Stuetzle Research Summary - February 2016

Christopher S. Stuetzle

July 27, 2016

1 Geographic Information Science

1.1 Abstract of the Project

Terrain surface data is relevant to a large variety of applications in Geographic Information Science (GIS), including observer siting (where to place cell-phone towers), GPS, floodplain mapping, erosion, and formation simulation, among many others. Current representations of terrain surface data are severely lacking when it comes to accuracy, flexibility, and validity, and as such limit the applications to which they can be applied. This project aims to extend an already-established suite of mathematical operators that can be used to model terrain surfaces in a procedural manner, useful in applications such as gap filling, data compression, storage, and transmission, while simultaneously providing a more accurate, realistic, and hydrologically valid representation of terrain data. The work also focuses on feature-aware error metrics used to validate the new model of terrain, as well as the development of parallel algorithms to operate on terrain.

1.2 Goal of the Project

This project (currently) has the following goals:

1. **Long Term:** To develop a deep, robust, and accurate mathematically-based representation of terrain surface data that is based on real world processes (such as digging or eroasion) and can be used to store terrain as a procedural representation.
2. **Long Term:** Once this representation is established, standard GIS algorithms (such as observer siting, path planning, gap filling, watershed mapping, etc.) will be adapted to work with the new representation.
3. **Short Term:** Since this representation will need to scale due to the ubiquity of "big data", or very large terrain datasets, parallel versions of common GIS

algorithms must be developed in order to accomodate this new data. This means that, in the short term, some common parallel algorithms will be investigated. Specifically, gap filling and watershed mapping, but also observer siting. This is the primary subject of Summer 2016's FDG.

1.3 State of the Project

- In August 2012, I presented "Representing Terrain With Mathematical Operators" at *Spatial Data Handling 2012* in Bonn, Germany. This paper was published in 2013.
- For two years, I moved on and focused on other work (described below - Human-Computer Interaction).
- In summer, 2015 I was awarded a faculty development grant to pick this work back up. I ported the code I had written in graduate school to a friendlier version, made effiency upgrades in the running time of algorithms and the data structures, and took some new screen shots. Summer 2015 laid the groundwork for much of the work I was to continue until Tenure.
- In spring 2016, I was contacted by my graduate adviser, an expert in the field, and after some discussion we came to the conclusion that the work I was focusing on (namely the "Drill Operator" and all GIS-centric algorithms using that terrain representation) was too small scale to be universally interesting in the area of GIS. Instead, he instructed me to focus on Big Data and push toward the long term goal of a deeper and more universal representation of terrain surface data. This setback will bear fruit eventually, but it will take a while for the program to be up-and-running.
- In spring 2016, Jacob Wilkins finished a directed study with me working on measuring a series of terrain-surface error metrics and visualizing them. This project will directly inform future work in the area in terms of being able to measure the error of gap filling.
- I received a Faculty Development Grant (FDG) to course-correct for summer 2016, and will be working on parallel algorithms in GIS with regard to hydrology and terrain surface representation during that time.

1.4 Future of the Project

Over the next two years, I will attempt develop a series of parallel algorithms for previously un-parallelized common procedures in GIS (as described above). This summer, the groundwork will be laid for this project, including finding and adapting big datasets, learning CUDA and OpenMP APIs, developing fundamental shared-memory parallel applications, and implementing several already-existing algorithms. Once performance

baselines have been determined by this process, parallel algorithms will become the focus.

By the end of summer 2016, I hope to have at least one of these algorithms completed. Over the next year beyond that (by the end of summer 2017) more will follow suit. There is a demand for this work in the GIS community and it, assuming it is successful, will be publishable.

2 Pedagogy

I have developed several novel pedagogical tools over the course of my classwork development. Two such techniques have been published:

- The Human Computer (CCSCNE 2014) - An in-class exercise (witnessed by Mary Noonan) in which students play the role of subroutines in a computer program. They pass data back and forth on index cards, which represents parameter passing (both by reference and by value) and return statements. They can declare variables with new index cards. The students are asked to stand and solve a semi-complex problem with this method, as the teacher looks on and corrects issues as they aris.
- Public Debate Format for the Development of Soft Skill Competency in Computer Science Curricula (CCSCNE 2015) - A paper describing the use of public debates of controversial topics in a human-computer interaction class. Several soft skills are practiced and developed during public debates, and this paper argues in favor of incorporating these soft skill competencies in computer science curricula.

Over the next two years I plan to write two more pedagogy papers:

- How Do Humans Sort? - A study chronicling and investigating how humans naturally sort objects. Cards (with words and also with numbers) will be arranged for student subjects to sort on tape, and the sorting technique the students use will be identified. We will analyze the sorting methods used and determine whether textbooks that claim that "insertion sort" is the "natural sort" are correct, or whether our intuition about sorting may be incorrect. This work is planned with Zachary Kissel.
- Grade Weights - I use a unique weighting process when determining the grades for my classes, in that I do not weight anything and instead rely on blanket point values to naturally determine weighting. I will survey students have had both myself and other members of the department to determine which grading technique is their preferred method, and why, and make a strong case in favor of a weight-less points-based grade system.

3 Democratization

3.1 Abstract of the Project

This project brings computer simulation techniques, specifically cellular automata, to the study of democratization and democracy. While there have been some attempts to bring and use simulation in Comparative Politics, the field leaves much unexplored ground, some of which we hope to cover. We model several competing mechanisms of regime change, both internal (for example economic development ala modernization) and external (neighborhood effects and diffusion mechanisms), and illustrate how their interaction can produce a world similar to ours in key dimensions - such as spatial interconnectedness and the overall level of democracy in any given year.

Dr. Bilev presented our preliminary results at a panel titled ‘‘Diffusion and Regime Change’’ within the ‘‘Comparative Democratization’’ section of the flagship Political Science national conference in early September , 2015 (APSA) in San Francisco. We received feedback from the panel discussant and comments from two of the other presenters. The main recommendations were to include more data or simply utilize additional datasets; to try to conceptualize what a ‘‘neighbor’’ means in different ways (one of the other papers on the panel was on cultural neighbors); and to very carefully explain which portions of the paper use observed data from the real world and where we deviate with the computer simulation and what the benefits of simulation are in this case. Overall, the feedback was positive and encouraging.

3.2 State of the Project

- In summer 2015, we built an intialize simulation of democratization that was used to determine the level of “external influence” vs. “internal influence” on the regime value (Democracy or Autocracy) of a particular country. The external influence has long been ignored or marginalized in political science literature, and this project aimed at determining if this was the right approach and, if not, what the right weighting value is.
- In fall 2015, Gavril presented our initial abstract at APSA (as stated above) and received a great deal of positive feedback on the future of the project.
- In spring 2016 I submitted the following abstract to the International Conference on Information Technology and Computer Science (ICITCS), taking place in Athens, Greece. It was accepted, and I will be presenting this work in May 2016.

Title: Adapting Cellular Automata Simulation Techniques to the Study of Democratization

Author: Christopher Stuetzle, Assistant Professor of Computer Science, Merrimack College, USA (stuetzlec@merrimack.edu)
Author: Gavril Bilev, Assistant Professor of Political Science, Merrimack College, USA (bilevg@merrimack.edu)

This work adapts popular techniques in cellular automata theory and application to the study of the spread of Democracy in the world, or Democratization. Much work in Comparative Politics has focused on determining the most important factors that lead to a country's regime (government) being either Autocratic or Democratic. We use a cellular automata-based simulation to model the spread of democracy from 1945 to 2011 (years for which comparison data is available), whose time steps are individual years. The two primary sources of influence on a country's regime value are Internal and External. Internal influence is determined by the Gross Domestic Product (GDP) of the country in question as well as a linear fitting of known data for a variety of parameters (including the country's stability, religious populations, and age), and the external influence is set by the regime value of all countries (weighted by their Internal influences) within its 'neighborhood.' Our simulation treats each country as a cell whose state is governed by a linear weighting between Internal and External influences, and its resistance to change by its own GDP (the higher it is, the more resistant to regime change).

A parameter sweep method is employed to test ranges of the parameters (each unique set of parameter values is called a parameter set) used to determine the regime score of each country each time step. These tested parameters are weights on internal vs. external influence, neighborhood sizes, neighborhood types, extent to which outside democracies are considered, and stability weighting. Each parameter set produces a configuration of the world in 2011 that is compared against the actual world configuration and a series of error values are calculated using Percent Correct, Moran's I, and Percent Democratic. 1000 trials are run for each parameter set and the median regime values are used. The global minimum error for each metric is calculated, determining the parameter set that most resembles those which are responsible for democratization in the world.

3.3 Future of the Project

- In spring 2016, Gavril and I are writing and submitting a paper to Political Geography, a well-known journal in Political Science that our work fits well. The paper will present our initial findings of the necessity of including an external influence weighting into a study of democratization. Authoring of this paper is currently underway.
- In spring 2016, students in CSC4910 Software Engineering have taken on the task of designing a new simulation for our project in which parameter sweeps will be used to determine, from a larger pool of parameters and values, which parameters are most important in determining the regime value (Democracy or Autocracy) of a particular country. These parameters include neighborhood (and how neighbors are determined), internal vs. external influence weighting, starting regime value, gross domestic product (GDP), and a number of other factors. With this more powerful and flexible tool (students have a great deal more time to work on it than we do when it is the primary focus of their semester) we hope to be able to draw even more substantial and influential results.
- By spring 2017, we hope to have taken these more influential and important results and converted them to another journal publication in Political Geography.

4 Human-Computer Interaction

4.1 Abstract of the Project

Abstract from most recent submission:

```
We present a system for developing multi-user, multi-cursor, multimedia,
single display groupware that allows for intuitive interaction
with visual graph datasets. Our software framework separates
the application and the input device modules with an interface that
handles generic gestures captured from different input devices. The
system's gesture set was designed to handle standard actions for
graph data interaction. To enhance the collaborative group interaction
experience for ease of graph data exploration, we present an automatic
method for adjusting the global focus and zoom level. Our
system allows simultaneous input from works for multiple USB
mice and laser pointers. To utilize laser pointers as input devices,
we also present the idea of a persistent laser personality of its infra-
red leak. Our system is extensible and scalable due to its heavy
emphasis on modularity, and works on both single output screens
and large-scale projection surfaces.
```

4.2 State of the Project

- In summer 2013, I was funded by my first FDG (entitled "Using Laser Pointer Personalities in Multiple User Interaction") to work for several weeks with Barbara Cutler from Rensselaer Polytechnic Institute on this laser-interaction research. During that summer, we produced a paper entitled "Identifying Inexpensive Off-the-Shelf Laser Pointers for Multi-User Interaction on Large Scale Displays" which was, during that time, rejected from Graphics Interface 2014.

- During summer 2014, we continued to expand and improve upon this work in my second FDG (entitled "Multiple User Interaction for Graph Datasets"), and by the end of the summer had produced an extended paper entitled "Multimedia Groupware for Graph Interaction and Visualization". This paper was rejected from Graph Interaction 2015.

- In Summer 2016, we submitted "Multimedia Groupware for Graph Interaction and Visualization" to SUI2016.

4.3 Future of the Project

This project is no longer actively being pursued and extended. However, there are two plans for the extensive work that has already been done.

1. The second paper ("Multimedia Groupware for Graph Interaction and Visualization") has been submitted to Graphics Interface 2016 and I am currently waiting to hear word of acceptance or rejection. If the paper is rejected, there are several other possible publication venues (International Conference on Advances in Computer-Human Interaction, International Conference on Computer Graphics and Visualization, or International Working Conference on Advanced Visual Interfaces, for example). All of these conferences and publications have due dates during 2016. This paper will be published somewhere before tenure.

2. Dr. Cutler also mentioned converting some of the work our partner did for a Master's thesis into a conference publication, of which I would be the second author as much of my work contributed to his thesis.

5 MerMAID Research Group

5.1 Abstract of the Project

```
During the course of this interdisciplinary effort, members of the
Electrical Engineering (EE), Computer Science (CS) and Mechanical
Engineering (ME) departments collaborated on energy related curricular
efforts.  Initially this was an EE/CS collaboration, with work carried
out developing and utilizing an inexpensive, open-source system for
```

```
measuring, storing, and displaying energy related data from across
campus.  Hardware and software components were chosen to be
open source or free for educational use. A low power Linux server
was utilized.  LAN enabled Arduinos included sensors to measure
energy related quantities such as power and temperature.  EE and
CS students were engaged in various aspects of the project { EE
focusing on the hardware, CS on the programming.  EE juniors worked
with clients to implement real world measurement and display solutions.
A CS student project focused on developing a JavaScript-based web
paged that visualizes sensor data by leveraging CanvasJS and JQuery
packages.  This web page development project will continue in spring
2016 as the work is significantly incorporated into the CS department's
Web Development class.  More recently, collaboration between the
Mechanical and Electrical Engineering Departments resulted in the
integration of this work with courses in the ME curriculum.  Mechanical
engineering students utilized the campus as a \living laboratory" as
part of coursework for the Heat and Mass Transfer and Thermo-
dynamics courses in spring 2015.  Students observed working HVAC
systems and were able to link coursework to the operating cycles of the
heating and cooling pumps and heat exchangers in various buildings on
campus.  Students were assessed via survey: response to this living lab
experience was overwhelmingly positive.  Use of campus as a living
laboratory will be extended in spring 2016: students will evaluate pump
energy usage using data collected by EE students and calculate energy
savings due to potential pump replacement.  Additionally, EE junior
projects currently underway (fall 2015) emphasize collaborations across
a wide variety of disciplines: projects include wetland environmental
factors (Biology), greenhouse environmental factors (Biology), pump
energy usage (ME), weather monitoring (Physics), classroom temperature
monitoring (Facilities).
```

5.2 State of the Project

- In spring 2016, we submitted a paper entitled "Incorporating Energy Related Concepts into EE, CS, and ME coursework with emphasis on Interdisciplinary Collaboration", submitted to the 2016 American Society for Engineering Education (ASEE) National Conference. It was conditionally accepted, but we decided to pursue other avenues.
- In spring 2016, a group of capstone students have selected the MerMAID research project as their semester-long project. During this time, the students will migrate all MerMAID server information to its permanent home on the CS Department Server. They will also redesign the web portal and data visualization for the sensor data.

- In spring 2016, we submitted a smaller version of the paper, entitled "Incorporating Energy Related Concepts into EE and CS laboratory work and coursework" (omitting the ME portion of the work) to ASEE's conference on "Revolutionizing Engineering Education". It was accepted and Prof. Adams presented the paper at the conference in Rhode Island April 28th-30th.

5.3 Future of the Project

- In fall 2016 or spring 2017 I hope to collaborate on another ASEE paper as an update and, hopefully, a recount of the success of the student collaboration wherein BSIT students were able to built and administrate the MerMAID system and, in the end, create a portable and robust system whereby other institutions can simply unpack and install a system that is ready to use given arduino boards housing environmental sensors.

6 Compression of Time-Lapse NEXRAD RADAR Image Data

During Summer 2016, I began a project to compress time-lapse NEXRAD RADAR images (like Doppler images). This work was begun in conjunction with Distributed Communications Systems, Inc., working on contract whose details are below.

```
Air Force Weather Phase II:
Integrating Tactical Weather Sensors with Mobile Devices

Contract No. FA8750-16-C-0172
```

6.1 Problem Statement

We would like to encode NEXRAD (or other) time-lapse weather imagery into as few bits as possible (3000 bits) using the Weather-Huffman (WH) algorithm introduced in [4] and analyzed and described in Document 1. We assume the input to this algorithm is a series of images representing time-lapse RADAR imagery, whose pixels have one of six (for precipitation, more for others) possible values representing weather severity (0 = none, 5 = severe).

6.2 State of the Project

FILL IN JANUARY 2017

6.3 Future of the Project

FILL IN JANUARY 2017

7 Encrypted Operations of Spatial Datasets

Dr. Zachary Kissel and I have been discussing possible processes and applications of mathematical operations performed on encrypted spatial data. For instance, given an encrypted terrain dataset that lives on the cloud, can you find the maximum elevation without decryption and without data leaking into the cloud?

8 Memory-Efficient Dictionary Data Structure

Dr. Zachary Kissel and I spent three weeks during summer 2015 building on an idea we had regarding memory-efficient dictionary data structures that store a language of strings. The data structure would store the characters of the langauge as nodes of a graph where temporal edges connected two nodes if a particular transition (from index i to index j) in at least one of the strings in the stringset contained the transition from the first node's character to the second. We made good progress until discovering a paper that proved theoretically that the memory bound we were shooting for was impossible to achieve, and thus the project was abandoned.

9 Motivational User Interface Design

Drs. Courtenay Dunn-Lewis, Allison Seitchik, and I are exploring user interface designs and methodologies that could help motivate users to exercise. We are still in very preliminary stages but are looking forward to exploring further. I have met with Dr. Seitchik and mapped out a strategy for exploring starting Fall 2016.

10 Summary Timeline of Research - Publications and Presentations

- Fall 2012 - Presented "Representing Terrain with Mathematical Operators" at *Spatial Data Handling 2012*
- Spring 2013 - Published "Representing Terrain with Mathematical Operators" in *Spatial Data Handling 2012*
- Summer 2013 - Awarded Faculty Development Grant for project entitled "Using Laser Pointer Personalities in Multiple User Interaction"
- Fall 2013 - Awarded Faculty-Led Initiative Award for project entitled "Powerful, Elegant, and Inexpensive Open System to Measure, Store and Visualize Energy Data Computer Science", work with Jack Adams
- Fall 2013 - Gave invited talk "Not Just For Making Pretty Pictures" at Northeastern University's ACM Student Chapter
- Spring 2014 - Published and presented "The Human Computer" at *Consortium of Computing Sciences in Colleges, Northeast Region*

- Spring 2014 - Gave invited talk "Diggin' It: Procedural Representation of Terrain with 'Drill' Operations" at Wheaton College, Norton, MA.
- Summer 2014 - Awarded Faculty Development Grant for project entitled "Multiple User Interaction for Graph Datasets" (a continuation of FDG from 2013)
- Spring 2015 - Published "Computer Erosion Modeling Considering Soil Hydraulic Conductivity" in *Journal of Geotechnical and Transportation Engineering*
- Spring 2015 - Published and presented "Public Debate Format for the Development of Soft Skill Competency in Computer Science Curricula" at *Consortium of Computing Sciences in Colleges, Northeast Region*
- Summer 2015 - Awarded Faculty Development Grant for project entitled "Representing Terrain Surface Data with Procedural Mathematical Operators"
- Fall 2015 - Began a directed study with Jacob Wilkins to investigate terrain error metrics and their visualizations. Continued through Spring 2016.
- Spring 2016 - Students in CSC 4910 developed advanced Democratization simulation as part of CSC 4910 Software Engineering Course.
- Spring 2016 - Students in CSC 4920 set up MerMAID web portal for storing, accessing, and visualizating data from MerMAID arduino sensors.
- Spring 2016 - Zach Kissel and I began data collection for our sorting study.
- Summer 2016 - Applied for Faculty Development Grant for project entitled "Development of Novel Mathematical Model for Terrain Surface Data' (a continuation of FDG from 2015)'
- Summer 2016 - Presented "Adapting Cellular Automata Simulation Techniques to the Study of Democratization" at the International Conference of Information Technology and Computer Science in Athens, Greece.
- Summer 2016 - Submitted "Multimedia Groupware for Graph Interaction and Visualization" to SUI2016.
- Summer 2016 - Began work on compression of NEXRAD weather data, plan to continue this through Fall 2016.

11 Summary of Future Goals

- Spring 2016 - Submit first Democratization paper to Political Geography.
- Summer 2016 - Publish HCI paper on cursors.

- Summer 2016 - Lay groundwork for parallel GIS algorithms for gap filling, path planning, and watershed mapping.
- Summer 2016 - Submit "Adapting Cellular Automata Simulation Techniques to the Study of Democratization" to ATINER.
- Fall 2016 - Submit sorting study paper to CCSSCE 2017.
- Fall 2016 - Submit laser personality paper to SIGSCE 2017.
- Spring 2017 - Submit 2nd Democratization paper to Political Geography journal.
- Spring 2017 - Submit 2nd MerMAID education paper to ASEE National Conference.
- Summer 2017 - Submit extensive parallel algorithm paper to SIGSPATIAL (ACM GIS) 2017.
- Fall 2017 - Submit grade weight paper to CCSCNE 2018.
- Fall 2017 - Submit GIS summary of all work to date in terrain modeling to GIS Journal TBD.
- Spring 2018 and beyond - Focus attention on deep mathematical analysis of terrain surface data, adopting parallel and non-parallel algorithms alike, along with statistical measurements and the already-established "drill operator" from SDH 2012 to work on developing a deep, mathematically rich representation of terrain surface data that is hydrologically valid and produces a suite of procedural steps used to generate and represent terrain data.