated for the next query that is unchecked in the table. This most important feature of this new organization structure is that it allows the questions to be asked in any order and it allows possible questions to remain unasked without jeopardizing the flow of the training scenario. This change allows experienced users, who recognize the medical emergency early in the scenario, to react as they would naturally and successfully complete the scenario.

FIGURE 4. PROPOSED STATE MODEL WITH ARRAY TABLE.



B. *Proposed Testing*

The new system will be tested by bringing in subject matter experts; that is, actual surgical team members. The surgeons would be residents in varying post-graduate years. The degree to which the surgeons with different experience levels can fluidly interact with the system will determine the new system success. This validation process will be used to determine whether the system accurately represents the scenario as it occurs in the genuine operating room, or whether further modifications will need to be made to improve simulation realism.

C. *Contribution*

Ultimately, with the implementation of the improved model proposed here, the training experience will be enhanced. Allowing more flexibility during the training scenarios will allow the same scenario to be appropriate for a wider level of skill sets; that is, the system will be able to accommodate earlier decisions made by more experienced surgical team members.

Due to this restructuring, it will also be easier for consumers to develop and expand scenarios using an interface requiring no programming knowledge. The new model consists of a general framework that can be easily modified to adhere to the training needs of institutions. An easy to use interface will be designed so that trainers can input their own agent dialogue and speech recognition keywords to customize scenarios and fit the model to many different contexts.

REFERENCES

[1] M. W. Scerbo, L. A. Belfore, H. M. Garcia, L. J. Weireter, M. Jackson, A. Nalu, and E. Baydogan, "The virtual operating room," Proceedings of the Interservice/Industry Training, Simulation and Education Conference, p. 1-9, 2006.

[2] M. W. Scerbo, L. A. Belfore, H. M. Garcia, L. J Weireter, M. W. Jackson, A. Nalu, E. Baydogan, J. P. Bliss, and J. Seevinck, "A virtual operating room for context-relevant training," Proceedings of the Human Factors & Ergonomics Society 51st Annual Meeting, p. 504-511, 2007.

[3] Nalu, M. W. Scerbo, and L. J. Weireter, "Individual personality characteristics for virtual agents in a virtual operating room," Modsim World, 2007.

[4] Baydogan, L. A. Belfore, M. W. Scerbo, and M. Sauray, "Virtual operating room team training via computer-based agents," International Journal of Intelligent Control and Systems, 14, p. 115-122, 2009.

Improving nursing infection control practices in a virtual intensive care unit

Lydia E. Wigglesworth-Ballard

***Abstract*—Hospital acquired infections are a critical issue for healthcare facilities, health professionals, and patients. The Centers for Medicare and Medicaid Service's (CMS) new conditions for coverage require that ambulatory surgical centers establish a fully integrated infection control program, which includes training. There are numerous computer simulations used in the training of healthcare personnel. These are primarily computer-based tutorials using simulated cases coupled with reading materials aimed at teaching initial or remedial knowledge on compliance or containment methods. They have limited transfer in promoting individual commitment to infection control. This study focuses on the use of a Simulated Infectious Disease (SID), undetectable in normal lighting conditions, to simulate infections and visually show how infections are spread and the impact of infection control lapses during the training. Ultraviolet light excites the SID substance showing where potential disease spread has occurred and how far it has spread from its original location. Simpler substances present logistical issues for complex training environments beyond simple hand washing. The specific goals are to incorporate SID into instructional scenarios to evoke a positive reaction and learning experience towards inflectional control, increase knowledge and skills, create a positive behavior change toward infection control practices, and increase commitment to infection control.**

***Index Terms*—Hospital-associated infections, infection control, healthcare simulations**

I. Introduction

Approximately 1.7 million hospital patients each year get an infection while being treated for another condition and almost 99,000 of them die as a direct or indirect cause of their infection [1]. Healthcare-associated infections (HAIs) not only cause increased morbidity and mortality, but also cost the United States almost $5 billion each year due to the extra days or weeks of hospitalization needed to treat the infection [2]. The individual costs can vary greatly depending on patient and type of infection. These costs can range from $600 for a urinary tract infection to $50,000 for prolonged bloodstream infection [3]. The patient's quality of life may also be impacted by HAIs by causing increased antibiotic use, lost time from work, increased financial burden, increase in hospital length of stay, and increased burden to the family. The increased burden to the family can be financial, physically, and mentally if the patient is unable to care for themselves.

Along with the excessive cost and lack of accountability a statement submitted by the Department of Health and Human Service's Center for Medicare Services which stated that beginning October 2008 payments will be withheld from hospitals for care associated with treating certain hospital acquired infections that are seen as highly preventable due to the existence of prevention guidelines [4]. Other insurance companies are following Medicare's lead and are beginning to implement the same cost cutting measures.

Infection control measures that are supposed to be practiced by healthcare workers are intended to prevent spread, transmission, and acquisition of infection between patients, healthcare workers to patients and from patients to healthcare workers. Infection control is a critical concern for patients, healthcare workers, facility administrators, and government agencies. Guidelines to prevent HAIs have been developed by the Center for Disease Control and the World Health Organization. These infection control measures are designed to combat everything from the spread of colds and flu to hepatitis B and C, SARS, HIV/AIDS, and other potentially life threatening diseases. However, there still remains the problem of translating these guidelines into practice.

One of the biggest challenges in implementing infection control precautions is the lack of the ability to assess compliance and identify failures. Most infections have an incubation period, so the infection may not manifest until after the patient is discharged from the hospital. Currently, healthcare workers may only learn about failures through total infection rates documented by the health facility. This makes it very difficult to

Manuscript received March 12, 2011.
Lydia E. Wigglesworth-Ballard is a doctoral student in Health Services Research in the College of Community and Environmental Sciences, Old Dominion University, Norfolk, VA 23529 USA (email: lwigg005@odu.edu).

determine the exact person or persons who are responsible for the transmission.

There are numerous computer simulations used in the training of healthcare personnel for improved infection control practices. These are primarily computer-based tutorials using simulated cases coupled with reading materials aimed at teaching initial or remedial knowledge on compliance and containment methods. They have limited transfer in promoting individual commitment to infection control.

Using SID for instruction and training with patient simulators, health care workers will be able to visually see the potential for disease spread by their actions.

II. Literature Review

Based on the current literature, some of the methods used to currently promote infection control practices in health care facilities are education, routine observation and feedback, reminders, administrative sanctioning, and change in hygiene agent [5]. To this date there is no universal standard for measuring compliance and among the major methods such as direct observation and self-reporting there are disadvantages. Direct observation is a great way to access specific circumstances as to why health care workers do or do not comply with infection control standards, however, they are also labor intensive and increase the chances for bias and limitations, such as the Hawthorne Effect. The Hawthorne Effect is a process where humans may change or modify their behavior because they are being observed. Another method, such as self-reporting, is one of the least expensive, but many previous studies have been called into question due to validity issues [6].

Several studies have concluded that the lack of adherence to infection control practice is due to several barriers. In the Nursing 2007 infection control survey report that was conducted to explore nurse's perceptions about infection control guidelines and standards, it was discovered that experience does not predict infection control knowledge [7]. A nurse's perception regarding hand hygiene and infection control is shown to affect their behavior. Among the nurses who participated in Nursing 2007 survey, 76% believed that if the patient did not show signs of an infection while under their care, then they believed they were doing a good job of preventing infections. This becomes a problem in the area of infection control because most infections have an incubation period, so the nurse may never know of the infection due to the shorter lengths of stay in the hospital [7]. Nursing staff perceptions in an Atlanta Veterans Affairs long term facility reported that 59% of the nurses perceived MRSA as being a threat to the patients and perceived an even less risk to themselves [8]. This study may point to the aim of infection control practice directed to protect the patient, however, another study observed, at two different intensive care units, that nurses were consistently better at hand hygiene post-contact as compared to pre-contact [9]. These results show just the opposite in that hand hygiene for some nurses may be for self-protection. Nurses perceived barriers to hand hygiene compliance as long hours, high workload, understaffing, skin conditions that are irritated by frequent hand washing, lack of knowledge, inaccessible supplies, gloves providing a false sense of comfort as a replacement for proper hand hygiene, and ignorance to guidelines [10].

III. What is SID?

Polyethylene microspheres, known as SID (Simulated Infectious Disease) were custom developed with a specific size, color, UV reflectance and shape. SID is not detectable under normal light in small amounts and has the ability to fluoresce under ultraviolet light. Due to the spherical nature there is the ability to distinguish between fluorescing SID microspheres and fluorescent foreign material. SID has been assessed and evaluated by the Cosmetic Ingredient Review (CIR) Expert Panel as a safe ingredient for use in cosmetics. The Food and Drug Administration has also approved this substance for use in chewing gum and food additives.

After the development of SID, initial tests were done to test its behavior in the environment. The microspheres were tested for the ability adhere to skin surface, solid surfaces, i.e., countertops, walls, computer equipment, and porous surfaces, i.e., fabrics. It was also tested for the ability to wash off using standard hand washing guidelines, to be removed from hands using alcohol gel, to transfer under direct and indirect contact, and to be able to clean up easily with common cleaning methods, which includes using cleaning wipes.

The SID microspheres did adhere very well to human skin. When one then touched a surface, some of the material transferred to the surface touched while some of the material remained on the skin. SID was also able to transfer with subsequent touching up to eight times depending on the type of surface. The material performed very well on non-porous surfaces but not as well on fabrics. After placement on the hands, washing was performed using a quick 5 sec wash, which showed that some traces of the material remained on the hand. Nearly all of the SID microspheres was removed from contaminated hands when using the standard hand washing guidelines. Very tiny traces remained under nails or on jewelry but were removed with a second hand wash. Alcohol-based hand rub only spread the material around on the hands and did not remove it. The material cleaned up very well on surfaces using cleaning wipes, water and paper towels, or cleaning solution and paper towels.

Simpler substances have been tested and present logistical issues for complex training environments beyond simple hand washing. Simpler substances are not suitable for repeated test, some are visible in daylight, and their consistencies (i.e. liquid makeup, and mediums) may mechanically or cosmetically damage delicate medical equipment and simulators. SID is very safe and does not pose any potential danger to any trainee or the delicate equipment that is used for medical simulations.

IV. Pilot Study

The author is currently conducting a pilot study at Old Dominion University School of Nursing. The purpose of this study is to examine if SID can be an effective training tool to visualize common infection control lapses, increase knowledge, and influence infection control practices. The Old Dominion University Institutional Review Board (IRB) application was approved for this study.

The specific aims for this study is to (1) incorporate SID into training scenarios using patient simulator mannequins to train nursing students in an interactive simulation environment by visually replicating spread of infectious bacteria while performing common clinical tasks in a virtual setting (2) evaluate healthcare student's performance by measuring the magnitude and distance of SID spread from the original location photographically, (3) compare outcomes of infection control knowledge, opinions, and behaviors after completing the training with SID, and (4) assessing long-term commitment to infection control in subsequent training.

V. Kirkpatrick's Levels of Evaluation

Kirkpatrick's Levels of Evaluation model will be used in this study to evaluate the simulation instruction and training (see Figure 1). Donald Kirkpatrick published a series of articles in 1959, where he described a four-stage model for evaluating training programs [11]. This model is well known in educational and instructional design and has been widely used for adult training programs.

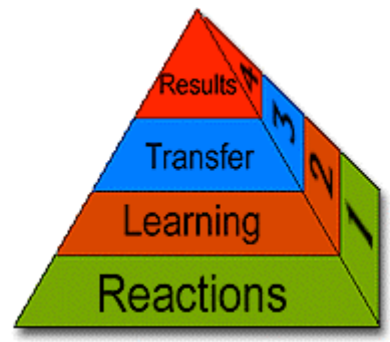


Figure 1. Kirkpatrick's Levels of Evaluation

Kirkpatrick identified four levels for training evaluation. Level I is known as the reaction stage. This level measures the reaction of the trainee to the training program in an effort to determine if the learners like the training, if it was a positive experience, and if the training was motivational. Kirkpatrick noted that in the reaction stage just because a learner has a positive experience it might not necessarily guarantee a learning outcome. Level II is the learning stage. This level measures if there was an attitude change, increase in knowledge or acquisition of new skills. Level III is the transfer stage. This stage measures the transfer of training and if the learner's new knowledge or acquired skills are being implemented in their workplace. Level IV is the result stage. This stage is measured at the organizational level and determines if the desired outcome of the training was achieved.

VI. The Role of SID in Infection Control Training

A. *Incorporation into training scenarios*

The scenario will require the student to react as he or she would in real-life circumstances. It is important the scenarios not differ from the experiences in the actual clinical setting that the student trainee normally works in, which could affect performance during the simulation. Applying scientifically acceptable infection principles are also important. The scenarios developed for this study include students interacting with both infected and uninfected simulated patients. SID was placed in areas on the patient simulator mannequins where infectious bacteria normally colonize. The scenarios consist of wound change of dressing, head-to-toe assessments, which includes taking vital signs, and cleaning of a tracheotomy.

B. Measurement of Spread

Once the scenarios were completed, photographic images were taken of the patient simulator mannequin and surrounding work area under ultraviolet light showing the distance and the magnitude of spread from its original location (see Figure 2).

Video camera recordings were also taken of the instructional session. The video recording will allow the author to monitor the student's performance and demonstration of skill. It will also be able to show the number of contacts the student has with the patient and surrounding area therefore giving the number of times the bacteria is potentially reintroduced to another part of the patient's body or surrounding area.

Figure 2. SID spread on Mannequin under UV light.

Once all data is collected the mannequins and surrounding areas used during the scenarios were cleaned with commercial cleaning wipes or water and paper towels to remove all traces of the SID substance. Before any future scenario is performed the area is scanned with ultraviolet light and photos taken to ensure there is no SID present and to provide a baseline in case any SID was missed during the previous cleaning.

After a thorough review of the photographs and videos a quantitative value will be assigned to each photograph by using a 3 x 6 grid that is superimposed over the photograph of the mannequin (see Figure 3).

Figure 3. Grid for calculating contamination

As shown in Table 1 each cell of the grid will represent an area on the mannequin giving 18 cells. There will be an additional 5 cells added to the chart for grid numbers and corresponding areas. These 5 cells will represent areas not immediately around the mannequin such as the computer, table, and monitors. This value will be derived by assigning a score of 1 to each cell of the grid that is contaminated. The total value of the photograph will range from 0 (no contamination) to 23 (contamination in all cells).

Grid Number	Area covered
1	Upper left bed area, includes top and left frame & pillow
2	Upper middle bed area, includes top frame & pillow
3	Upper right bed area, includes top and right frame & pillow
4	Upper left bed area, includes left frame and pillow
5	Patient's head and throat
6	Upper right bed area, includes right frame and pillow
7	Left bed area, patient shoulder and right arm, part of bed frame
8	Patient's chest area
9	Right bed area, patient shoulder and right arm, bed frame
10	Patient's right arm and surrounding bed area
11	Patient's lower torso and surrounding bed area
12	Patient's left arm and surrounding bed area
13	Lower bed area, near patient's right leg
14	Patient's legs, knees and surrounding bed area
15	Lower bed area, near patient's left leg
16	Lower left bed area, includes patient's right foot
17	Lower bed area, includes patient's inner portion of feet
18	Lower bed area, includes patient's left foot
19	Ambu-bag
20	Blood pressure cuff
21	Keyboard
22	Monitor
23	Table

Table 1. Grid numbers and locations

C. *Debriefing*

The students will then take part in a debriefing session that will include disclosing the presence of SID during the instruction, showing pictures of the locations were the SID was planted, taking a walk through the space to show the contamination, using the ultraviolet light to highlight any presence of SID on the student's hands and uniform, and providing additional instruction and feedback on ways to correct and prevent spread. The debriefing will conclude with a questionnaire asking how they felt about the training and if the training motivated them. The students will also have a chance to ask any question and express any concerns. By visually demonstrating the degree of contamination that nurses can be exposed to everyday, the goal is that they will be more motivated to improve infection control behaviors and apply them in their workplace.

VII. Discussion

One of the biggest challenges in understanding infectious disease is estimating the probability of spread. Pathogens are not visible to the naked eye, making it nearly impossible to estimate rates of spread. By incorporating SID with patient simulators, students and trainees will be able to visually see the potential for disease spread by their actions.

The role of debriefing is important in simulation-based training. Reflection on an event and subsequent analysis is the cornerstone of the experiential learning process and leads to life-long learning. However, the capability of naturally being able to analyze and process the event does not come naturally to everyone leaving gaps between experiencing the simulation training and making sense of it. Debriefing allows for facilitated reflection thus closing theses gaps.

Incorporation of SID into simulations for infection control instruction and training will lead to improved practices for nursing students who are getting ready to enter the workforce. This simulation can have a great impact on patient safety and lead to the ultimate goal of zero infections.

VI. Future Work

Future research will be conducted with nurses in a medical facility. Unlike the educational setting, many of these nurses will have a wide variety of educational experiences, year of practice and backgrounds.

To facilitate further studies additional variations of SID can be can be produced to include variable colors under ultraviolet light to help differentiate the source of the disease and other visibility capabilities such as infrared.

Acknowledgment

This work was conducted with the assistance and support of Dr. Holly Gaff and is gratefully appreciated. An appreciation and thank you also to Dr. Ginger Watson and Dr. Yiannis Papelis for assistance with this research.

REFERENCES

[1] R.M. Klevens, J.R. Edwards, M.S. Chesley et al., "Estimating health-care associated infections and deaths in U.S. Hospitals, 2002, *Public Health Rep*., Vol. 122, No. 2, pp. 160-166, Mar-Apr 2007.

[2] L. McKibben, T. Horan, J.L. Tokes, et al., "Guidance on public reporting of healthcare-associated infections: Recommendations of the healthcare infection control practices advisory committee", *American Journal of Infection Control*, Vol. 33, No. 4, pp. 217-226, May 2005.

[3] M. Hassan, H.P. Tucker, R.H. Patrick, D.S. Koutz, J.L. Kohn, "Cost of hospital-acquired infections", *Hospital Topics*, Vol. 88, No. 3, pp. 82-89, 2010.

[4] Department for Health and Human Services Center for Disease Prevention, Healthcare Infection Control Practice Advisory Committee, Atlanta, GA, February 11-12, 2008.

[5] The Centers for Disease Control and Prevention, "Guideline for hand hygiene in healthcare settings: Recommendations of the healthcare infection control practices advisory committee and the HICPAC/SHEA/APIC/IDSA hand hygiene task force", *Morbidity and Mortality Weekly Report*, 52 (RR-16), pp. 1-56, 2002.

[6] J.P. Haas and E.L. Larson, "Compliance with hand hygiene, American Journal of Nursing, Vol. 108, No. 8, pp. 40-43, 2008.

[7] K.M. Delahanty and F.E. Myers, "Infection Control Report", *Nursing 2007*, Vol. 37, No. 6, pp. 28-37, November 2010.

[8] R.L. Wolf, D. Lewis, R. Cochran, and C. Richards, "Nursing staff perceptions of Methicillin-resistant *Staphylococcus aureus* and infection control in a long-term care facility", *Journal of the American Medical Directors*, Vol. 9, No. 5, pp. 342-346, 2008.

[9] A. Bahal, D. Karamchandani, A.P. Fraise, M.L. MeLaws, "Hand hygiene compliance: universally better post-contact than pre-contact in healthcare workers in the UK and Australia", *British Journal of Infection Control*, Vol. 8, pp. 24-28, 2007.

[10] D. Pittett, "Improving adherence to hand hygiene practice: A multidisciplinary approach, *Emerging Infectious Diseases*, Vol. 7, No. 2, pp. 234-240, 2001.

[11] D.L. Kirkpatrick, "Evaluating training programs: The four levels, San Francisco, CA, Berrett-Koehler, 1994.

Using Modeling and Simulation to Improve Oral Health Services Delivery in Norfolk Public Health District Dental Clinic, Little Creek

Author(s): Mohammad Alzahrani, Holly Gaff, Deanne Shuman, Rani Kady

Keywords: oral healthcare delivery, public health performance, modeling and simulation, public health district dental clinic

Abstract: The purpose of this study is to examine the system performance in delivering oral health services in a public health district based on the Conceptual Framework to Measure Performance of the Public Health System (PHS). Using modeling and simulation, a predictive model based on the conceptual framework dimensions: mission, structural capacity, processes, and outcomes was developed to predict the performance of public health district in delivering oral health services.

This is a retrospective longitudinal study. Independent and dependent variables under investigation were categorized based on a comprehensive conceptual framework, Public Health System Performance Framework (PHS) dimensions.

The main objective of this study is to use the modeling and simulation approach to developed a simulating model to predict the performance of public health district dental clinic in delivering oral health services. Specifically, the following performance metrics were examined: the average number of patients; visits per day at a public health district dental clinic, the average number of diagnostic and preventive dental services delivered by the dentist or the dental hygienist per day at a public health district dental clinic, the average number of corrective services provided by the dentist per day at a public health district dental clinic and the average total dental services. The scenarios based on the existing structural capacities and personnel were modeled and simulated using Rockwell Automation software, Arena ® version 13.5.

Purposeful sample consists of previously selected homogeneous five public health district dental clinics of Hampton Roads for the fiscal years, 2005- 2010. For the purpose of this study the following five public health district dental clinics were chosen: Norfolk, Virginia Beach, Hampton, Peninsula, and Western Tidewater. Norfolk Health District operates two sites: Little Creek and Park Place. Virginia Beach District operates two sites: Birdneck and Pembroke. Western Tidewater Health District operates two sites: Isle of Wight and Southampton.

Data analysis revealed that adding a new healthcare provider (a dental hygienist) to the system has a statistically significant increase in delivering oral health services at a public health district dental clinic in the following performance metrics: diagnosis and preventive services, corrective services, total number of dental services and number of patients; visits per day ($p < 0.05$).