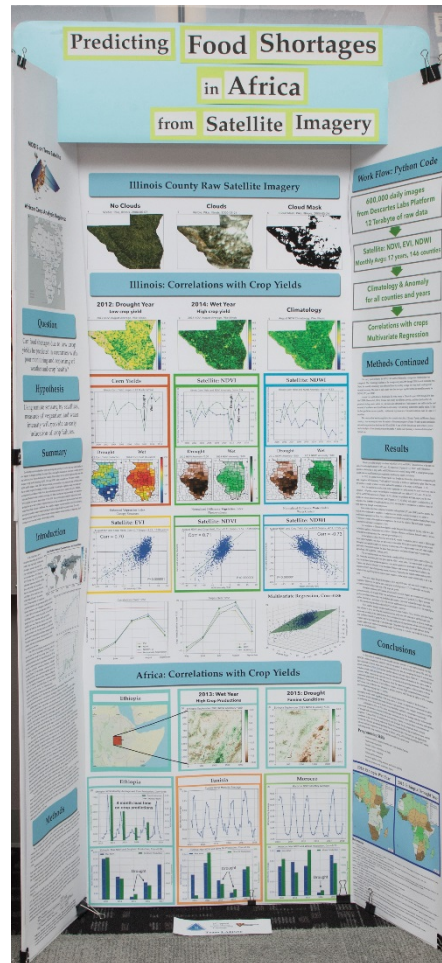




Finalist Reports 2017-2018



www.supercomputingchallenge.org



Printed by the
Los Alamos National Laboratory
High Performance Computing Group (HPC-ENV)

**Cover: Predicting Food Shortages in Africa
from Satellite Imagery**

Los Alamos High School Team 56
Lillian Kay Petersen

Mentors: Daniela Moody and Rick Chartrand
Winner in the Technical Poster Competition



Notification: These final reports are presented in an abridged form, leaving out actual code, color, and appendices where appropriate. Complete copies of most of the final reports are available from the archives of Supercomputing Challenge web site:
<http://www.supercomputingchallenge.org> .

New Mexico Supercomputing Challenge 2017 – 2018 Finalist Reports

Table of Contents

About the New Mexico Supercomputing Challenge

For more information, please visit our website at

<https://www.supercomputingchallenge.org> 2

2017—2018 Challenge Awards 4

Scholarship winners 7

Sponsors 9

Participants 10

Judges 16

Finalist Reports 19

1. *What's Missing?* Los Lunas High School Team 47
2. *Predicting Food Shortages in Africa from Satellite Imagery*,
Los Alamos High School Team 56,
3. *RSA Based Primality Test*, Los Alamos High School Team 71
4. *Epidemiology Simulation*, Academy For Technology and the Classics Team 50
5. *A-Maze-Ing*, Jackson Middle School Team 32
6. *Implementing an Artificial Neural Network to Recognize Handwritten Digits*,
Los Alamos High School Team 201
7. *Spread of Herbicide through Crops*, Monte Del Sol Team 185
8. *Efficient Evacuation of At-Risk Populations*,
Santa Fe High School Team 162 (**Technical Writing Award winner**)
9. *Modeling Straw Mulch in Taos Agriculture*, Taos High School Team 31
10. *MitoSensor*, Taos High School Team 159



Supercomputing Challenge Vision

The Vision of the Supercomputing Challenge is to be a nationally recognized program that promotes computational thinking in science and engineering so that the next generation of high school graduates is better prepared to compete in an information-based economy.

Supercomputing Challenge Mission

The Mission of the Supercomputing Challenge is to teach teams of middle and high schools students how to use powerful computers to analyze, model and solve real world problems.

About the Supercomputing Challenge

The Supercomputing Challenge (the Challenge) is an exciting program that offers a truly unique experience to students in our state. The opportunity to work on the most powerful computers in the world is currently available to only a very few students in the entire United States, but in New Mexico, it is just one of the benefits of living in the "Land of Enchantment."

The Challenge is a program encompassing the school year in which teams of students complete science projects using high-performance supercomputers. Each team of up to five students and a sponsoring teacher defines and works on a single computational project of its own choosing. Throughout the program, help and support are given to the teams by their project advisors and the Challenge organizers and sponsors.

The Challenge is open to all interested students in grades 5 through 12 on a nonselective basis. The program has no grade point, class enrollment or computer experience prerequisites. Participants come from public, private, parochial and home-based schools in all areas of New Mexico. The important requirement for participating is a real desire to learn about science and computing.

Challenge teams tackle a range of interesting problems to solve. The most successful projects address a topic that holds great interest for the team. In recent years, ideas for projects have come from Astronomy, Geology, Physics, Ecology, Mathematics, Economics, Sociology, and Computer Science. It is very important that the problem a team chooses is what we call "real world" and not imaginary. A "real world" problem has measurable components. We use the term Computational Science to refer to science problems that we wish to solve and explain using computer models.

Those teams who make significant progress on their projects can enter them in the competition for awards of cash and scholarships. Team trophies are also awarded for: Teamwork, Best Written Report, Best Professional Presentation, Best Research, Creativity and Innovation, Environmental Modeling, High Performance, Science is Fun and the Judges' Special Award, just to name a few.

The Challenge is offered at minimal cost to the participants or the school district. It is sponsored by a partnership of federal laboratories, universities, and businesses. They provide food and lodging for events such as the kickoff conference during which students and teachers are shown how to use supercomputers, learn programming languages, how to analyze data, write reports and much more.

These sponsors also supply time on the supercomputers and lend equipment to schools that need it. Employees of the sponsoring groups conduct training sessions at workshops and advise teams throughout the year. The Challenge usually culminates with an Expo and Awards Ceremony in the spring at Los Alamos National Laboratory. In 2017 we started alternating the Expo and Awards Ceremony between Albuquerque and Los Alamos.

History

The New Mexico High School Supercomputing Challenge was conceived in 1990 by former Los Alamos Director Sig Hecker and Tom Thornhill, president of New Mexico Technet Inc., a nonprofit company that in 1985 set up a computer network to link the state's national laboratories, universities, state government and some private companies. Sen. Pete Domenici, and John Rollwagen, then chairman and chief executive officer of Cray Research Inc., added their support.

In 2001, the Adventures in Supercomputing program formerly housed at Sandia National Laboratories and then at the Albuquerque High Performance Computing Center at the University of New Mexico merged with the former New Mexico High School Supercomputing Challenge to become the New Mexico High School Adventures in Supercomputing Challenge.

In 2002, the words "High School" were dropped from the name as middle school teams had been invited to participate in 2000 and had done well.

In the summer of 2005, the name was simplified to the Supercomputing Challenge.

In 2007, the Challenge began collaborating with the middle school Project GUTS, (Growing Up Thinking Scientifically), an NSF grant housed at the Santa Fe Institute.

In 2013, the Challenge began collaborating with New Mexico Computer Science for All, an NSF funded program based at the Santa Fe Institute that offers a comprehensive teacher professional development program in Computer Science including a University of New Mexico Computer Science course for teachers.

2017—2018 Challenge Awards

Jen Marie Phifer and Forest Good win top honors, teams from Los Alamos High School take second and third place.



April 26, 2018

28th Annual New Mexico Supercomputing Challenge winners Forest Good and Jen Marie Phifer of Los Lunas High School.

LOS ALAMOS, N.M., April 25, 2018—Jen Marie Phifer and Forest Good of Los Lunas High School won top honors on Tuesday at the 28th Annual New Mexico Supercomputing Challenge held at Los Alamos National Laboratory. Their project, “What’s Missing?” looks at a way to create a systematized method for finding missing pieces of data. The project has implications for the fields of accounting, logistics, and cybersecurity.

Second place went to Lillian Petersen of Los Alamos High School for her project, “Predicting Food Shortages in Africa from Satellite Imagery.” She created a tool to predict crop yields so that international aid organizations can be better prepared for humanitarian relief operations.



Elijah Pelofske of Los Alamos High School took third place with his project, “RSA-Based Primality Test.”

This year, 60 teams representing 26 schools from around the state submitted final reports. New Mexico Education Secretary Christopher Ruszkowski attended the awards ceremony and presented one of the awards to the finalists and also announced that computer sciences courses would now be allowed to count for high school graduation credit.

“The goal of the yearlong event is to teach student teams how to use powerful computers to analyze, model and solve real-world problems,” said David Kratzer of Los Alamos National Laboratory’s High Performance Computing division, and executive director of the Supercomputing Challenge. “Participating students improve their understanding of technology by developing skills in scientific inquiry, modeling, computing, communications and teamwork.”



In addition to presenting their projects at the Laboratory, students took part in tours, talks and demonstrations with Laboratory scientists. A complete list of all winning student teams is available at the New Mexico [Supercomputing Challenge website](https://supercomputingchallenge.org/17-18), (<https://supercomputingchallenge.org/17-18>) and all final student reports are also available online (<https://www.supercomputingchallenge.org/17-18/finalreports/submitted.php?type=Final>).

The Supercomputing Challenge is open to any New Mexico student in grades 6 through 12. More than 300 students and teachers from schools around the state spent the school year researching scientific problems, developing sophisticated computer programs and learning about computer science with mentors from the state’s national laboratories and other organizations.

Scholarships worth more than \$18,000 were awarded to students participating in the Challenge. Many other awards were also distributed ranging from random \$100 gifts for finishing the academic marathon to team prizes for teamwork, research, programming prowess, and environmental impact.

Expo Activities

Students presented their projects to groups of Expo Judges in 30 minute time slots. They displayed their poster in the Santa Clara Gallery for public viewing and then they toured the Los Alamos National Laboratory seeing the supercomputer centers and hearing talks by Laboratory scientists about some of the work being done at the Laboratory. In conjunction with the judging Expo and Award Events, the Supercomputing Challenge held a reception at the Bradbury Science Museum for Challenge participants.

Awards Ceremony

Teams spent the school year researching scientific problems, developing sophisticated computer programs, and learning about computer science with mentors from the state's national laboratories and other organizations. The culmination of the program include written and oral presentations of their projects and recognition for all their work.



All the finalist teams received plaques for their schools, a large banner suitable for hanging at their schools, and other gifts. More than one third of the teams at the Awards Ceremony received some kind of an award. Five random door prizes of a \$100 bill were given out in addition to many “goodies” from Supercomputing Challenge sponsors.



Scholarship winners

More than \$18,000 in scholarships were awarded to 17 seniors:

Robert Borrego from Monte del Sol, Rowan Cahill from Santa Fe High, Matuke Fomukong from Saturday Science and Math Academy, Steven Herrera-Luna from Monte del Sol, Rowan Kinney from Taos High, Oscar Lopez-Rivera from Capital High, Jasmine Marquez from McCurdy Charter School, Andrea Martin from Melrose High, Ramona Park from Santa Fe High, Mackenzie Perkins from Melrose High, Jen Marie Phifer from Los Lunas High, Ryan Romero from Capital High, Rebecca Rush from Melrose High, Ruby Salazar from McCurdy Charter School, Luke Shankin from Santa Fe High, Ben Thorp from Academy for Technology and the Classics, and Hannah Wofford from Melrose High.

Since its inception, the Supercomputing Challenge has engaged more than 11,000 New Mexico students in computational science projects that prepare them for future endeavors in many science and high-technology fields. Past participants have succeeded in private industry and national laboratories.

For more than a quarter century the Challenge has:

- Helped state high school graduates go on to succeed at college in STEM areas;
- improved the information-based economy of New Mexico by promoting computational thinking;
- increased New Mexico’s international competitiveness in smart manufacturing;

- helped middle and high school students meet common core standards with academic excellence in math modeling, science and technical writing;
- and created a culture of collegiality and professional development among the New Mexico educator community.

The Challenge's positive impact on New Mexico is widely acknowledged. See the report by KRWG at <http://krwg.org/post/how-one-small-school-district-embraces-challenge-computer-science-education>.

About the Supercomputing Challenge

The New Mexico Supercomputing Challenge teaches computational thinking, computer modeling and 21st century skills such as research, teamwork, project management, written and oral communication to middle and high school students throughout the state. Any New Mexico middle-school or high-school student, including home schooled students, is eligible to participate in the Challenge. Students conduct research projects on subjects of their own choosing. The Supercomputing Challenge is sponsored by Los Alamos National Laboratory, Los Alamos National Security, LLC, Sandia National Laboratories, and generous industry and educational partners across the state and country. A complete list of sponsors and supporters of the Challenge is on its website (<https://supercomputingchallenge.org/>).

About Los Alamos National Laboratory

Los Alamos National Laboratory, a multidisciplinary research institution engaged in strategic science on behalf of national security, is operated by Los Alamos National Security, LLC, a team composed of Bechtel National, the University of California, BWXT Government Group, and URS, an AECOM company, for the Department of Energy's National Nuclear Security Administration.

Los Alamos enhances national security by ensuring the safety and reliability of the U.S. nuclear stockpile, developing technologies to reduce threats from weapons of mass destruction, and solving problems related to energy, environment, infrastructure, health, and global security concerns.

Sponsors

The Supercomputing Challenge is sponsored by Los Alamos National Laboratory and Los Alamos National Security, LLC.

A Platinum Partner is Sandia National Laboratories.

Commercial partners are Abba Technologies/Hewlett Packard, Cray Inc, DataONE, Lockheed Martin, Air Force Research Laboratory, Gulfstream Group and bigbyte.cc, LANL Foundation, VanDyke Software, Albuquerque Journal, Ideum, New Mexico Network for Women in Science and Engineering, New Mexico Technology Council, Redfish Group, OpenMake Software, The MathWorks.

Educational partners include The Center for Connected Learning (CCL) and Computer-Based Modeling/NetLogo, Central New Mexico Community College Ingenuity, Eastern New Mexico University, MIT StarLogo, New Mexico EPSCoR, New Mexico Highlands University, New Mexico Institute of Mining and Technology, New Mexico Public Education Department, New Mexico State University, NMSU-Dona Ana Community College, NM Technology in Education (NMTIE) in conjunction with NMCHECS, Northern New Mexico College, San Juan College, Santa Fe Community College, the University of New Mexico and the UNM Center for Advanced Research Computing.



Locations of the 2017-2018 Supercomputing Challenge teams

Teams Finishing the Challenge and submitting final reports:

Team ATC50, Academy For Technology and the Classics, *Infectivity Model*

Team Members: Alex Baten, Ben Thorp, Teddy Gonzales

Sponsor: Jenifer Hooten, Mentor: John Thorp

Team ATC55, Academy for Technology and the Classics, *A Safer Way to Track Forest Fires*

Team Members: Etta Pope, Mindy Bilbo, Savannah Valerio, Shyla Sharma

Sponsor: Jenifer Hooten

Team BMES204, Barranca Mesa Elementary School, *Code versus Cholera*

Team Member: Andrew Morgan

Mentor: Nathaniel Morgan

Team CHS33, Capitol High School, *The Spread of Red Imported Fire Ants*

Team Members: Alyssa Huerta, Oscar Lopez-Rivera Jr., Ryan Romero

Sponsors: Irina Cislaru, Barbara Teterycz

Team GADMS113, Gadsden Middle School, *Tuberculosis Outbreak*

Team Members: Carlos Mendoza, Christian Martinez, Denisse Delgado, Abel Romero, Victor Castillo

Sponsor: Danielle McFarland

Team JMS32, Jackson Middle School, *A-Maze-Ing*

Team Members: Brendan Kuncel, Quentin Dye

Sponsor: Karen Glennon, Mentor: Patty Meyer

Team JMS58, Jackson Middle School, *Stuxnet*

Team Members: Nancy Avila, Tiffany Chau, Laisbiel Garcia, Brandon Pham

Sponsors: Karen Glennon, Sharee Lunsford, Mentors: Jane Haagensen, Patty Meyer

Team JMS59, Jackson Middle School, *Glioblastoma*

Team Members: Israel Chirino, Santiago Herrera

Sponsor: Sharee Lunsford, Mentors: Nick Bennet, Karen Glennon

Team JMS60, Jackson Middle School, *How Small Towns are Affected by Dengue Fever?*

Team Members: Qwenevere Caouette, Kyreen White

Sponsors: Karen Glennon, Sharee Lunsford

Mentors: Nick Bennett, Steven B. Bradfute, Susan Gibbs, Patty Meyer

Team JMS62, Jackson Middle School, *Traffic Flow at the Albuquerque International Balloon Fiesta*

Team Members: Kiara Onomoto, Ayvree Urrea, Delaney Montoya

Sponsor: Karen Glennon, Mentor: Nick Bennett

Team LAHS201, Los Alamos High School, *Implementing an Artificial Neural Network to Recognize Handwritten Digits*

Team Member: Sam Crooks

Sponsor: Jacob Stinnett

Team LAHS56, Los Alamos High School, *Predicting Food Shortages in Africa from Satellite*

Team Member: Lillian Kay Petersen

Mentors: Daniela Moody, Rick Chartrand

Team LAHS71, Los Alamos High School, *RSA Based Primality Test*

Team Member: Elijah Pelofske

Sponsor: Alan Didier

Team LAHS72, Los Alamos High School, *Hydroponic Agriculture: Commercial vs. Individual Growth in New Mexico*

Team Members: Gabriel Holesinger, Duncan Fuehne, Uriah Sanchez, Xavier McTeigue

Sponsor: Allan Didier

Mentors: Donald Davis, Adam Drew

Team LAMS155, Los Alamos Middle School, *Interpreting and Classifying Music*

Team Members: Andrew Corliss, Maximilian Corliss, Phillip Ionkov, Ming Lo

Sponsor: Ellie Simons, Mentor: Li-Ta Lo

Team LLHS47, Los Lunas High School, *What's Missing?*

Team Members: Jen Marie Phifer, Forest Good

Sponsor: Ann Loveless

Mentors: Nick Bennett, Creighton Edington, Annette Hatch, James Taylor

Team LLHS53, Los Lunas High School, *Effects of Drag in Different Environments*

Team Members: Brandon (Everest) Sonnenberg, Chloe Sawatzki

Sponsor: Anne Loveless

Team MCS158, McCurdy Charter School, *Healing the Ozone Hole*

Team Members: Haven Hennelly, Jasmine Marquez, Ruby Salazar

Sponsors: Melissa Berryhill, Roger Duroux

Team MCS69, McCurdy Charter School, *Bug Infestation in Carson National Forest*

Team Members: Isaiah Marquez, Ramses Martinez, Christopher Maestas, Jesus "Chuy"

Gonzales, Michael Trujillo

Sponsors: Melissa Berryhill, Roger Duroux

Team MDS185, Monte Del Sol Charter School, *Spread of Dicamba Through Resistant and Not Resistant Crops*

Team Members: Robert Borrego, Lia Fukuda, Stephen Herrera

Sponsor: Rhonda Crespo, Mentors: Susan Gibbs, Paige Prescott

Team MELES41, Melrose Elementary School, *Crop dust? We must!!!*
Team Members: Lily Macfarlane, Heidi Macfarlane, Eoghen Knight
Sponsor: Alan Daugherty, Mentor: Seaver Tate

Team MELHS19, Melrose High School, *The 4U App*
Team Members: Gracie Sanchez Chantilly Fulgham
Sponsor: Alan Daugherty, Mentor: Dewayne Fulgham

Team MELHS3, Melrose High School, *Bioterrorism*
Team Members: Hannah Wofford, Mackenzie Perkins, Rebecca Rush, Shannon May
Sponsor: Alan Daugherty, Mentor: Randall Rush

Team MELHS34, Melrose High School, *Beyond the Shore*
Team Members: Madison Garrett, Evelyn Garrett
Sponsor: Alan Daugherty

Team MELHS35, Melrose High School, *Blood Ivory II*
Team Members: Dennis Roldan, Caitlyn Nicholson, Joshua Littell
Sponsor: Alan Daugherty

Team MELHS36, Melrose High School, *Volcanoes*
Team Members: Marilyn Lopez, Anjalina Sanchez
Sponsor: Alan Daugherty

Team MELHS42, Melrose High School, *PTSD in Babies Brains*
Team Member: Andrea Martin
Sponsor: Alan Daugherty

Team MELMS45, Melrose Middle School, *Sunsational Sundials*
Team Members: Brady Sorgen, Brandon Garcia, Brody McAlister, Xander Chavez
Sponsor: Alan Daugherty

Team MMS139, Mesa Middle School, *Solar Energy: Is It Worth It?*
Team Member: Jennifer Lopez
Sponsor: Tracie Mikesell

Team MMS184, Mesa Middle School, *Effect of Human Interaction on Wildlife*
Team Members: Jade Collins, Gabriel Agnew
Sponsor: Tracie Mikesell

Team MULTI61, nex+Gen Academy, Eldorado High School, *The Farmer in the Optimal Dell*
Team Members: Isabella Montoya, Reyanna Fromme, Savannah Phelps
Sponsor: Karen Glennon
Mentors: Neil Haagensen, Suzy Andrego

Team NFS26, New Futures High School, *Our Carbon Footprint*
Team Members: Serenity Pizarro, Ashlan Coe
Sponsors: Mary Rafferty, Joe Vertrees

Team NMSA134, New Mexico School for the Arts,
The Economic Effects of Implementing Environmental Full Cost Accounting
Team Member: Occam Kelly Graves
Sponsor: Jennifer Black

Team PESSF118, Pinon Elementary School,
Data Science for Social Media in Emergency Management
Team Members: Annisa Berendzen, Maylin Gonzalez, Logan Martinez, Adnan Mufleh, Amulya
Mulakala, Kairav Sharma
Sponsor: Delara Sharma, Mentor: Joel Berendzen

Team PHS73, Portales High School, *Diapetetics*
Team Members: Samuel Gore, Kerstiy Laman, Stephen Villanueva
Sponsor: Jack Willis

Team RRCA1, Rio Rancho Cyber Academy, *Ice Ice Power Baby*
Team Members: Dylan Martinez, Christopher Meyer, Gabe Turner
Sponsor: Harry Henderson

Team SFHS109, Santa Fe High School, *The Real Bee Movie*
Team Members: David Salazar-Reyes, Ricky Martin
Sponsor: Brian Smith

Team SFHS120, Santa Fe High School, *Feral Boars in New Mexico*
Team Members: Tenzin Lungtok, Mirabai Venkatesh, Justin Reynolds, Gabriela Armendariz
Sponsor: Brian Smith

Team SFHS122, Santa Fe High School, *Traffic Model*
Team Members: Elijah Sandoval, Omar Lux
Sponsor: Brian Smith

Team SFHS162, Santa Fe High School, *Efficient Evacuation of At-Risk Populations*
Team Members: Rowan Cahill, Ramona Park, Luke Shankin
Sponsor: Brian Smith, Mentor: Hope Cahill

Team SJCHS167, San Juan College High School, *Intelligence VS Cognition*
Team Members: Alex Hamilton, Brianna Morrow
Sponsors: Dena Burgert, Geizi Llanes/Dejka

Team SJCHS168, San Juan College High School, *O₂ as Time Flies*
Team Members: Brendon Johnson, Jacob Ledbetter
Sponsors: Dena Burgert, Geizi Llanes/Dejka

Team SJCHS169, San Juan College High School, *Radiation Effect on Foods*
Team Members: Aaliyah Nakai, Alishah Nakai
Sponsor: Dena Burgert

Team SJCHS170, San Juan College High School, *Dangers of Fire*
Team Members: Cameron Collyer, Hyrum Harris
Sponsors: Dena Burgert, Geizi Llanes/Dejka

Team SJCHS171, San Juan College High School, *The Accuracy of Facial Recognition*
Team Members: Samantha Smith, Robyn Curtis
Sponsors: Dena Burgert, Geizi Llanes/Dejka

Team SJCHS172, San Juan College High School, *Curing Breast Cancer Using Sloth Fur Fungi*
Team Members: Isaiah Joe, Jaron Yazzie
Sponsors: Dena Burgert, Don Lorett

Team SJCHS23, San Juan College High School, *On the Concept of Negative Mass*
Team Member: Johndenmyr Mendoza
Sponsor: Dena Burgert

Team SPS128, Sandia Preparatory School,
Solar System Formation: How Common is a Planet Like Earth?
Team Members: Cal Boye-Lynn, Noah Peterson, Parker Willis
Sponsor: Neal Holtschulte

Team SPS132, Sandia Preparatory School, *Re-creation of the RSA algorithm*
Team Members: Logan Pierson, Sarah Maurice, Houston Fett, Evan Custer
Sponsor: Neal Holtschulte

Team SSMA200, Saturday Science and Math Academy, *Animatrix*
Team Members: Sarai Rankin, Dominique James
Sponsor: Eric Rankin

Team SSMA48, Saturday Science and Math Academy, *Hurricane Effect on Disease Transfer*
Team Members: Naomi Rankin Matuke Fomukong
Sponsor: Debra Johns, Mentor: Eric Rankin

Team SVA144, South Valley Academy, *M.A.Z.K*
Team Members: Makayla Paiz, April Sandoval, Zaira Enriquez, Kimberly Kolander
Sponsor: Brian Dickinson

Team THS116, Taos High School, *Disaster Simulator*
Team Members: Jenay Barela, Kineo Memmer, Jeremiah Martinez, Zak Willis
Sponsor: Tracy Galligan

Team THS31, Taos High School, *Modeling Straw Mulch in Taos Agriculture*
Team Members: Rowan Kinney, Eben Bellas
Sponsor: Tracy Galligan, Mentor: Gregory Rael

Team THS131, Taos High School, *The Dangers of CTE*
Team Members: Cyrus O'Hern, James Ryan Cox, Sawyer Solfest, Skylar Spriggs
Sponsor: Tracy Galligan
Mentors: Andrew Leonard, Nick Bennett

Team THS159, Taos High School, *MitoSensor*
Team Members: Sierra Ferguson, Delaney Galligan
Sponsor: Tracy Galligan

Team THS203, Taos High School, *Tune the Fire*
Team Members: Joshua Fambro, Geneva Tafoya, Kyle Totman
Sponsor: Tracy Galligan

Team UBMS187, Manzano High School, *Path to Pathology*
Team Members: Seth Griffin, Danny Luong, Torrey Luong, Nigone Phommachack, Manoj Subedi, Jerrel White
Sponsors: Karen Glennon, Sharee Lunsford, Mentor: Kurtis Griess

Team YWiC163, Young Women in Computing, *Vaccines and Viruses*
Team Members: Fareeda Alhalawany, Isabel Shu, Larissa Zhou
Sponsor: Clara Graham, Mentor: Gabi Hernandez

Team YWiC165, Young Women in Computing, *How Can We Reduce the Rising Temperatures?*
Team Members: Ella Gamache, Tiffany Monarco
Sponsor: Gabi Hernandez

Judges

Ashar Ali, Air Force Research Laboratory
Ed Angel, Professor Emeritus UNM
Claudia Aprea, Northern New Mexico College
Caitlin Ard, New Mexico State University
Charlene Arias, Sandia National Laboratories
Zachary Barraza, New Mexico State University
Richard Barrett, Sandia National Laboratory
Nick Bennett, CNM STEMulus Center
Rusmir Bilalic, Air Force Research Laboratory
Patrick Bridges, University of New Mexico
Bianca Bologna, University of New Mexico
Zanoni Contreras, New Mexico State University
Shaun Cooper, New Mexico State University
Steve Cox, Northern New Mexico College
Ken Cradock, Eastern New Mexico University
Jesse Crawford, Sandia National Laboratories
Matthew Curry, Sandia National Laboratories
Geoff Danielson, Sandia National Laboratories
Mike Davis, Cray Inc
Sharon Deland, Sandia National Laboratories
Sean Dillon, Risk Sense
Juergen Eckert, University of South Florida
Drew Einhorn
Sandy Frost, Los Alamos National Laboratory
Chris Garasi, Sandia National Laboratories
Ayana Ghosh, Los Alamos National Laboratory
Susan Gibbs, Project GUTS
Kaley Goatcher, Honeywell Aerospace
Zachary Harding, Risk Sense
Michael Hawthorne, City of Clovis
Kevin Horst, AmeriCorps VISTA
David Janecky, Los Alamos National Laboratory
Elizabeth Yakes Jimenez, University of New Mexico
Adjei Jones, Eastern New Mexico University
Philip Jones, Los Alamos National Laboratory
Mia Kalish, New Mexico State University
Randall Keeswood, San Juan College
Cole Kendrick, New Mexico Institute of Mining and Technology
Larry Kilham
Amy Knowles, New Mexico Institute of Mining and Technology
Chris Koch, Google
Rebecca Koskela, DataONE, University of New Mexico
David Kratzer, Los Alamos National Laboratory

Peter Lamborn, Los Alamos National Laboratory
Maximo Lazo, Central New Mexico Community College
Dale Leschnitzer, Los Alamos National Laboratory
David Liebel, San Juan College
Laura Liebel, San Juan College
Chris Lindemann, Eastern New Mexico University
Lena Lopatina, Los Alamos National Laboratory
Joan Lucas, University of New Mexico-Los Alamos
John McDermon, Los Alamos National Laboratory Foundation
Karen McKinnon, Descartes Lab
Patty Meyer, Supercomputing Challenge
Vince Meyer
Nathaniel Morgan, Los Alamos National Laboratory
Monique Morin, Los Alamos National Laboratory
Heidi Morning, Los Alamos National Laboratory
Michaela Negus, Sandia National Laboratories
Paul Nelson, City of Clovis
Nimemi Niemand, University of New Mexico
James Overfelt, Sandia National Laboratories
Stephen Pate, New Mexico State University
Veena Parboteeah, Eastern New Mexico University
Paige Prescott, Code.org/Projects GUTS
David Price, New Mexico State University
Maureen Psaila-Dombrowski, Los Alamos National Laboratory
Jacquelyn Quesenberry
Eric Rankin
Lonnie Rednour, San Juan College
Eric Ried, Presbyterian Healthcare Services
Dana Roberson
Teri Roberts, Los Alamos National Laboratory/retired
Chrysm Watson Ross, University of New Mexico
Kim Selvage, Los Alamos National Laboratory
Albert Simon
Ray Stubbs, New Mexico State University
Daniel Swinford, Eastern New Mexico University
Lydia Tapia, University of New Mexico
Steve Taylor, OpenMake Software
Tim Thomas, University of New Mexico
Ann Thompson, NM Technology in Education
William Tolley, University of New Mexico
David Torres, Northern New Mexico College
Michael Trahan, Sandia National Laboratories
Laure Triplett, Los Alamos National Laboratory
Barbara Trujillo, Los Alamos National Laboratory
Eleanor Walther, Sandia National Labs/retired

Talaya White
Thomas (Mark) Willmon, Eastern New Mexico University
Simon Woodruff, Woodruff Scientific, Inc.
Janelle Zamie, Eastern New Mexico University
Christopher Zapotocky, University of New Mexico



Do you want to become a supporter of the Supercomputing Challenge?
Please email us at consult@supercomputingchallenge.org for details.

Project Title:
“What’s Missing?”

New Mexico
Supercomputing Challenge
Final Report
April 4, 2018

Team Members:
Jen Marie Phifer
Forest Good

Teacher:
Anne Loveless

Project Mentors:
Creighton Edington
Annette Hatch
Nick Bennett
James Taylor

Acknowledgements:

We would like to thank the following people for assisting us in our project:

Nick Bennett: For mentoring us and assisting us with the “Set-finding code” of our project and for encouraging us to delve deeper into some of the questions this project posed

Creighton Edington: For mentoring us and always assisting us with any questions we had and for guiding our project to higher levels

Annett Hatch: For mentoring us, helping us think about the different ways to approach this problem, and for always welcoming any questions we had

Anne Loveless: For sponsoring us

James Taylor: For helping us review the basic conceptual theories of this problem

Table of Contents

Introduction **4**

Materials and Methods **4**

Regular Game Play **5**

The Mathematics of a SET **6**

Code **8**

One Missing Card **11**

Code **12**

Two Missing Cards **14**

Problem Analysis **14**

Problem Exploration and Solution **15**

Code **19**

Results **22**

Discussion **22**

Optimization **23**

Code **25**

Trial, Results, and Discussion **26**

Applications and Final Conclusions **27**

Bibliography **28**

Introduction

Information is powerful; it is a retention of knowledge, an instructional guide, and a key to learning. Therefore, databases that hold information are crucial. However, what happens when pieces of data go missing? How would these pieces be recovered? Typically, if one missing piece of information is lost, it's not too hard to fill in. We would simply go through all the records, and based on what we have, we can fill in the missing "blanks." However, what happens when there are two missing pieces of information? At this point, the problem is that the missing pieces of information can be a combination of different things.

The question is, how do we determine what the missing pieces are? How do we use what we know to find what we don't know? To explore this problem, we looked at the card game SET. Specifically, we looked at situations when one or two cards went missing from the deck and how we could determine what those cards were without having to go through the whole deck.

Materials and Methods:

To explore this problem, we created a computer simulation using the software NetLogo.

The three general steps involved in solving this problem were:

1. Finding a way to create a program that plays the game
2. Finding a way to find what one missing card is
3. Finding a way to find what two missing cards are

We conducted our experiment by using NetLogo's built in "BehaviorSpace" function for our trials.

Regular Game Play

There are 81 cards, each with 4 attributes (number, color, shading, and shape) that each have 3 variations (3 different numbers, 3 different colors, 3 different shadings, and 3 different shapes).

The variation for each attribute is as follows:

| Attributes | Number | Color | Shading | Shape |
|--------------------|--------|--------|---------|----------|
| Variation 1 | 1 | Green | Open | Diamond |
| Variation 2 | 2 | Purple | Hashed | Oval |
| Variation 3 | 3 | Red | Solid | Squiggle |

The game starts by placing 12 cards face up on the table. From there, players try to find a “SET,” which is comprised of 3 cards in which the attributes have either all the same variation or all different variations. For example, take the layout below:

This is a SET

Let's check why:
(remember that the variation of each attribute has to be either all the same or all different)

- ✓ Number – All Different (1,2,3)
- ✓ Color – All the Same (red)
- ✓ Shading – All the Same (open)
- ✓ Shape – All the Same (Ovals)

This is a NOT a SET

Let's check why:
(remember that the variation of each attribute has to be either all the same or all different)

- ✓ Number – All Different (1,2,3)
- ✓ Color – All Different (purple,green,red)
- X Shading – Two solids but one hashed
- X Shape – Two ovals but 1 diamond

Figure 1: The SET graphics are taken from (“Harris Spiral”). The arrows, circles and explanations are the competitors’

Essentially, there are 4 different variation combinations for each attribute. Using “D” as an abbreviation for “Different” and “S” as an abbreviation for “Same,” the 4 combinations are:

- 4D
- 3D, 1S
- 2D, 2S
- 1D, 3S

**Note: The SET shown in Figure 1 was combination 1D, 3S. All the variations for one attribute (number) was different while the rest were all the same.*

*Hint/side note for newer players: The most common way most people find SETs is by pairing 2 cards together and looking for the third card that will complete the SET. This is known as the Fundamental Theorem of SET: “Given any pair of cards, there is a unique card that completes a SET with the pair” (Gordon, 13).

Once a player spots a SET, the player yells “SET” and proceeds to collect his or her cards. From there, the dealer deals 3 more cards. If there are no sets, the dealer deals out groups of 3 cards until there is a SET, going only up to a maximum of 21 cards on the table — research guarantees a SET at 20 cards (Klarreich). The players go through the whole deck and whoever has the most sets wins.

The Mathematics of a SET

A SET’s validity (whether it is a SET or not) can not only be visually checked, but also mathematically checked. For example, say the following cards were grouped as a SET:

- Card 1: 1 purple solid diamond

- Card 2: 3 red solid diamonds
- Card 3: 2 green hashed diamonds

Typically, a visual checklist is conducted for all the attributes: *Are all the number variations either all the same or all different? Are all the color variations either all the same or all different? Etc.* In this case, the three cards are not a SET because the shading is not all the same or all different. Visually, this would be easy enough to see.

However, a mathematical check can also be conducted. To do this, each attribute variation must be assigned a value of 0, 1, or 2. For this project, the following assignment variation was used:

| Attributes | Number | Color | Shading | Shape |
|----------------|--------|--------|---------|----------|
| Value 0 | 3 | Green | Open | Diamond |
| Value 1 | 1 | Purple | Hashed | Oval |
| Value 2 | 2 | Red | Solid | Squiggle |

**Note: It does not matter how the values are assigned. For example, whether red is 2 instead of 0 is of no consequence.*

After an assignment is made, the sum of each attribute's variation must be totaled. If the modulo 3 (the remainder after dividing by 3) of each attribute's sum equals 0, then it is a valid SET. Looking at the previous card example (which was *not* a SET), if each card's attribute variation is added, the following result is obtained:

- Number sum: 3
- Color sum: 3
- Shading sum: 5

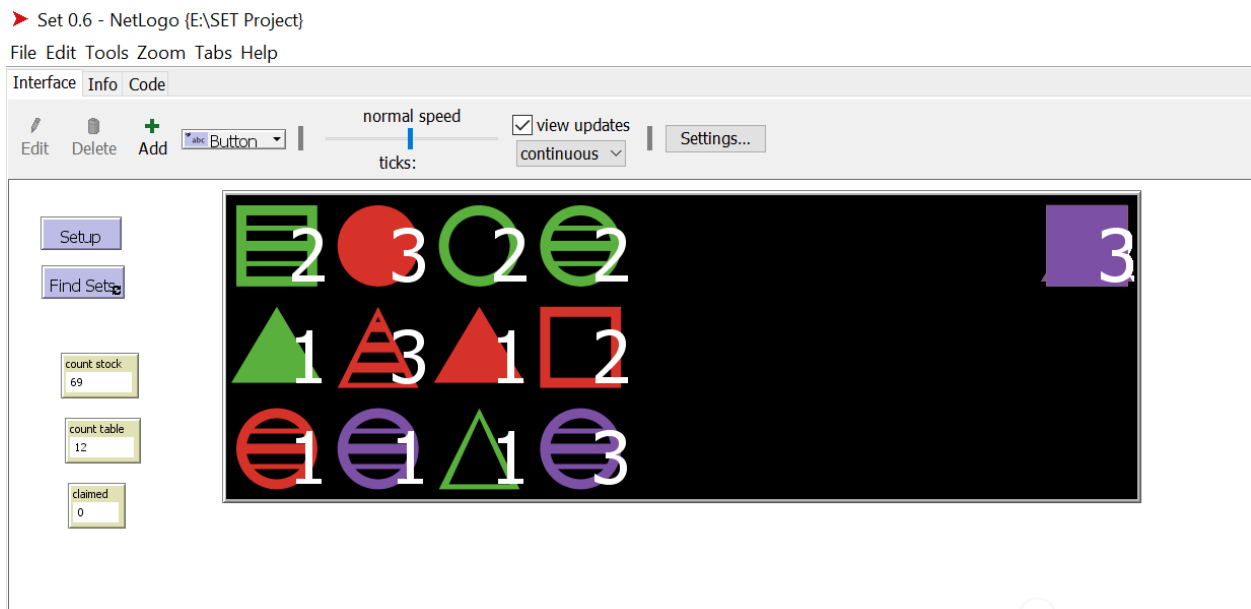
- Shape sum: 0

The modulo 3 of each sum with respect to the order above is: 0, 0, 2, and 0. Because the shading's modulo 3 is not equal to zero, it is not a valid SET. However, if Card 3 was replaced with 2 green *solid* diamonds, it would become a valid SET because the shading sum would change to 6, making the modulo 3 equal 0.

*Note: Every valid SET follows this mathematical rule.

Code

Using the mathematics of a SET, we created a program that played the game. The following is a screenshot of the first working version we created:



Graphically, it is slightly different than the original game. Instead of the shapes being oval, diamond, or squiggle, it is now circle, triangle, and square respectively. Furthermore, instead of having multiple shapes (1, 2, or 3) there is a label signifying the quantity.

The following is a screenshot of the set-finding code:

```
10
11 to-report is-a-set? [:candidates]           ;; looks at all three attribute
12   report (                                 ;; if yes, it is a set
13     ((sum [coloring] of :candidates) mod 3 = 0)
14     and ((sum [shading] of :candidates) mod 3 = 0)
15     and ((sum [form] of :candidates) mod 3 = 0)
16     and ((sum [number] of :candidates) mod 3 = 0))
17   end
18
19 to-report all-sets [:pool]
20   let :sets-found []                       ;; Creat
21   let :pool-list (sort :pool)
22   let :card1-list (sublist :pool-list 0 (length :pool-list - 2)) ;; Proce
23   let :card1-indexes (n-values (length :card1-list) [i -> i])   ;; as a
24   (foreach :card1-list :card1-indexes [                          ;; value
25     [:card1 :index1] ->                                         ;; a "hc
26     let :card2-list (sublist :pool-list (1 + :index1) (length :pool-list - 1)) ;; all 3
27     let :card2-indexes (n-values (length :card2-list) [i -> i + 1 + :index1])
28     (foreach :card2-list :card2-indexes [                        ;; For e
29       [:card2 :index2] ->                                       ;; Cards
30       let :card3-list (sublist :pool-list (1 + :index2) (length :pool-list))   ;; the n
31       let :card3-indexes (n-values (length :card3-list) [i -> i + 1 + :index2]) ;; formi
32       (foreach :card3-list :card3-indexes [                       ;; the s
33         [:card3 :index3] ->
34         let :candidates (turtle-set :card1 :card2 :card3)
35         if is-a-set? :candidates [
36           set :sets-found (lput :candidates :sets-found)
37         ]
38       ])
39     ])
40   ])
41   report :sets-found
42   end
43
```

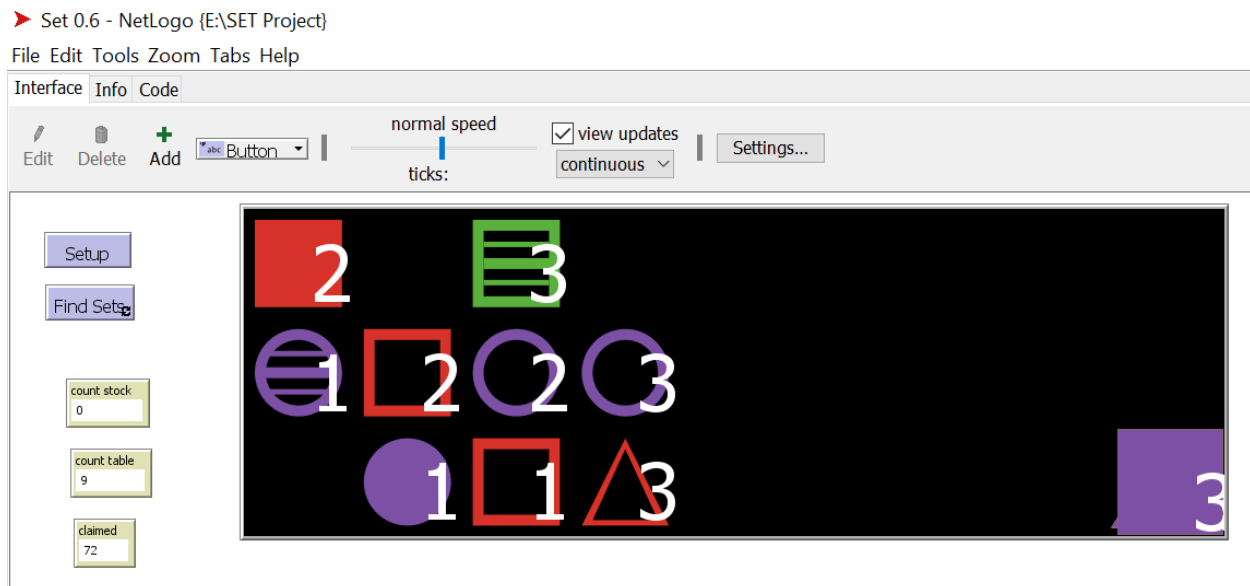
**Quick background* — In this code, there were 3 distinct variables we used to classify the cards: table, stock, and claimed. “Table” refers to the cards that have been dealt, “stock” refers to the deck, and “claimed” refers to the cards that have been grouped into SETs.

To find SETs, we used anonymous procedures. The 12 cards on the table are first placed into all the possible 3-card combinations. The variation of each attribute is then summed, and if the modulo 3 of each of the sums is 0, then it is considered a SET and placed into the list “sets-found.”

Once all the possible SETs are found, the program removes the first SET in the list and places them in the claimed category. More cards are then dealt to replace the cards claimed and

the process repeats, looping until there are no more possible SETs and no more cards left in the stock. If there is not a SET when 12 cards are on the table and there are still cards in the stock, it will deal more cards until there is a SET, going only up to 21 cards since there is guaranteed a SET at 20 cards (Klarreich).

The following is a screenshot of the program after it has been played:



**Note: When the game is played, there are typically cards left over that did not get grouped into a SET and cannot be grouped into a SET with one another; rarely are there no leftover cards.*

One Missing Card

Aside from the regular game play, there is a version called “End Game SET,” where the only difference is that a random card is taken out at the beginning of the game. After the card is taken out, the game is played normally. At the end, using the cards left over, an algorithm can be used to find the missing card.

To do this, a modified process of checking a SET’s validity is used. The value of each attribute’s variation must still be assigned, and the sum of each card’s attributes must also still be added. However, because we want to find the missing card, we must find the value (when added to the current sum) that causes the modulo 3 of each attribute’s total sum to equal 0, meaning that the total sum must equal a multiple of three to have a modulo 3 equal to zero. Once we know the values, we can use the assignment variation to find the corresponding attributes of the missing card.

**Note: The algorithm and method for finding one missing card has already been previously discovered.*

An example:

Say the following 5 cards are left:

- 3 red open squiggles
- 1 green solid oval
- 3 green solid diamonds
- 3 green open ovals
- 3 red solid ovals

The sum of all the cards' attribute is:

Number: $0 + 1 + 0 + 0 + 0 = 1$

Color: $2 + 0 + 0 + 0 + 2 = 4$

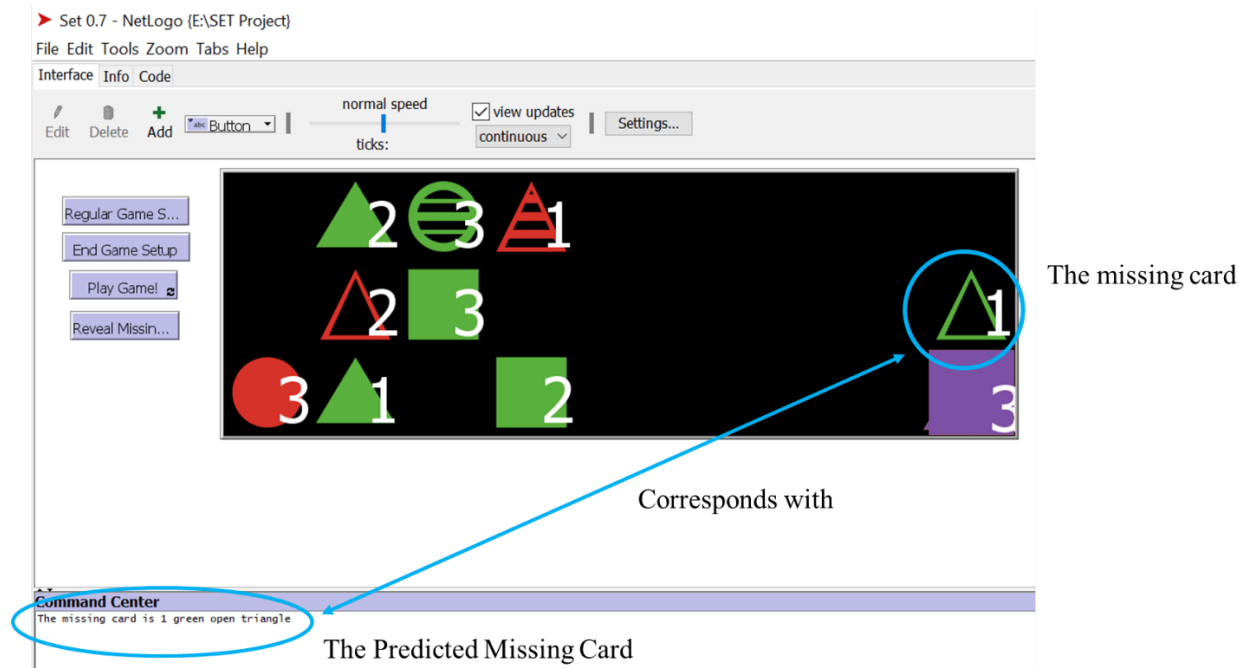
Shading: $0 + 2 + 2 + 0 + 2 = 6$

Shape: $2 + 0 + 1 + 0 + 0 = 5$

In order of the above attributes, the value that will cause the modulo 3 of each of the above sums to equal zero are: 2, 2, 0, 1. Looking at the assignment variation, the missing card is 2 red open ovals.

Code

We added the algorithm for finding a missing card into our program. The following screenshot is of the updated version:




```

131 to missing-card                                ;; Finds the
132   let needed_number 0                          ;; Creates :
133   let needed_coloring 0
134   let needed_shading 0
135   let needed_form 0
136   let total 0
137
138   set total (sum [number] of table)             ;; Sets the
139   let y 0
140   foreach [1 2 3]                               ;; Finds the
141   [                                               ;; mod 3 of
142     ifelse ((y + total) mod 3 = 0)
143     [
144       set needed_number y
145     ]
146     [
147       set y (y + 1)
148     ]
149   ]
150

```

A random card is taken out at the beginning of the game. The game is then played normally. Using the cards left over, the variation value of each card's attribute is summed. In the screenshot above, the program tests all the values (0, 1, 2) to see what value, when added to the total value for number, gives back a modulo 3 equal to 0. This same algorithm is used to find the color, shading, and shape of the missing card.

Two Missing Cards

**Note: As mentioned before, the method for finding one missing card has already been discovered. However, a method for finding two missing cards has not been previously discovered.*

Problem Analysis

Finding one missing card is fairly easy. However, the problem is when we try to look at two missing cards. The possible combination choices are too ambiguous; there are too many ways to make the sum of the variations equal a modulo 3 of zero. For example, in a hypothetical situation, say two cards at the beginning of the game are taken out. After playing, one card is left, meaning there are two possible combinations that will return a modulo 3 equal to zero. In this case, assume the leftover card is 1 green hashed oval. The variation assignment would be (with the same order of number, color, shading, and shape): 0, 1, 1, 0.

For each attribute variation there can be two possibilities that give a modulo 3 equal to zero:

Number: 0 and 0 / 1 and 2

Color: 0 and 2 / 1 and 1

Shading: 1 and 1 / 0 and 2

Shape: 0 and 0 / 1 and 2

It would be difficult to narrow down the choices, and even it was narrowed down (say it's 0 and 0 instead of 1 and 2 for number), it is still unclear which card has which variations. For example, say the first possibility for each attribute is right. It is unclear if Missing Card 1 is (0, 0, 1, 0) or if it's (0, 2, 1, 0).

Problem Exploration and Solution

To tackle this problem, we tried simplifying it and broke it down to a group of 9 cards with the same shape and color. If laid out correctly, a hyperplane that shows the 12 different sets that can be made with the 9 cards is created (see Figure 2 below).

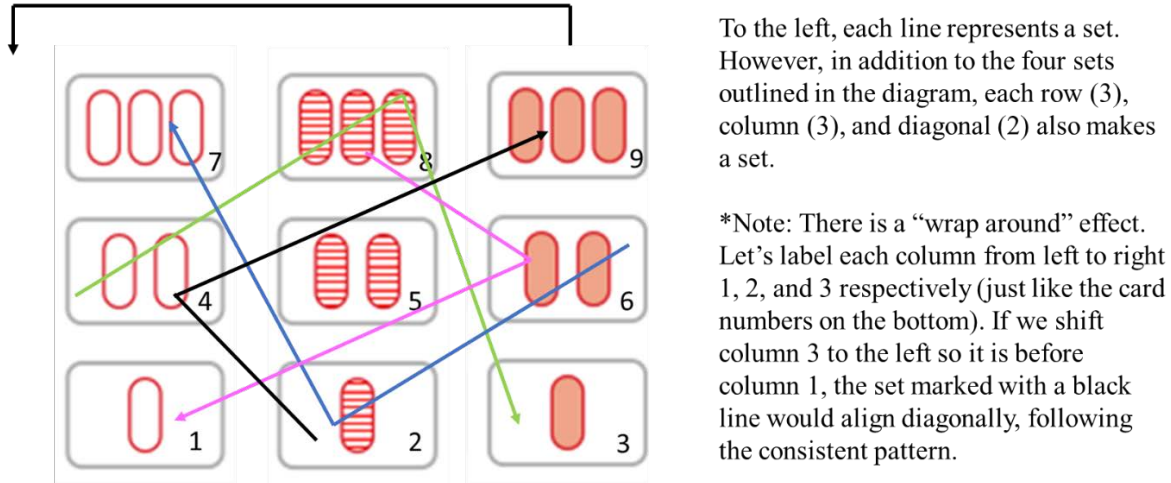


Figure 2: The SET graphics are taken from (“Game of Set”). The arrows and explanations are the competitors’

A list of the 12 possible SETs:

- | | |
|---------|---------|
| [1 2 3] | [1 5 9] |
| [4 5 6] | [3 5 9] |
| [7 8 9] | [2 6 7] |
| [1 4 7] | [3 4 8] |
| [2 5 8] | [1 6 8] |
| [3 6 9] | [2 4 9] |

Using the cards shown in Figure 2, if we only had 9 cards and 2 are missing, we would figure out what the missing cards are by grouping the cards that could form a SET together, as if playing a regular game. Then, we would look at all the possible set combinations and start drawing cards from the claimed pile. Using the cards that were drawn, we would begin

eliminating some of the 12 set possibilities. In our case (see Figure 3 below), we decided to make cards 2 and 5 the missing cards. In the best-case scenario, cards [1, 4, 7] would make a set and [9, 6, 3] would also make a set, leaving card 8 behind.

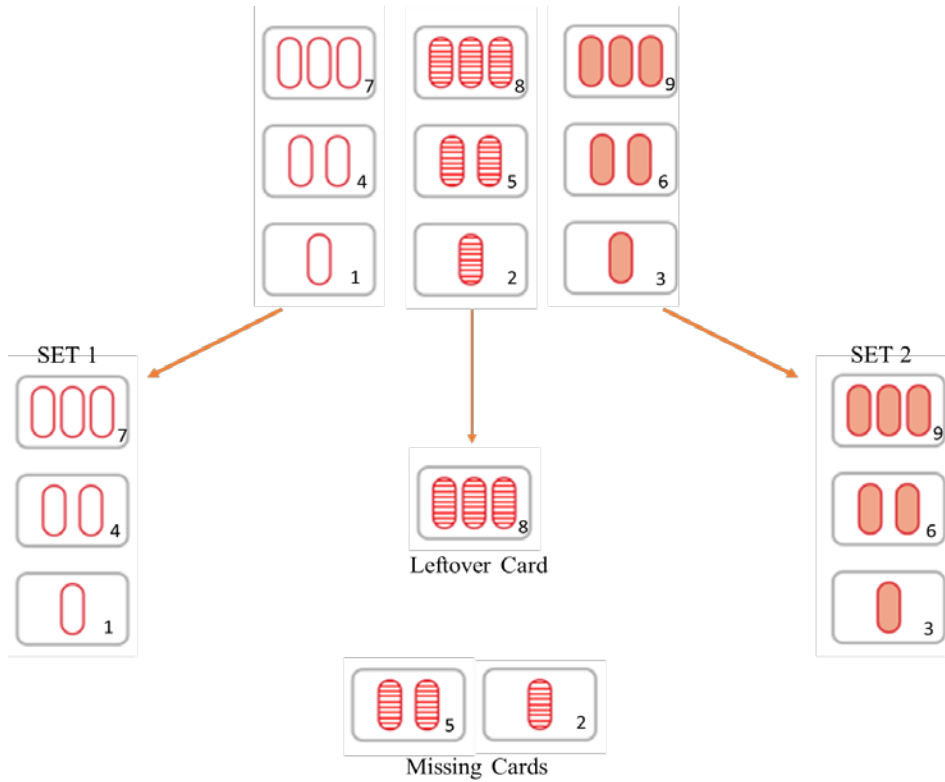


Figure 3: The SET graphics are taken from (“Game of Set”). The arrows and explanations are the competitors’

Now, when we look at the 12 possible sets that can be made, we must look for sets containing an 8 because the missing cards must form a SET with card 8 (we know this based on our hyperplane). Those sets are:

- The 4 possible SETs:
- | | |
|---------|---------|
| [7 8 9] | [2 5 8] |
| [3 4 8] | [1 6 8] |

We then go through the claimed cards to start eliminating the answer choices. For example, if card 7 was the first card that was revealed, then we can eliminate [7, 8, 9] as an answer choice. If card 3 was revealed next then we can eliminate [3, 4, 8]. If after that card 6 was revealed, then we can eliminate [1, 6, 8] and be left with [2, 5, 8]. That means the missing cards would have to be 2 and 5. However, this again is the best-case scenario in that no card in a set already eliminated appeared. For example, if card 4 was revealed after card 3 was revealed, it would not help because we had already eliminated set [3, 4, 8]. In other words, this all depends on how lucky we are when we draw the cards.

Overall, while this method did provide us with some valuable insight, it was difficult to apply it to a real game. To begin with, only having 9 cards that form a hyperplane is too convenient. In a real game, there are 81 cards — which makes 1080 total possible sets — and when we play the game, we are most likely not going to be left with only one card. Therefore, we would have great difficulty trying to narrow down which of the 1080 set possibilities holds the two missing cards. Furthermore, there is no guarantee that one of the cards left over will form a set with the two missing cards (like card 8 did in the previous example); the card that would complete the two missing cards' SET could have been grouped into another SET. In the previous example, we could have easily taken out SET [7, 8, 9] and had cards 1, 3, 4, and 6 leftover instead.

As a result, we were simply left to ponder for a while. However, we managed to realize something. Using the Fundamental Theorem of Set, we knew that for any pair of cards, there is only one card that makes the pair a SET. Therefore, we know that for the two missing cards, there is only one unique card that forms its SET. As we were looking at the mathematics of the

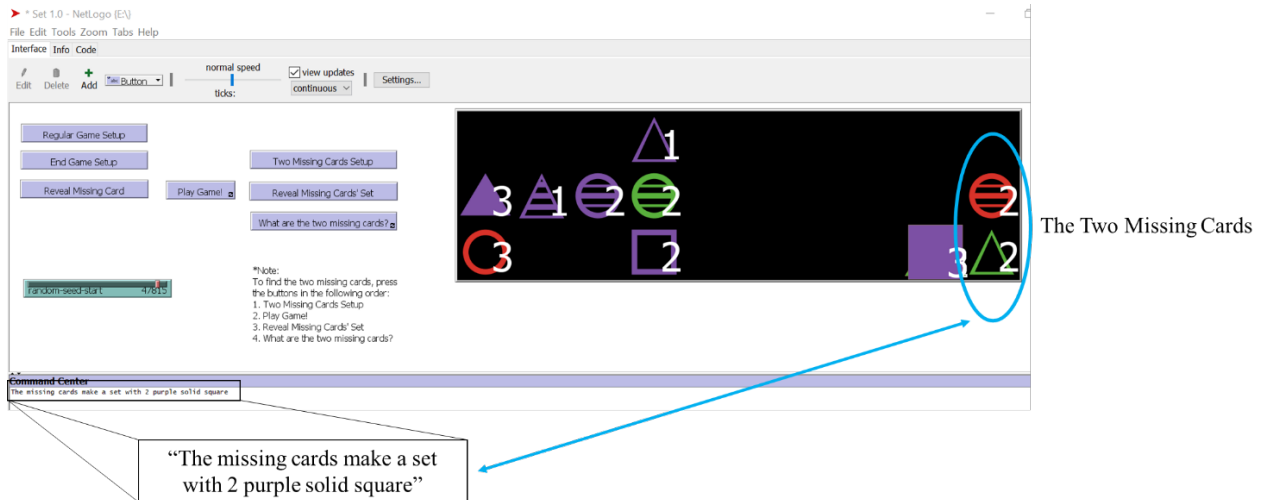
game, we saw that we could determine what this unique card was simply by “reversing” the formula used to find one missing card.

To determine the attribute of one missing card, we essentially look for the higher multiple of 3. However, in the case of two missing cards, to find the unique third card, we must look at the lower multiple of 3 instead. For example, for one missing card, if the sum of the color attribute was 10, then we would say the missing card’s color is red because the next multiple of 3 is 12, and we must add 2 to 10 to get there; 2 corresponds to red. However, in the case of two missing cards, we would say that the unique third card’s color is purple because we look at the lower multiple of 3, which is 9. We must subtract 1 from 10 to get there, and 1 corresponds to purple.

The significance of this find is that if we know what that third card is, then instead of having to look through all 1080 set possibilities, we would only have to go through 40 set possibilities because each card can make 40 different sets (Gordon, 31). Furthermore, after narrowing it down to the 40 sets, we can then use the same method of drawing cards from the claimed pile to start narrowing down the 40 possible sets.

Code

The following is a screenshot of the updated version, version 1.0, we created that finds the 2 missing cards:



Before the two missing cards are determined, the unique third card is determined first. As mentioned previously, this is done by “reversing” the algorithm originally used to find one missing card.

Original Algorithm

```
131 to missing-card ;; Finds the
132 let needed_number 0 ;; Creates
133 let needed_coloring 0
134 let needed_shading 0
135 let needed_form 0
136 let total 0
137
138 set total (sum [number] of table) ;; Sets the
139 let y 0
140 foreach [1 2 3] ;; Finds the
141 [ ;; mod 3 of
142 [ ifelse ((y + total) mod 3 = 0)
143 [ set needed_number y
144 ]
145 [
146 [ set y (y + 1)
147 ]
148 ]
149 ]
150
```

“Reversed” Algorithm

```
364 to missing-cards ;; Finds
365 let needed_number2 0 ;; Creat
366 let needed_coloring2 0
367 let needed_shading2 0
368 let needed_form2 0
369 let total 0
370
371 set total (sum [number] of table) ;; Sets 1
372 let y 0
373 foreach [1 2 3] ;; Finds
374 [ ;; mod 3
375 [ ifelse ((total - y) mod 3 = 0)
376 [
377 [ set needed_number2 y
378 ]
379 [
380 [ set y (y + 1)
381 ]
382 ]
383 ]
```

Then, we created an algorithm that generated all the set possibilities for that one card (done by modifying the original set finding code as show below).

| Original Set-finding Code | Modified Set-finding Code |
|---|---|
| <pre> 110 to-report is-a-set? [:candidates] 111 ; looks at all three attributes 112 report (113 ((sum [coloring] of :candidates) mod 3 = 0) 114 and ((sum [shading] of :candidates) mod 3 = 0) 115 and ((sum [form] of :candidates) mod 3 = 0) 116) 117 end 118 119 to-report all-sets [:pool] 120 let :sets-found [] 121 let :pool-list (sort :pool) 122 let :card1-list (sublist :pool-list 0 (length :pool-list - 2)) 123 let :card1-indexes (n-values (length :card1-list) [i -> i]) 124 (foreach :card1-list :card1-indexes [125 [:card1 :index1] -> 126 let :card2-list (sublist :pool-list (1 + :index1) (length :pool-list - 1)) 127 let :card2-indexes (n-values (length :card2-list) [i -> i + 1 + :index1]) 128 (foreach :card2-list :card2-indexes [129 [:card2 :index2] -> 130 let :card3-list (sublist :pool-list (1 + :index2) (length :pool-list)) 131 let :card3-indexes (n-values (length :card3-list) [i -> i + 1 + :index2]) 132 (foreach :card3-list :card3-indexes [133 [:card3 :index3] -> 134 let :candidates (turtle-set :card1 :card2 :card3) 135 if is-a-set? :candidates 136 set :sets-found (lput sort :candidates :sets-found) 137]) 138]) 139]) 140 }) 141 report :sets-found 142 end </pre> | <pre> 129 to-report all-sets2 [:pool] 130 let :sets-found [] 131 let :pool-list (sort :pool) 132 let :card1-list (sublist :pool-list 0 (length :pool-list - 2)) 133 let :card1-indexes (n-values (length :card1-list) [i -> i]) 134 (foreach :card1-list :card1-indexes [135 [:card1 :index1] -> 136 let :card2-list (sublist :pool-list (1 + :index1) (length :pool-list - 1)) 137 let :card2-indexes (n-values (length :card2-list) [i -> i + 1 + :index1]) 138 (foreach :card2-list :card2-indexes [139 [:card2 :index2] -> 140 let :card3-list (sublist :pool-list (1 + :index2) (length :pool-list)) 141 let :card3-indexes (n-values (length :card3-list) [i -> i + 1 + :index2]) 142 (foreach :card3-list :card3-indexes [143 [:card3 :index3] -> 144 let :candidates (turtle-set :card1 :card2 :card3) 145 if is-a-set? :candidates and member? third-card [who] of :candidates 146 ifelse member? revealed-card [who] of :candidates 147 [148 set :sets-found (fput sort [who] of :candidates :sets-found) 149] 150 [151 set :sets-found (lput sort [who] of :candidates :sets-found) 152] 153]) 154]) 155 }) 156 report :sets-found 157 end </pre> |

From there, the program starts eliminating the sets the leftover cards make. Once all the leftover cards' SET have been checked for, the program then draws cards from the claimed pile to start eliminating possibilities.

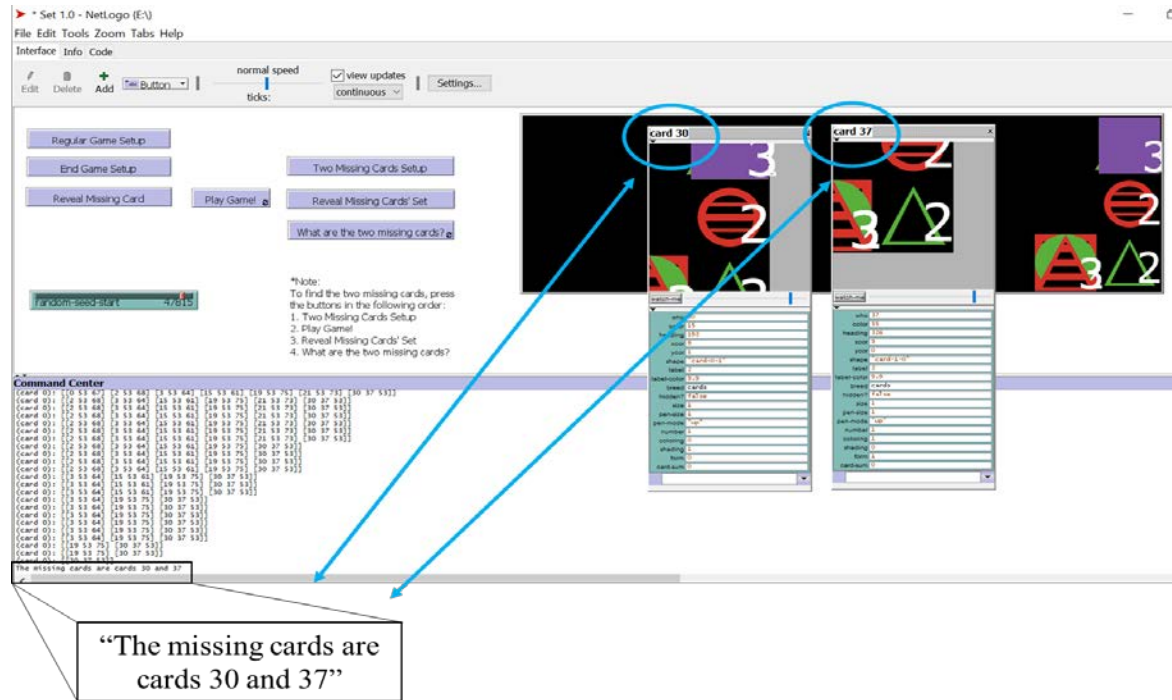
```

57 ask one-of claimed
58 [
59     set revealed-card [who] of self
60     setxy 9 2
61 ]
62
63
64 ask turtle 0
65 [
66     set stock (cards with [xcor <= 9])
67     set all-1080-sets all-sets2 stock
68     if repetition = 0
69     [
70         set all-1080-sets2 all-1080-sets
71     ]
72     set repetition repetition + 1
73     let check []
74     set check sublist all-1080-sets 0 1
75     let check2 []
76     set check2 item 0 check
77     set all-1080-sets2 remove check2 all-1080-sets2
78     show all-1080-sets2
79 ]
80
81 if length all-1080-sets2 = 1
82 [
83     let check3 []
84     set check3 item 0 all-1080-sets2
85     set check3 remove third-card check3
86     type "The missing cards are cards " type item 0 check3 type " "
87     ask cards with [xcor = 9 and ycor >= 0]
88 [
89     show-turtle

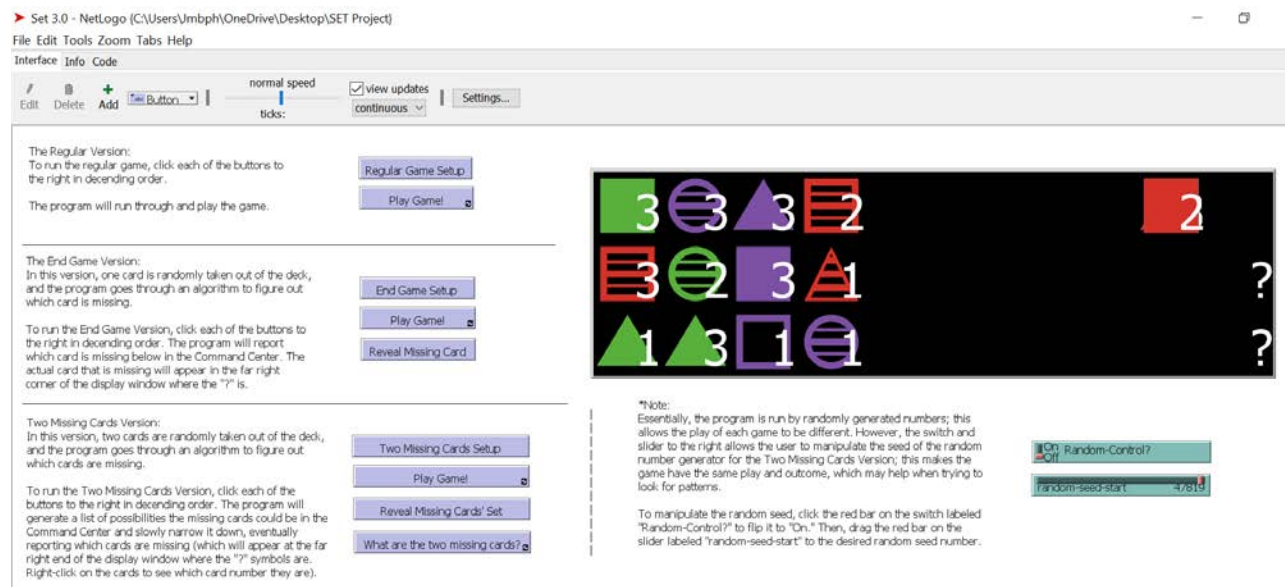
```

The code draws a random card from the claimed pile and sets it as the “revealed card.” One of the turtles then checks to see if the revealed card is in one of the 40 possible SETs list, labeled “all-1080-sets2”. If it is, the turtle removes the SET from the list. Once the list has been narrowed down to one SET, the program types out which cards are the missing cards.

Below is a screenshot of the interface at the end of the program's run:



Below is a screenshot of the "final version" (3.0) of our program. It's been cleaned up and organized to make it more user-friendly:



Results

The program was able to accurately determine what the two missing cards were. However, we wanted to determine how many claimed cards the program had to go through before it was able to determine what the two missing cards were. To do this, we ran the program 100 times through NetLogo's built-in BehaviorSpace feature. This allowed us to record the average number of claimed cards left after each trial. On average, the number of claimed cards left was approximately 11. The number of cards on the table that did not make a SET was approximately 8. Therefore, it had to go through approximately 85% of the claimed cards before it was able to figure out which cards the missing cards were.

Discussion

By finding out which card the two missing cards group with to make a SET, we could determine the possible set combinations the two missing cards have to be in. From there, using the claimed cards, we could slowly start narrowing down the possibilities. Through this method, we only had to go through approximately 85% of the claimed cards. Questions about how helpful this is can vary.

Depending on how large the claimed cards (the equivalent of records/what we know in the real world) are, it can be fairly helpful or not helpful at all. For example, if our claimed cards stack was 10000 cards, then saving 15% would help quite a bit. We would only have to go through 8,500 cards, saving ourselves from looking through 1,500 of them. However, if our claimed cards stack was 10, then saving 15% would not be helpful, saving ourselves from having to go through about 2 cards.

Optimization

We were dissatisfied that there were 11 cards left in the claimed pile; it seemed like too low of a number. Therefore, we tried looking at alternate ways to narrow down which cards could be the missing cards.

We had two ideas in mind when we were looking at how we could optimize this process so we would have to go through less cards. One idea was to examine a cube so we could determine if the connections a SET can make can eliminate other SETs. A cube is three hyperplanes stacked on top of one another. When that happens, SETs can not only be made on one plane, but across planes as well. Figure 4 demonstrates a few different SETs the middle card on Plane 1 can make across all three planes.

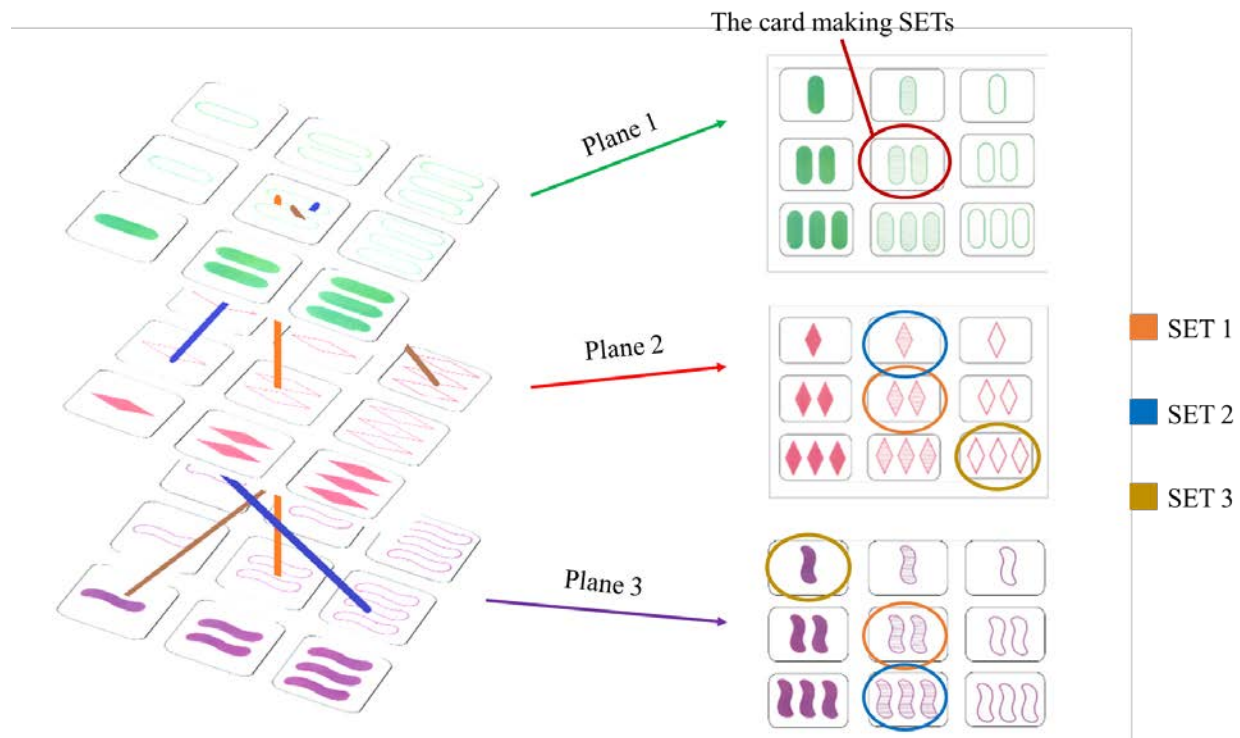


Figure 4: The SET graphics are cards that have been scanned from the original card game. The arrows, circles, and explanations are the competitors'

Figure 4 only outlines a few SETs one card can make. However, if all the SETs were “connected” in the figure above, there would be many lines. If one of those lines were to intersect with another line, could we eliminate the other SET connected to that line? That was one idea we had. However, as we were trying to test it, we came across incidents where we would have eliminated a SET that contained one of the missing cards. Therefore, we had to move on to our second idea.

The second idea was to look at probabilities. Instead of having a 100% accuracy rate where we can confirm which two cards are missing, we thought we could have a 70 – 90% accuracy rate instead. This way, we could say something similar to “Cards 1, 4, 7, 8” have an 80% chance of being the missing cards. This would likely increase the number of cards we wouldn’t have to go through.

Doing some more research, we found the following table:

TABLE 2.1.
Final counts (including percentage of total deck).

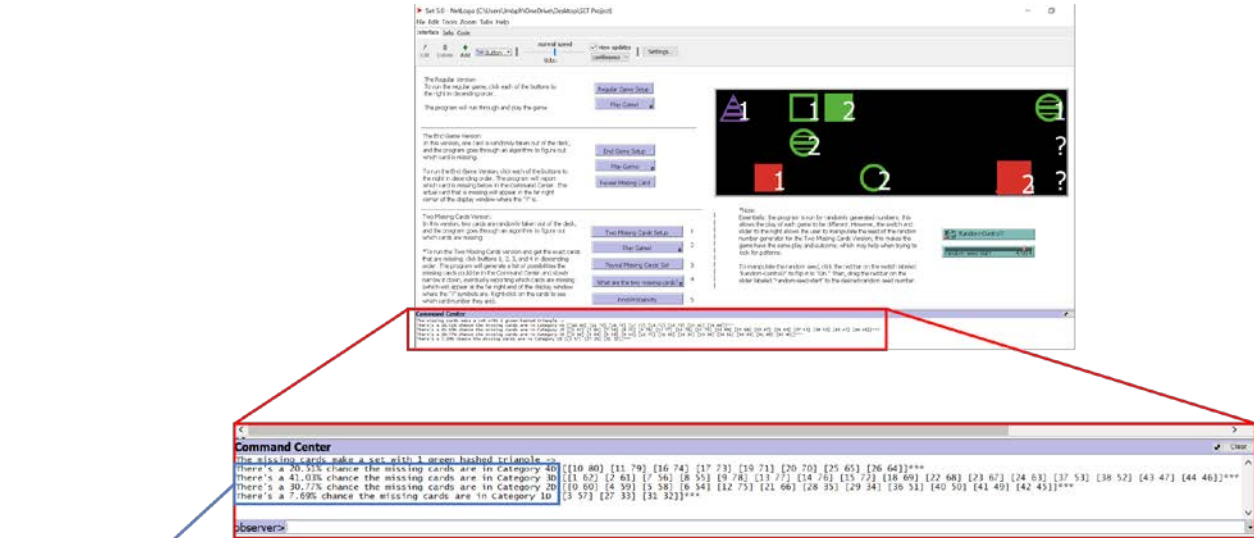
| <i>Kind of SET</i> | <i>Number</i> | <i>Fraction</i> | <i>Percentage</i> |
|--------------------|---------------|---------------------|-------------------|
| All different | 216 | $\frac{216}{1080}$ | 20% |
| Three different | 432 | $\frac{432}{1080}$ | 40% |
| Two different | 324 | $\frac{324}{1080}$ | 30% |
| One different | 108 | $\frac{108}{1080}$ | 10% |
| Total | 1080 | $\frac{1080}{1080}$ | 100% |

Figure 5: Table taken from page 38 of The Joy of SET book listed in the bibliography

With reference to the explanation on page 6 of this report explaining the types of SETs that can be made (such as “4D,” “3D,” etc.), Figure 5 shows the probability for each type of SET, which we incorporated into our model.

Code

The following are screenshots of the optimized version of the program and part of its code:



There's a 20.51% chance the missing cards are in Category 4D
 There's a 41.03% chance the missing cards are in Category 3D
 There's a 30.77% chance the missing cards are in Category 2D
 There's a 7.69% chance the missing cards are in Category 1D

```

800 set prob-length ((length D4-pt) + (length D3-pt) + (length D2-pt) + (length D1-pt))
801 type "There's a " type (precision (((length D4-pt) / prob-length) * 100) 2) type "% chance the missing cards are in Category 4D " type D4-pt
802 print "****"
803 type "There's a " type (precision (((length D3-pt) / prob-length) * 100) 2) type "% chance the missing cards are in Category 3D " type D3-pt
804 print "****"
805 type "There's a " type (precision (((length D2-pt) / prob-length) * 100) 2) type "% chance the missing cards are in Category 2D " type D2-pt
806 print "****"
807 type "There's a " type (precision (((length D1-pt) / prob-length) * 100) 2) type "% chance the missing cards are in Category 1D " type D1-pt
808 print "****"
809

```

The interface shows the probability of the missing cards being in each category. This is done by separating the 40 possible SETs into their respective categories then dividing the amount of SETs in each category by the total number of SETs. For example, when dividing the 40 SETs into their categories, the number of SETs initially in each category is as follows:

| SET Category | Number of SETs | Explanation |
|--------------|----------------|----------------|
| 4D | 8 | 20% of 40 = 8 |
| 3D | 16 | 40% of 40 = 16 |
| 2D | 12 | 30% of 40 = 12 |
| 1D | 4 | 10% of 40 = 4 |

When we divide the number of initial SETs in 4D by the total number of SETs, we would get 20% (corresponding to its initial percentage). Therefore, there is a beginning 20% chance the missing cards will be in the 4D category. To narrow down which category the missing cards are in, we follow the same algorithm of drawing a card from the claimed pile to eliminate SETs. As SETs are eliminated, the number of SETs in each category will decrease, therefore also decreasing or increasing the chances the missing cards are in each category. For example, let's say the first card we draw from the claimed pile is in category 4D. That would mean there are now 7 SETs in 4D and 39 total SETs. Therefore, if we divide 7 by 39, we would get a percentage of approximately 17.95; its chances have decreased. However, the chances of the missing cards being in the other categories will increase because their SET number stayed the same while the total number of SETs decreased.

Trial, Results, and Discussion

We ran the model in BehaviorSpace 1000 times. This yielded a significant amount of data. However, due to our limited time constraint, we were not able to conduct an in-depth analysis of our data. Nevertheless, skimming through some of the data, we did not see any obvious patterns that stood out to us. However, before looking through our data, we hoped that the 1D column could provide us with some useful information. Because the chances of the missing cards being in that category was so low, we expected that after we drew a certain number of cards and 1D had not been completely eliminated yet, then it had a high likelihood of being the missing card. We did not have the chance to truly try to prove this theory, but we did see some trends to suggest this. Overall, we have the intention of seeking mathematicians who would be willing to help us analyze our data.

Applications and Final Conclusions

Because finding missing pieces of information is crucial, there are many potential applications for this project. For fields such as accounting or logistics, missing pieces of data can be a hassle to deal with. Take accounting for example. What happens if money, receipts, or invoices are missing? Could we use the method for finding a missing card in this field to find the missing pieces of information on those receipts or invoices? If we managed to find a way to scale it up, perhaps we can.

Additionally, this might be applied to cybersecurity as well. Missing pieces of data can be vital in dealing with threats. Often times, there are known variables and unknown variables that must be taken into account of. There are the *knowns* (what we know), the *unknowns* (what we don't know), the *known unknowns* (what we know we don't know), and the *unknown unknowns* (what we don't know we don't know). In the case of the card game SET, when we had one or two missing cards, we realized that we had a known unknown (we knew that we were missing 'x' number of cards but didn't know which card(s) it was). However, we managed to use our knowns (cards left over or the cards we had) to find our unknown (the missing card(s)).

To conclude, we were able to accurately determine what the two missing cards were by generating the possibilities they could be then using the information we had to narrow down those possibilities. We were not able to analyze our optimization method. However, our overall objective of this project was to look at a way to create a systemized method for finding missing pieces of data.

Bibliography

End Game SET. 2017. Web. <<https://www.setgame.com/end-game-set>>.

Game of Set. 8 August 2013. Web. <<https://whieldon.wordpress.com/2013/08/08/game-of-set/>>.

Gordon, Gary, et al. *The Joy of SET*. Princeton: Princeton University Press, 2017. Print.

Harriss Spiral, Math Snacks, and SET. 15 Jan 2015. Web.

<<https://mathmunch.org/2015/01/15/harriss-spiral-math-snacks-and-set/>>.

Klarreich, Erica. *Simple Set Game Proof Stuns Mathematicians*. 31 My 2016. Web.

<<https://www.quantamagazine.org/set-proof-stuns-mathematicians-20160531/>>.

What is modular arithmetic? 2017. Web. <<https://www.khanacademy.org/computing/computer-science/cryptography/modarithmetic/a/what-is-modular-arithmetic>>.

1 **Predicting Food Shortages in Africa from**
2 **Satellite Imagery**

3
4 New Mexico

5 Supercomputing Challenge

6 Final Report

7 April 3, 2018

8 LAHS56

9 Los Alamos High School

10 Lillian Kay Petersen

11 Mentors: Daniela Moody and Rick Chartrand

12 **Contents**

13 **1 Executive Summary 4**

14 **2 Introduction 5**

15 **3 Methods 7**

16 **4 Results 10**

17 4.1 Illinois 10

18 4.2 Africa 14

19 4.3 Africa: Prediction of Future Crop Production 16

20 **5 Conclusions 19**

21 **6 Personal Statement 20**

22 **7 Acknowledgments 21**

23 **List of Figures**

24 1 Illinois and Ethiopia yield over time 6

25 2 MODIS snapshots 8

26 3 Illinois NDVI: drought year versus wet year 9

27 4 Illinois corn yield and NDVI anomaly 11

28 5 Corn yield, NDVI anomaly, and NDWI anomaly: drought year versus wet year . . . 11

29 6 Correlations between Illinois corn yield and NDVI, NDWI, and EVI anomaly . . . 12

30 7 Correlations over time 12

31 8 Multivariate regression 13

32 9 Ethiopia, Tunisia, Morocco: NDVI over time 13

33 10 Correlations to crop production in Ethiopia, Tunisia, and Morocco 15

34 11 Ethiopia: drought year versus wet year 15

35 12 Boxes of satellite data examined in Africa 17

| | | | |
|----|----|---|----|
| 36 | 13 | Summary of African correlations | 17 |
| 37 | 14 | Africa: Real-time crop predictions | 18 |
| 38 | 15 | Farm fields: Ethiopia versus Illinois | 19 |

39 **List of Tables**

| | | | |
|----|---|---------------------------------------|---|
| 40 | 1 | NDVI, EVI, NDWI definitions | 9 |
|----|---|---------------------------------------|---|

1 Executive Summary

Developing countries often have poor monitoring and reporting of weather and crop health, leading to slow responses to droughts and food shortages. Here, new satellite data analysis tools were created to monitor crop health in Africa. The method was first tested in Illinois where there is reliable, high-resolution crop data. Measures of vegetation health were computed from 120-meter resolution MODIS satellite imagery since 2000. The author wrote 4000 lines of python code to process 12 terabytes of data. Correlations were computed between corn yields and monthly satellite index anomalies for every county and year, and a multivariate regression using every index and month (1600 values) produced a correlation of 0.86 with a p-value $<1e-6$. The high correlations in Illinois show that this model has good forecasting skill for crop yields. Next, the method was applied to three countries in Africa: Ethiopia, Tunisia, and Morocco for each country's main crop. All three had high correlations with the maximum monthly satellite index during the rainy season, at 0.99, 0.97, and 0.73 respectively. The satellite analysis methods and software tools developed here can be used to predict crop production two to four months before the harvest, and many more months before official crop data is published. Satellite imagery was then processed for every African country, and a publicly viewable interactive map displaying real-time crop predictions was posted online. This method is unique because it can be applied to any location, crop, or climate, making it ideal for African countries with small fields and poor ground observations. The author is actively engaged with several international aid organizations that are interested in using this early warning system of impending food shortages.

2 Introduction

In the United States, there is exceptional monitoring and reporting of weather and crop health, with thousands of weather stations and county-level crop yield data from the USDA that has been recorded since 1910 (Hamer et al., 2017; Menne et al., 2012). With this substantial amount of publicly available data, crop yields may be predicted based on historical records. However, not all parts of the world have open, reliable data (McKinnon, 2016). The availability of weather and crop data depends on the government's ability to collect it, financial resources, and willingness of authorities to share it. Lack of data is an especially important problem in developing countries where crop yields are less stable and droughts can lead to famines, death, government instability, and war. Therefore, there is a major need to monitor crop health in the developing world. Satellites provide coverage over the entire earth and certain bands may be used to assess plant health and drought conditions. This would enable scientists to monitor risks of food shortages and alert governments and international aid organizations in real time.

Crop yields in developing countries do not benefit from the same level of agricultural technology as the US, and therefore have much lower yields. Since 1970, corn yields have doubled in the US from 80 bu/acre to 160 bu/acre due to improvements in agricultural technology such as irrigation, pesticides, herbicides, fertilizers, and plant breeding (Figure 1a). In developing countries, crop yields are both much lower and much more variable than in the US, both geographically and in time (Mann and Warner, 2017b). For example, Ethiopia's corn yield has increased from 15 to 55 bu/acre since 1960 (Figure 1b), which is still one-third the corn yield of the US. Farmers in poor countries lack the financial resources and education to use the advanced technology used by the American and European farm industries. Therefore, crop yields in African countries are much more susceptible to the dangers of heat waves and droughts.

Satellite imagery has been extensively used for crop monitoring for decades. The majority of these studies are in the United States, where there is an immense amount of yield and production data at high resolution. Such data significantly improves agricultural research, but is only affordable by developed countries. The US also has large fields of a small number of individual crops, mainly corn, soybeans, and wheat. This allows research to be specific to individual crops and locations, and uniform crops within each satellite pixel. For example, Johnson (2016) developed algorithms

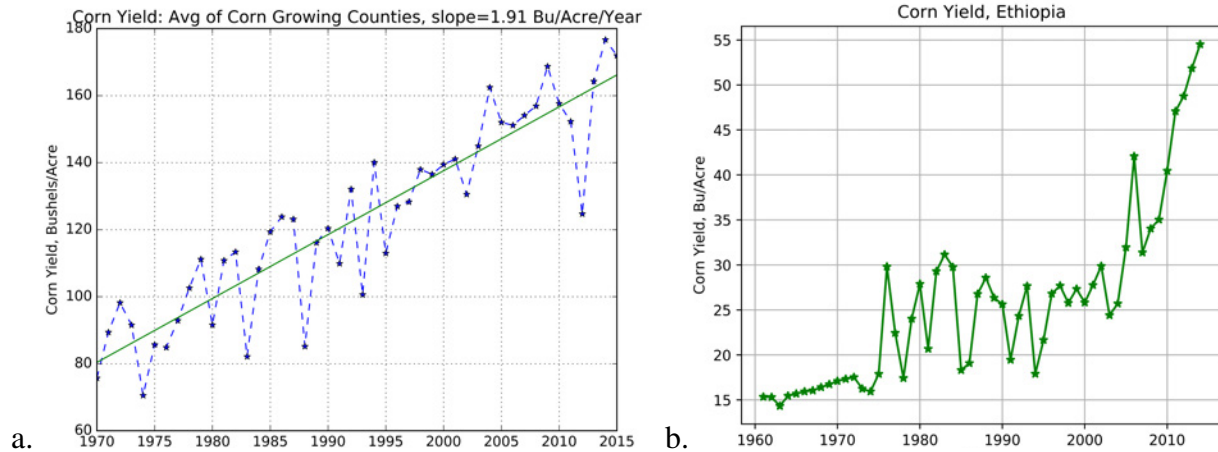


Figure 1. Illinois (a) and Ethiopia (b) corn yield over time. Both have improved significantly, but the yield in Ethiopia is still one-third of the U.S. All plots in this paper were created by the author.

90 to identify crops in the US from MODIS imagery and analyzed each crop individually. [Gao et al.](#)
 91 (2017) utilized week-by-week plant growth data in Iowa to design a method to monitor the growth
 92 stages of corn and soybeans from satellite imagery.

93 These types of studies are not possible in Africa because there is minimal reporting of crop
 94 health and yields; farms consist of very small plots of varied crops interspersed with buildings; and
 95 the continent contains a vast number of different climates, growing seasons, and crops. Despite
 96 these difficulties, a few studies have examined the climatology of specific countries or regions in
 97 the developing world. [Gissila et al. \(2004\)](#) correlated seasonal rainfall in Ethiopia with sea-surface
 98 temperature anomalies in the Indian ocean and the central Pacific. [Tadesse et al. \(2014\)](#) predicted
 99 NDVI (Normalized Difference Vegetation Index) 1–3 months in the future from multiple indices
 100 (land cover, standardized deviation of NDVI, etc.) as a means of forecasting droughts. Other studies
 101 develop models that forecast crop yields. NDVI/yield regressions for cereals at national level have
 102 been developed for specific countries in northern Africa ([Rembold et al., 2013](#)). [Mann and Warner](#)
 103 (2017b) use kebele (district) level data collected by the Ethiopian government, including crop
 104 damage, elevation, fertilizer use, population density, and road density, to estimate wheat output
 105 per hectare. I contacted the Ethiopian Central Statistical Agency, Mann, and Warner in an attempt
 106 to obtain this detailed, high-resolution data. Unfortunately, the Ethiopian government refuses to
 107 release data, even for agricultural research. Mann and Warner were only able to obtain this data
 108 under strict conditions and after years of collaboration ([Mann and Warner, 2017a](#)). These factors

109 all contribute to the difficulty of developing predictive tools for crop yields in Africa.

110 The method of predicting crop yields developed here differs from previous work in the U.S. and
111 Africa because it is an overall measure of relative vegetation health compared to the mean climate
112 on a per-pixel bases. Unlike previous studies, it may be applied anywhere in the world—it does not
113 depend on special tuning for the particular crop, region, or climate of interest. The method was
114 created for developing countries where detailed monitoring on the ground simply does not exist,
115 but was successfully validated against extensive crop data in Illinois.

116 **3 Methods**

117 The overall goal of this research is to create a predictive measure of crops computed from satellite
118 data. Python code was written by the author in order to obtain satellite images, mask out clouds,
119 calculate vegetation and water indices, compute monthly anomalies since 2000, and correlate the
120 anomalies of the satellite indices with crop yield anomalies for every county in Illinois and then
121 apply the same method to three countries in Africa.

122 MODIS (Moderate Resolution Imaging Spectroradiometer) imagery was obtained from the
123 Descartes Labs satellite platform at a resolution of 120 meters (Figure 2a, 2b). MODIS, hosted
124 on the satellites Aqua and Terra, has a return time of one day, giving almost continuous imagery
125 of every location on earth since 2000. The instruments capture 36 spectral bands ranging from
126 wavelengths of 0.4 μm to 14.4 μm (Jenner, 2015).

127 Clouds and snow in images can disrupt data and distort values. In order to account for cloud
128 contamination, clouds were identified based on the values of the bands blue, red, NIR, and SWIR.
129 Pixels with clouds or snow were not included in monthly averages and images with over 80% clouds
130 were thrown out (Figure 2c).

131 To measure the health of crops throughout the growing season, three indices were computed:
132 NDVI, EVI, and NDWI (Table 1). All three indices range from -1 to 1. Areas containing dense
133 vegetation show high NDVI and EVI values, between 0.4 and 0.8, desert sands will register at about
134 zero, and snow and clouds are negative. NDVI is sensitive to chlorophyll, which absorbs visible
135 light, from 0.4 to 0.7 μm , for use in photosynthesis. In contrast, EVI detects canopy structural
136 variations, including leaf area, canopy type, and canopy architecture (Herring and Weier, 2000).

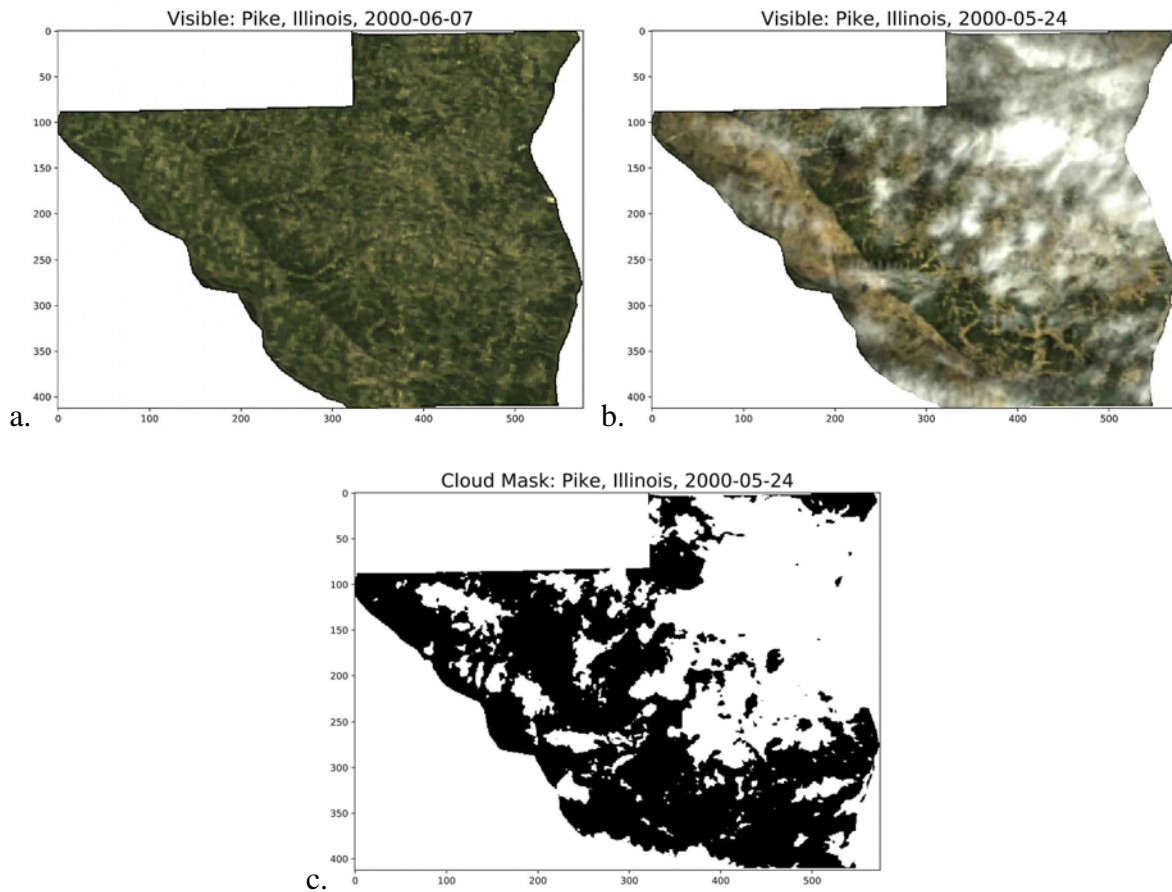


Figure 2. Snapshots of two MODIS satellite passes over Pike county, Illinois (a, b) and the cloud mask for the second image (c).

137 NDWI detects water content. Combined, all three indices complement each other on the detection
 138 of vegetation changes.

139 For every pixel in Illinois, the NDVI, EVI, and NDWI monthly averages and climatologies were
 140 computed. The climatology is defined as the average over years 2000 through 2016 for each month
 141 and pixel. Next, the monthly climatology was subtracted from the monthly average for every pixel,
 142 resulting in the monthly anomaly. The pixels in each county were then averaged together to find
 143 the monthly anomaly for NDVI, EVI, and NDWI.

144 Annual corn yield data was downloaded for every county in Illinois for years 2000 through
 145 2016 from the USDA (Hamer et al., 2017). Because each county has different growing conditions
 146 (soil quality, hills, proximity to large water bodies, etc.), the mean was subtracted out of each
 147 county's corn yield to find the yield anomaly. Correlations were found between each county's corn

| Index | Description | Formula |
|-------|--|--|
| NDVI | Normalized Difference Vegetation Index | $NDVI = \frac{NIR-Red}{NIR+Red}$ |
| EVI | Enhanced Vegetation Index | $EVI = G * \frac{NIR-Red}{NIR+C1*Red-C2*Blue+L}$ |
| NDWI | Normalized Difference Water Index | $NDWI = \frac{Green-NIR}{Green+NIR}$ |

TABLE 1. Definitions of indices to measure crop health. NIR is near infrared, L is the canopy background adjustment that addresses non-linear, differential NIR and red radiant transfer through a canopy, and C1, C2 are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band.

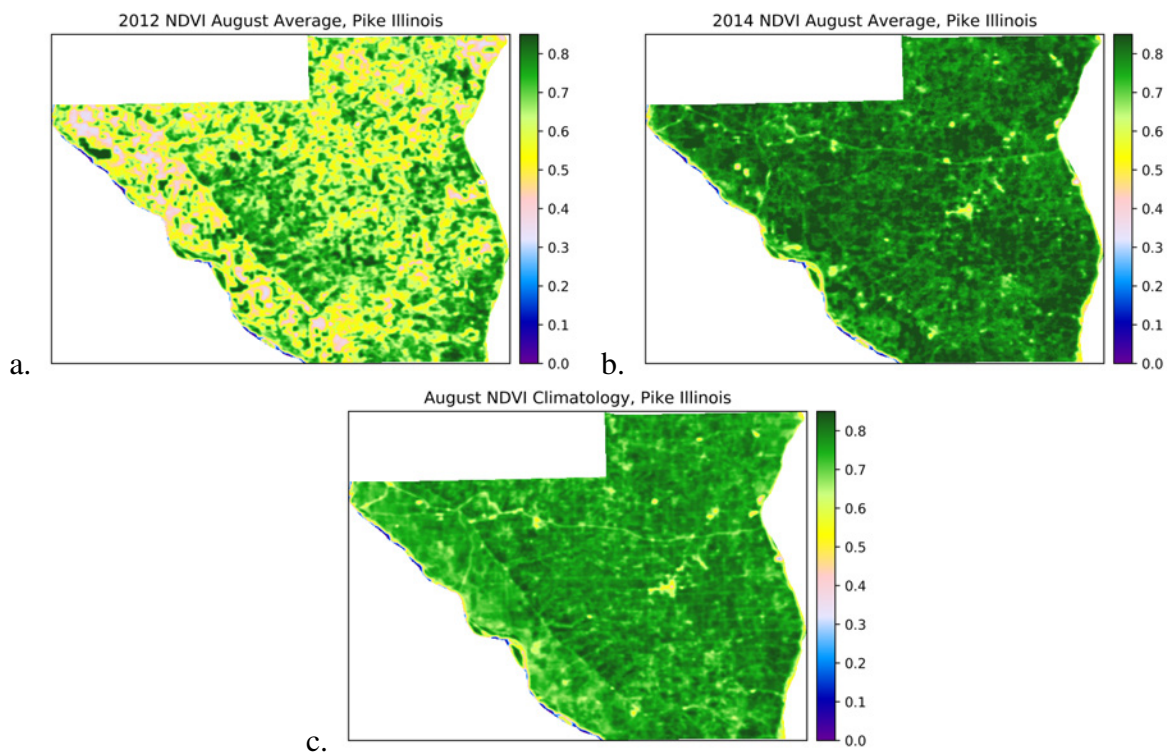


Figure 3. August average NDVI for a drought year (a) and a wet year (b), and the NDVI August climatology (c).

148 anomaly and the three satellite indices. To find the best prediction measure possible, a multivariate
 149 regression was fit to each month and index for a total of 15 variables.

150 The same method was then applied to three countries in Africa: Ethiopia, Tunisia, and Morocco,
 151 and later to every country in Africa. In each country, a box was analyzed where the majority
 152 of the crops are grown (Figure 12) and was then correlated to national crop production data
 153 from USDA (2018). A total of 4000 lines of code were written to process twelve terabytes of

154 raw data and produce the graphs. A code repository is maintained at the author's GitHub site:
155 <https://github.com/lillianpetersen/CropPredictionFromSatellite2018>.

156 **4 Results**

157 The method was first validated in Illinois and then applied in Africa.

158 **4.1 Illinois**

159 Illinois corn yield is highly correlated with NDVI, EVI, and NDWI. The correlations at the state
160 level are extremely statistically significant at 0.9, 0.85, and -.92 respectively (Figure 4). NDVI and
161 EVI both have a positive relationship to crop yields, and NDWI is inversely related. Strong NDWI
162 in critical growing stages could indicate insufficient evapotranspiration, resulting in a negative
163 correlation.

164 In 2012, the central United States was hit by a drought and Illinois had a lower than average
165 corn yield and a negative NDVI anomaly. Yields and NDVI anomalies in 2014 were significantly
166 higher. These two years are used as examples to show corn yield and satellite anomalies at the
167 county level (Figure 5).

168 Next, the satellite anomalies were plotted against the corn yield anomaly for every county and
169 year, for a total of 1559 points. August has the highest correlations with corn yields at 0.7, 0.71, and
170 -.73 for EVI, NDVI, and NDWI respectively (Figure 6). July has less predictive skill than August,
171 and the other months are almost uncorrelated with yields (Figure 7). All of July and August's
172 correlations have a P value less than 0.000001 (GraphPadSoftware, 2018), meaning there is less
173 than one in a million chance of them occurring through a random process.

174 Correlations have been computed with three indices (NDVI, EVI, and NDWI) and five months,
175 for a total of fifteen independent variables. In order to create a single predictive measure of corn
176 yields, a multivariate regression was fit to every index and every month using a Python machine
177 learning library. Figure 8 shows an example of the multivariate regression for two of the variables.
178 The multivariate regression improved the individual correlations to 0.86.

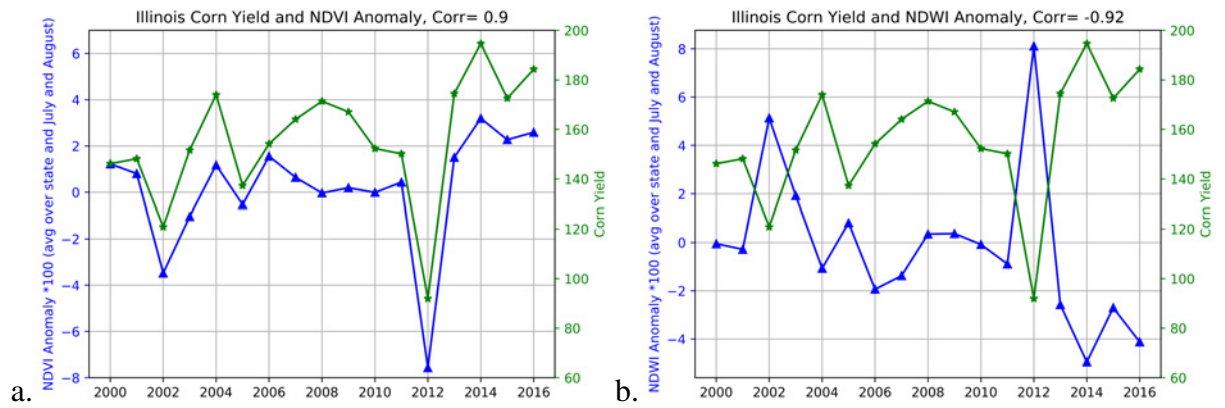


Figure 4. Illinois mean corn yield since 2000 (green) correlated with NDVI (a, blue) and NDWI (b, blue).

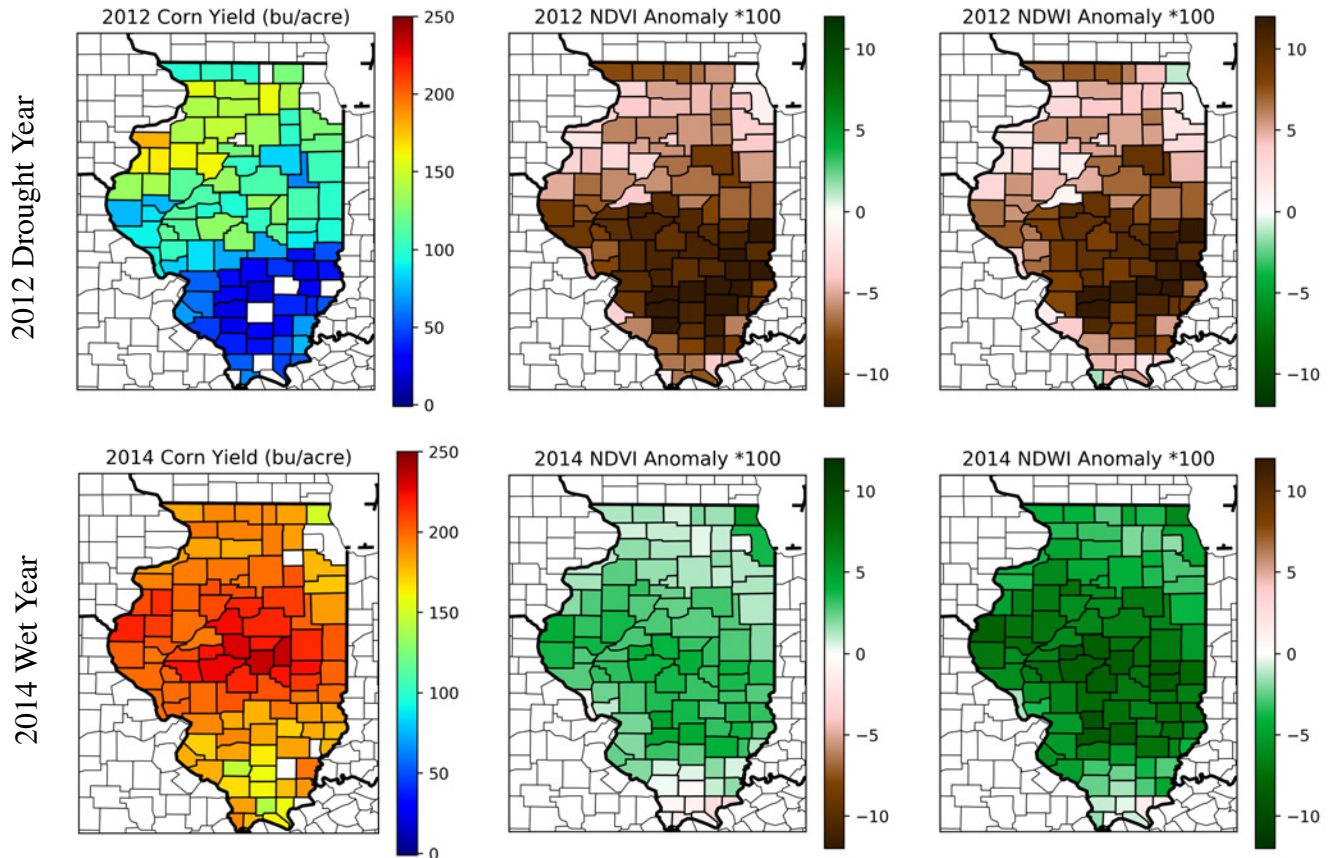


Figure 5. Corn yield (left), NDVI anomaly (center), and NDWI anomaly (right) by county in Illinois for the drought year 2012 (top) and for the wet year 2014 (bottom). During the drought year, there are low yields, low NDVI anomalies, and high NDWI anomalies, while the drought year is opposite.

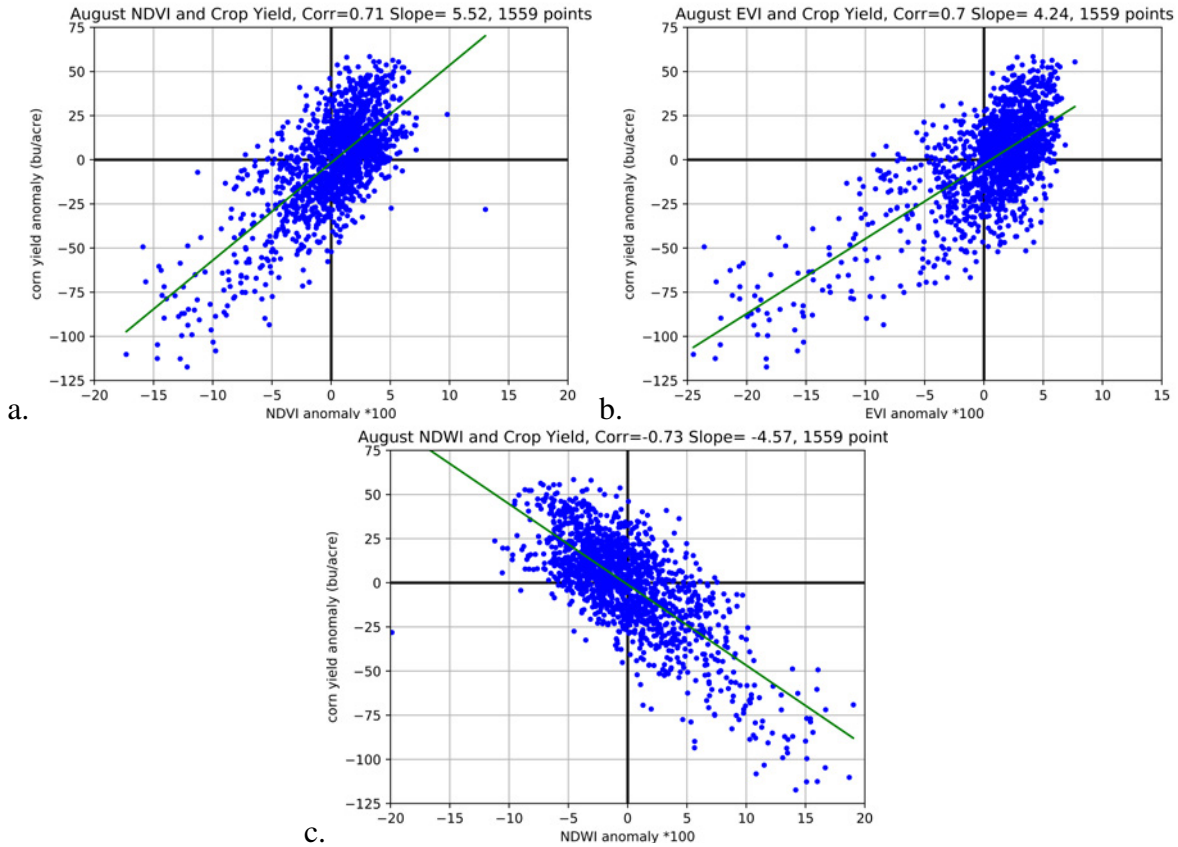


Figure 6. Correlations between Illinois corn yield and August average NDVI (a), EVI (b), and NDWI (c). All correlations are extremely significant with P values of <0.000001 . August had the highest correlations to yields out of all the months.

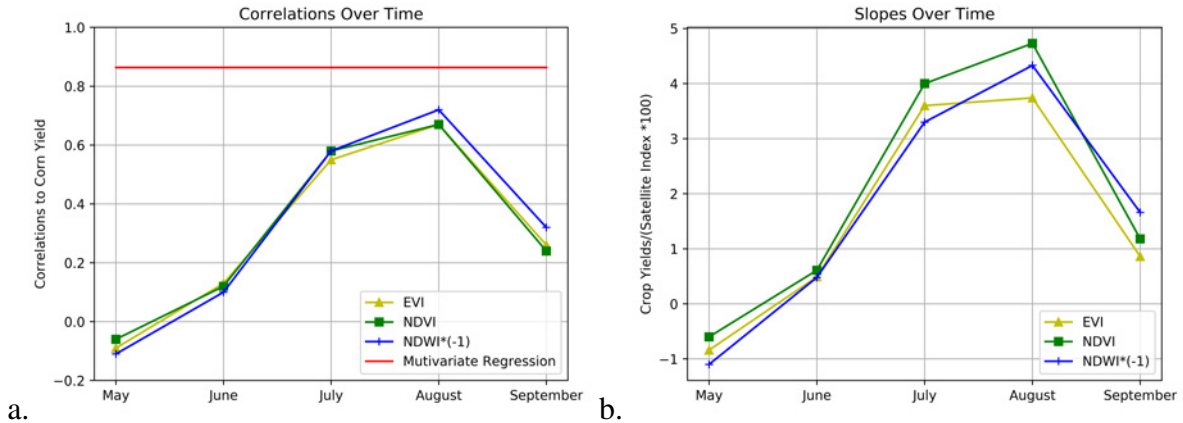


Figure 7. The absolute value of the correlations (a) and slopes of the linear regressions (b) for each month between Illinois corn yield and NDVI (green), EVI (yellow), and NDWI (blue). July and August have the highest predictive skill for crop yields which are harvested in October, meaning there is a two to three month lead time on yield estimates. The red line shows the correlation of the multivariate regression, which is higher than any individual month.

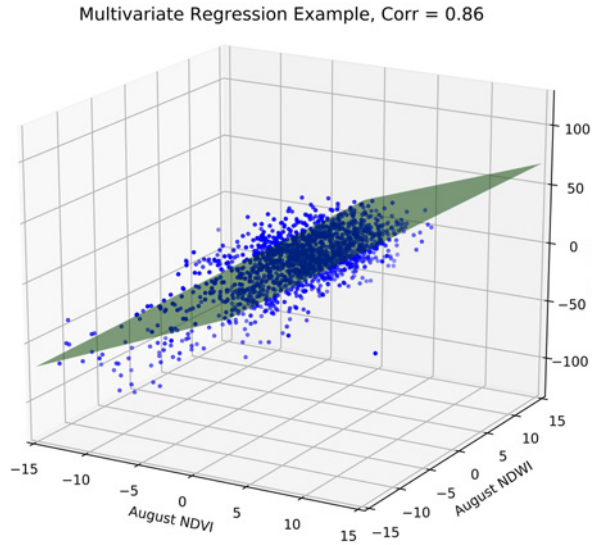


Figure 8. An example of the multivariate regression comprised of all three satellite indices and months. The multivariate regression improved the individual correlations to 0.86.

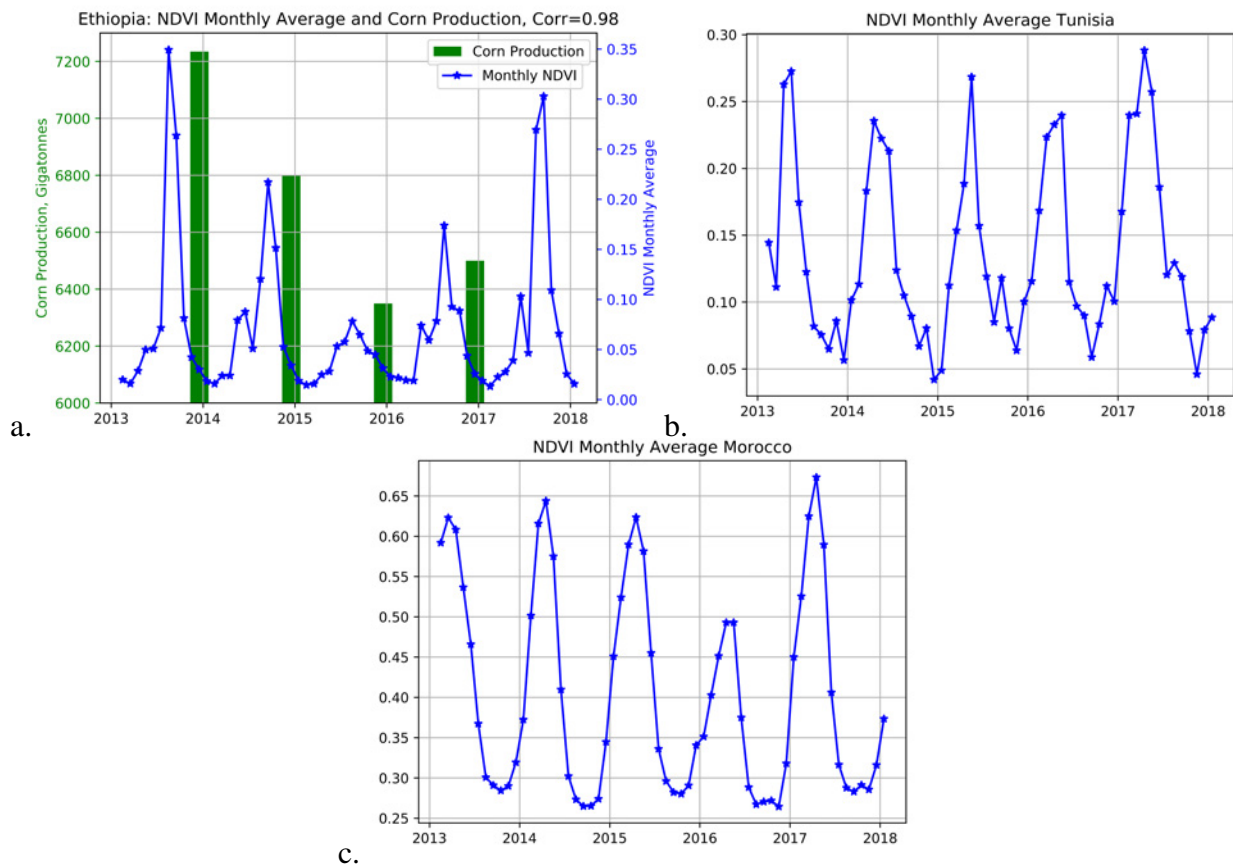


Figure 9. NDVI monthly average for Ethiopia (a), Tunisia (b), and Morocco (c). The annual rainy season produces high NDVI values and corresponds to the crop-growing months. Ethiopia also has the corn production overlaid, which has an almost perfect correlation to maximum NDVI at 0.98.

4.2 Africa

The high correlations in Illinois show that this model has good forecasting skill for crop yields. Next, this method was applied to three countries in Africa: Ethiopia, Morocco, and Tunisia. For each country, a box was analyzed where the majority of the crops are grown (Figure 11a).

In most places in Africa, there is a wet and a dry season. For example, the wet season in Ethiopia spans from June to September, and crops are harvested in December. This is known as the Meher growing season. Ethiopia's core agriculture and food economy is comprised of five major cereals: corn, teff, wheat, sorghum, and barley. These cereals accounted for about three-quarters of total area cultivated and 29 percent of the agricultural GDP in 2005/06 (Taffesse, 2012).

The wet and dry seasons are evident in the monthly NDVI values for all three countries (Figure 9). During the wet season, the crops green and the NDVI values spike. During the harvest, the values drop. The crop examined in each country was chosen based on the production quantity. Corn and sorghum were evaluated in Ethiopia, and wheat was examined in Tunisia and Morocco. It was found that Ethiopia and Morocco have the best correlation to the maximum NDVI value of the growing season, while Tunisia has the highest correlations to NDWI.

There was a major drought in Ethiopia in 2015, and 2013 was a very wet year by comparison. These vegetation differences can also be seen on the pixel level (Figure 11). The anomalies are especially evident in the Rift Valley where most of the crops are grown.

Ethiopia's maximum NDVI values, which usually occur in August, are extremely well correlated with grain production, at 0.98 and 0.99 for corn and sorghum respectively (Figures 9a, 10a). That is an almost perfect correlation between the crop production harvested in December and satellite imagery four months earlier. Tunisia has a correlation of 0.97 and Morocco has a correlation of 0.73 for wheat (Figure 10b, 10c), showing high predictive skill of satellite indices in all three countries.

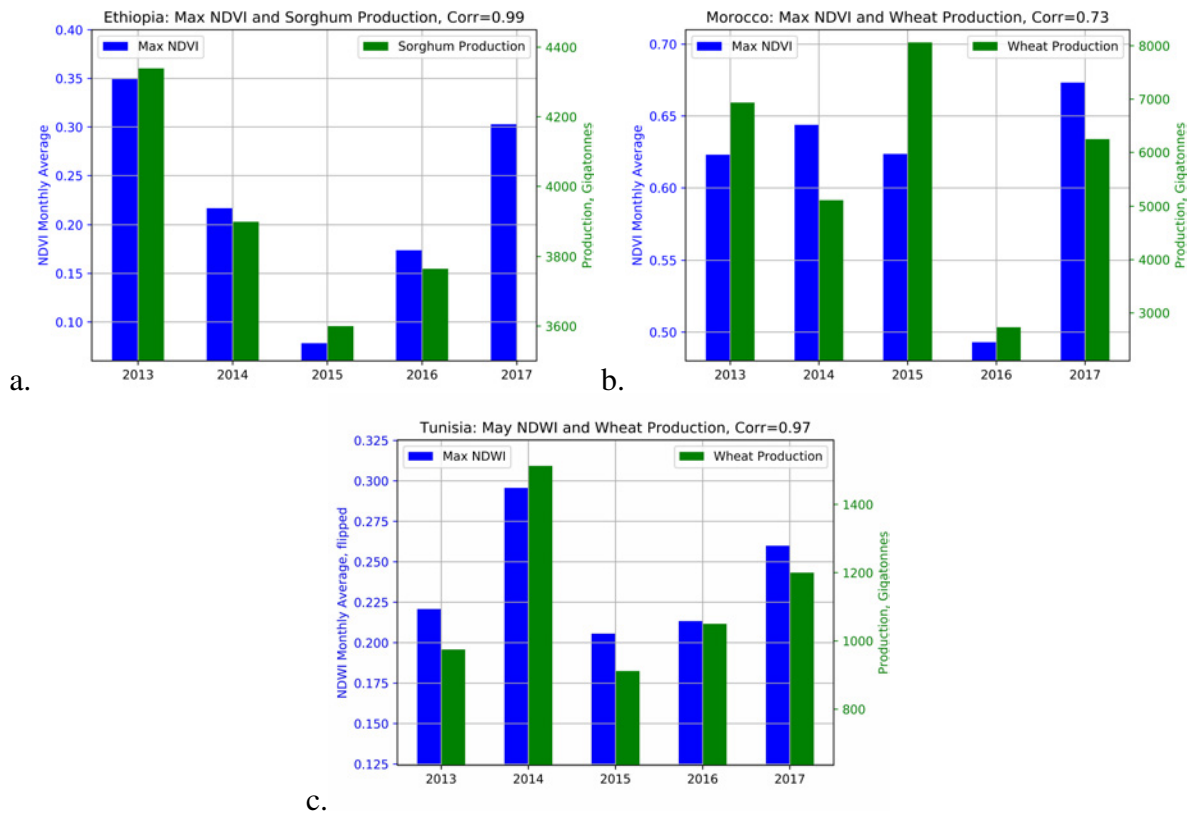


Figure 10. Maximum NDVI value of the growing season (green) with crop production (blue). All countries have significant correlations ranging from 0.99 to 0.73. Ethiopian production data for 2017/2018 has not been published because crops are harvested from November to February.

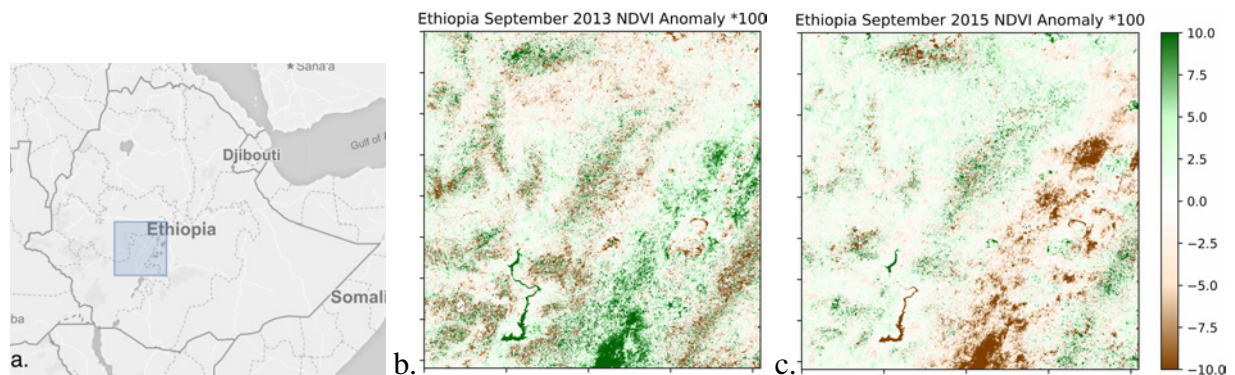


Figure 11. The box examined in Ethiopia (a) during a wet year (b) and a dry year (c). The NDVI anomalies are especially high in the rift valley, where farming is the most dense.

4.3 Africa: Prediction of Future Crop Production

Satellite imagery was processed for every African country. First, a box in an agricultural region was selected in every one of the 46 countries in Africa and a total of 12 terabytes of daily satellite imagery was processed according to the method above. Correlations and linear regressions were computed in every country and every crop. Difficulties in finding accurate correlations include:

- False reporting of production in some countries, due to lack of resources, poor oversight, or corruption (e.g. DR Congo, Eritrea, Libya)
- Multiple growing seasons in central countries (Rwanda, Somalia)
- Growing seasons across the December - January year boundary (Tanzania, Botswana)
- Clouds every day for months at a time in central African countries (Gabon, Cameroon)
- Time delays and misclassification of harvests in October–December, where production is incorrectly reported in the following calendar year (Nigeria, Sudan)

In every African country, correlations were computed between six indices (NDVI, EVI, NDWI, averages and anomalies) and for every crop. The highest correlation in each country was examined. Despite the above difficulties, two thirds of the correlations are considered to be statistically significant ($r \geq 0.75$ for five years, Figure 13)

Satellite imagery was then processed up to the current date for countries that are in growing season. Real-time predictions were computed for each of these countries and their highest correlating crop from the linear regressions (Figure 14). Next, an interactive map of the predicted production for harvests in the next few months was created and is now publicly viewable through https://lillianpetersen.github.io/africa_satellite. This map can give governments and aid organizations advance notice to see which countries are at the highest risk of a food shortage in order to better prepare supplies, transportation, and manpower for a rapid response.

The author is currently engaged with several international aid organizations who are interested in this product, including the International Food Policy Research Institute (IFPRI), the US Dept. of Agriculture (USDA), and the Global Agricultural Monitoring (GEOGLAM) group. The author has been invited to give hour-long talks to these institutions in Washington, DC.



Figure 12. A box was chosen in the densest agricultural region for each country in Africa.

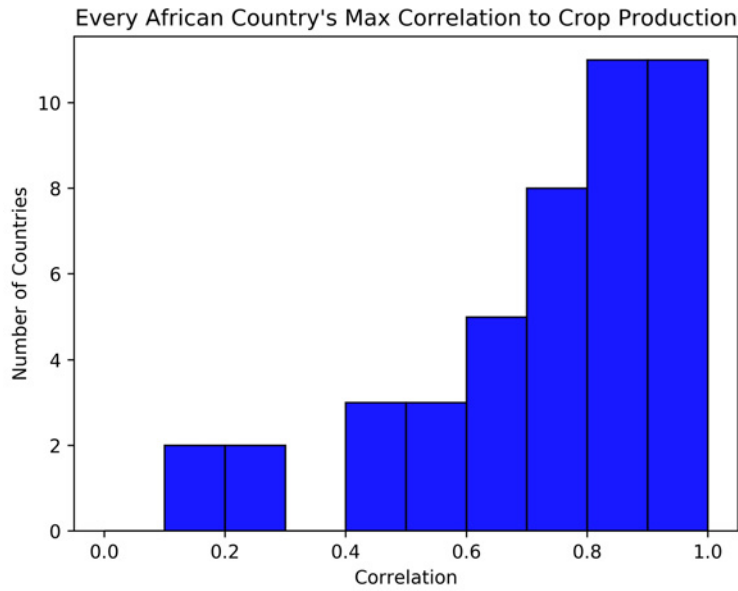


Figure 13. The highest correlations in every country in Africa. Two thirds are considered to be statistically significant.

Production Predictions: March 2018

For countries with height of growing season in October - February

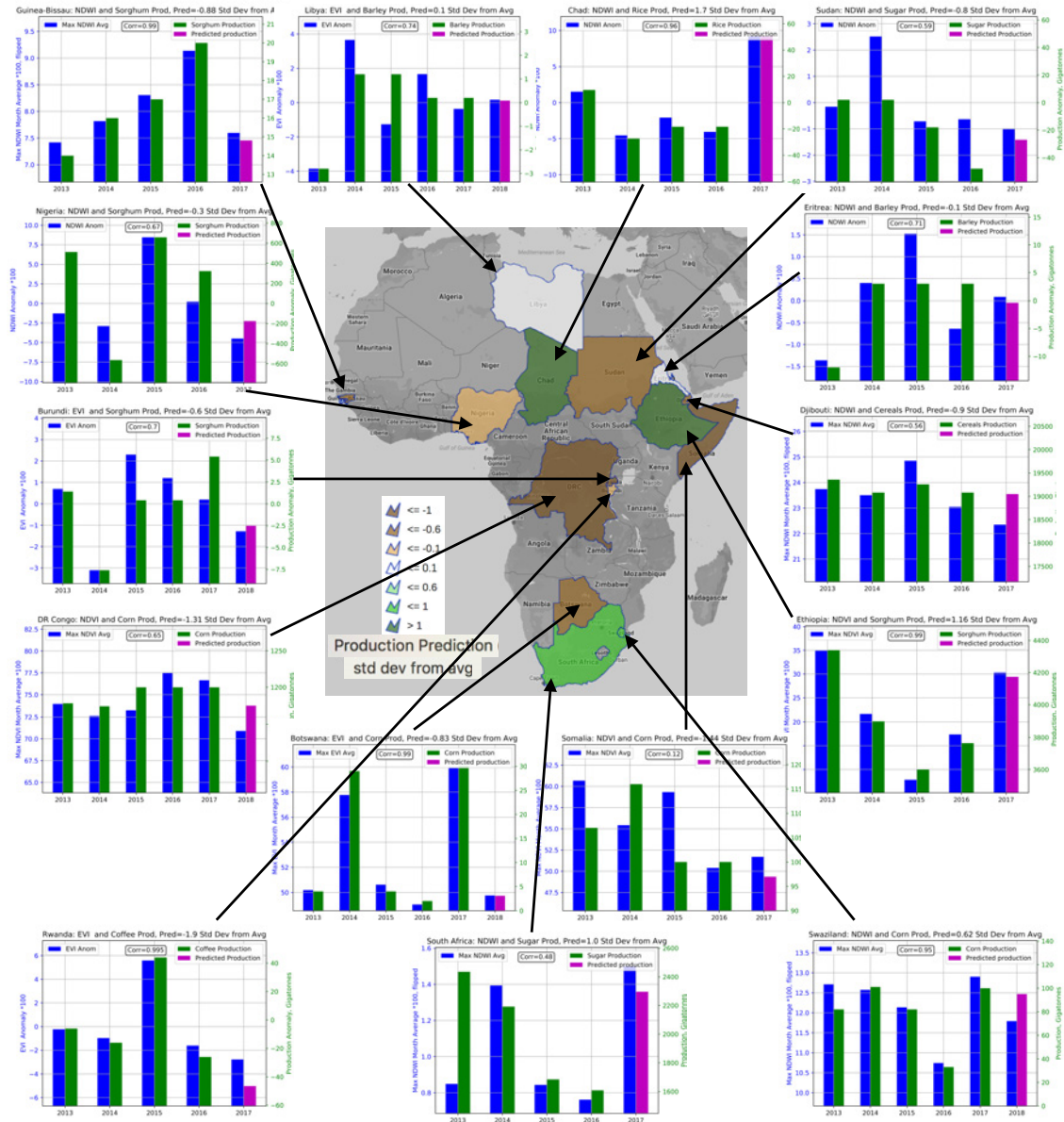


Figure 14. The map displaying the predicted production for every country currently in season in standard deviations from the average (middle). Surrounding the map are plots showing each country's highest correlation (crop and satellite index, green and blue) and predicted production (pink).



Figure 15. Farm fields by satellite in Ethiopia and Illinois at the same resolution. The small farm fields (smaller than a MODIS pixel) and poor ground truth data increase the difficulty of analyzing and predicting crop yields in Africa.

5 Conclusions

In this research, a method was developed to use three measures of crop health computed from daily MODIS satellite imagery as a predictive tool for crop yields 2–4 months before the harvest. The model was first validated in Illinois, where there is high-resolution yield data, by computing the linear fit between harvest yields in October (USDA, 2010) and the satellite indices in July and August. That is a three month prediction window, which could give farmers and insurance companies valuable information on the market months in advance. When a multivariate regression was fit to all months of the growing season and all three indices, the correlation peaked at 0.86 for 1600 data points. Next, the method was applied to three countries in Africa (Ethiopia, Tunisia, Morocco), all with different climates and crops. High correlations between maximum satellite indices and crop production were calculated in all three countries, with Ethiopia the highest at 0.99 to sorghum. After this success, satellite imagery was analyzed in every African country, and two thirds of the correlations proved to be statistically significant. Real-time crop predictions are now computed for every African country and are displayed on an interactive online map at https://lillianpetersen.github.io/africa_satellite.

244 Satellite imagery has been used to monitor and predict crop yields since the mid-1990s. How-
245 ever, most of these studies are completed in developed countries (e.g. US and Europe) because of
246 large amounts of ground truth data and large crop fields. Therefore, the method can be tuned to
247 specific crops and growing seasons. In Africa, it is almost impossible to tune the method because
248 of numerous crop types, climates, and growing seasons, as well as small farms and little to no crop
249 yield data (Figure 15). In the literature, there is no general measure of crop prediction that can be
250 applied to any crop, location, or climate. The method developed in this research is unique because
251 of its versatility, and has been shown to accurately predict crop yields across an entire continent.
252 It can be applied anywhere because it computes an overall measure of relative vegetation health
253 compared to the mean climate on a per-pixel bases.

254 In Ethiopia in 2015 and 2016, there was a major drought and food shortage, and eight million
255 people were at risk of starvation. However, the Ethiopian government did not have sufficient
256 monitoring and reporting of drought and crop conditions during the growing season, “leading to a
257 crucial delay in the international response.” (Laing, 2016). The satellite analysis tools developed
258 for this project can observe drought conditions as they develop and predict crop failures up to
259 four months before the harvest and many more months before the Ethiopian government publishes
260 the crop production data. This could give aid organizations advance notice to organize an early
261 response to famine. Luckily, 2017 has had much higher NDVI values, indicating healthy crop
262 conditions, and hopefully an end to the current crises.

263 **6 Personal Statement**

264 The most significant result of this project was creating a predictive system of crop yields for every
265 country in Africa two to four months before the harvest. I have been told by researchers in this area
266 that my method is unique because it may be applied for any location, crop, or climate. After seeing
267 the predictive results of my model, several international aid organizations invited me to visit their
268 offices. For example, I will be giving an institution-wide seminar to the International Food Policy
269 Research Institute on May 2, which will be advertised to the larger Washington DC metro area. I
270 hope that they will be able to use my model in order to better predict future famines and save lives
271 through a faster response.

7 Acknowledgments

I would like to acknowledge to assistance of: Daniela Moody and Rick Chartrand, my mentors, who advised me on satellite data retrieval and data analysis methods; Mark Petersen, who helped me choose my research topic; and Phillip Wolfram for recommendations on my plots and presentation.

REFERENCES

- Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., Mueller, R., Johnson, D. M., and Prueger, J. H. (2017). “Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery.” *Remote Sensing of Environment*, 188, 9–25.
- Gissila, T., Black, E., Grimes, D. I. F., and Slingo, J. M. (2004). “Seasonal forecasting of the Ethiopian summer rains.” *International Journal of Climatology*, 24(11), 1345–1358.
- GraphPadSoftware (2018). “P value calculator, <<https://www.graphpad.com/quickcalcs/pvalue1.cfm>>.
- Hamer, H., Picanso, R., Prusacki, J. J., Rater, B., Johnson, J., Barnes, K., Parsons, J., and Young, D. L. (2017). “USDA/NASS QuickStats US crop data, <<https://quickstats.nass.usda.gov>>.
- Herring, D. and Weier, J. (2000). “Measuring Vegetation (NDVI & EVI) : Feature Articles.” NASA, <<https://earthobservatory.nasa.gov/Features/MeasuringVegetation/>> (August).
- Jenner, L. (2015). “Terra Spacecraft.” NASA (March).
- Johnson, D. M. (2016). “A comprehensive assessment of the correlations between field crop yields and commonly used MODIS products.” *International Journal of Applied Earth Observation and Geoinformation*, 52, 65–81.
- Laing, A. (2016). “Ethiopia struggles with worst drought for 50 years leaving 18 million people in need of aid.” *The Telegraph*.
- Mann, M. L. and Warner, J. M. (2017a). “AgSS Ethiopia wheat yield data.” *personal communication* (December).

295 Mann, M. L. and Warner, J. M. (2017b). “Ethiopian wheat yield and yield gap estimation: A
296 spatially explicit small area integrated data approach.” *Field Crops Research*, 201, 60–74.

297 McKinnon, K. (2016). “GHCN-D: Global Historical climatology Network daily temperatures |
298 NCAR - Climate Data Guide.” *Climate Data Guide*, <[https://climatedataguide.ucar.edu/climate-
299 data/ghcn-d-global-historical-climatology-network-daily-temperatures](https://climatedataguide.ucar.edu/climate-data/ghcn-d-global-historical-climatology-network-daily-temperatures)>.

300 Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012). “An Overview
301 of the Global Historical Climatology Network-Daily Database.” *Journal of Atmospheric and
302 Oceanic Technology*, 29(7), 897–910.

303 Rembold, F., Atzberger, C., Savin, I., and Rojas, O. (2013). “Using Low Resolution Satellite
304 Imagery for Yield Prediction and Yield Anomaly Detection.” *Remote Sensing*, 5(4), 1704–1733.

305 Tadesse, T., Demisse, G. B., Zaitchik, B., and Dinku, T. (2014). “Satellite-based hybrid drought
306 monitoring tool for prediction of vegetation condition in Eastern Africa: A case study for
307 Ethiopia.” *Water Resources Research*, 50(3), 2176–2190.

308 Taffesse, A. S. (2012). “Crop production in Ethiopia.” *Food and Agriculture in Ethiopia Progress
309 and Policy Challenges*, University of Pennsylvania Press, 53–83.

310 USDA (2010). “Field Crops Usual Planting and Harvesting Dates.” *Report No. 628*,
311 <<http://usda.mannlib.cornell.edu/usda/current/planting/planting-10-29-2010.pdf>> (October).

312 USDA (2018). “Agricultural Production Statistics by Country,
313 <<https://www.indexmundi.com/agriculture/>>.

RSA Based Primality Test

New Mexico

Supercomputing Challenge

Final Report

April 4th, 2018

LAHS71

Los Alamos High School

Student:

Elijah Pelofske

Teacher:

Alan Didier

Abstract

Efficient and accurate primality testing is a key mechanism used to ensure digital security in the modern world. Ensuring accurate primality is critical for guaranteed accurate message encryption and decryption, accurate identity verification, as well as a secure cryptosystem, whose security is guaranteed by the discrete logarithm problem. There are two types of algorithms that determine whether an integer is prime or composite, known commonly as primality tests. The two types are, probabilistic which in general are faster but less accurate, and deterministic tests which are slower but guaranty accuracy.

The purpose of this project is to create a primality test that is based upon the asymmetric cryptosystem RSA. The RSA Based Primality Test works by using the first N primes respectively as both a prime factor of the modulus and encryption exponent. The integer being tested for primality is the other assumed prime factor of the modulus. A random message, M , that is less than N , is then encrypted and decrypted using the constructed cryptosystem. If that M is equal to original M , the test increments one on the count of true instances, otherwise it increments one on the count of false instances.

It was found that count of false instances for 2 factor RSA was the most efficient form of this test - for x supplied known primes between 2 and N , and y being the first odd counterexample, the line $y = 3.3 * x^{2.7}$ describes the RSA Based Primality Tests properties with a correlation coefficient, r , of 0.97.

Purpose

The purpose of this project is to create a primality test algorithm which is based on the RSA cryptosystem. The advantage of this type of test, be it probabilistic or deterministic, is that it relies on proven hypothesis and concepts, and would not function the same, i.e. have the same flaws, as other types of probabilistic primality tests. Additionally, so far this exact type of test has not been implemented in this manner.

Introduction

The accurate and efficient primality testing is critical for constructing asymmetric cryptosystems. Accuracy of primality testing is important because encryption and decryption of every message less than N must be accurate. In addition to inaccurate encryption, inaccurate primality testing can result in cryptographically insecure keys. Efficiency of primality testing is important because cryptosystem construction must not take an intractable amount of computation time. Digital signatures, secure message transmission, and identity verification are all incredibly useful applications of asymmetric cryptography, as well as being a vital part of modern digital industry on the internet.

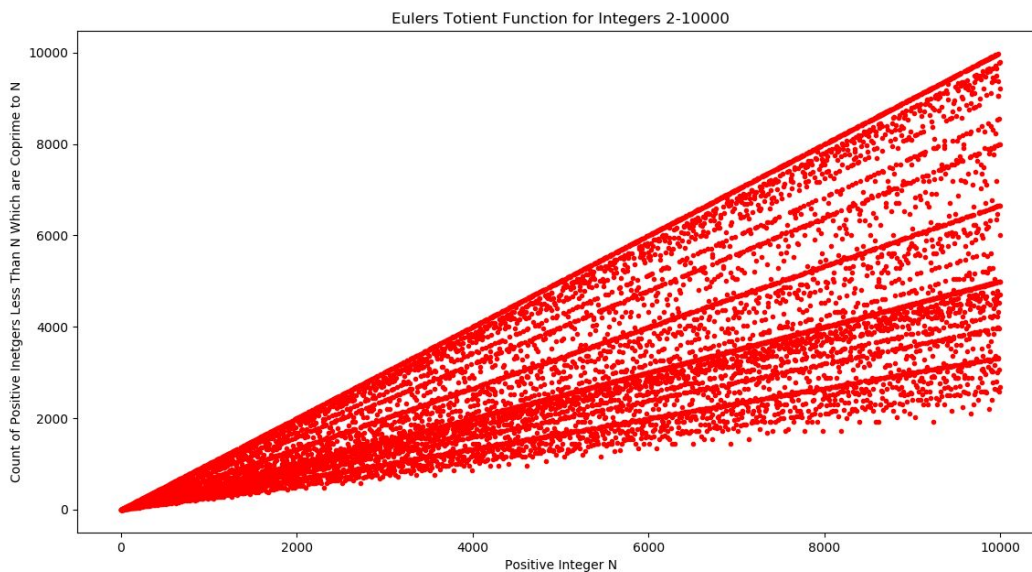
Examples of asymmetric cryptosystems (aka Public-Key cryptography) or key exchanges include RSA, Diffie-Hellman key exchange, ElGamal encryption, Paillier cryptosystem, and the Elliptic Curve cryptographic versions of all of these. In summation, efficient and accurate primality testing is critical for modern digital security.

The RSA cryptosystem works using two key pairs; (e, N) and (d, N) . The two key pairs are inverse permutations of each other via the simple function of raising the

message to encryption or decryption exponent (e or d) remainder division the modulus, N. As long as d is kept private, RSA can be used as identity verification as well as a secure cryptosystem. The operation of raising a number, M, to the power of e modulo (remainder division) N can be speed up on computers using a technique known as modular exponentiation. The modulus N is constructed from two or more known prime factors, and thus the Euler Totient function, denoted as $\phi(N)$, or the count of integers less than N which are coprime to N, can be constructed from the known prime factorization of the modulus, N. Figure 1 shows Euler's Totient function for all positive integers less than 10000. The public (encryption) exponent, designated as e, must be coprime to N and $\phi(N)$. The decryption exponent, d, must satisfy the equation $1 = d * e$ modulo $\phi(N)$, and can be computed efficiently using the Extended Euclidean Algorithm.

Probabilistic primality tests do not guarantee a certificate of primality. Most probabilistic tests can determine if a number is definitively composite, but sometimes will output a false positive for primality. Other probabilistic primality tests simply use mechanisms which have not been fully proven to guarantee primality. Deterministic primality tests give a definitive output of either primality or compositeness.

Figure 1



Experiment/Model

The RSA Based Primality Test works by using the first N primes respectively as both a prime factor of the modulus and encryption exponent. The integer being tested for primality is the other assumed prime factor of the modulus. A random message, that is less than N, is then encrypted and decrypted using the constructed cryptosystem. If that M is equal to original M, the test increments one on the count of true instances, otherwise it increments one on the count of false instances. It should be noted that earlier and less reliable versions of the RSA Based Primality Test, such as the several possible multiprime implementations, are not included in this final version. There are 18 earlier versions versions which include some of these many elements that were found to

be not as useful. Figure 2 below is the final version of the RSA Based Primality Test coded in Python3.

Figure 2

```
import random
import time
from fractions import gcd
#The Miller Rabin primality test is used to find the known primes for the RSABPT
def miller_rabin(n):
    k = 10
    if n == 2:
        return True
    if n % 2 == 0:
        return False
    r, s = 0, n - 1
    while s % 2 == 0:
        r += 1
        s //= 2
    for _ in range(k):
        a = random.randint(2, n - 1)
        x = pow(a, s, n)
        if x == 1 or x == n - 1:
            continue
        for _ in range(r - 1):
            x = pow(x, 2, n)
            if x == n - 1:
                break
        else:
            return False
    return True
def egcd(a,b):
    u, u1 = 1, 0
    v, v1 = 0, 1
    while b:
        q = a // b
        u, u1 = u1, u - q * u1
        v, v1 = v1, v - q * v1
        a, b = b, a - q * b
```



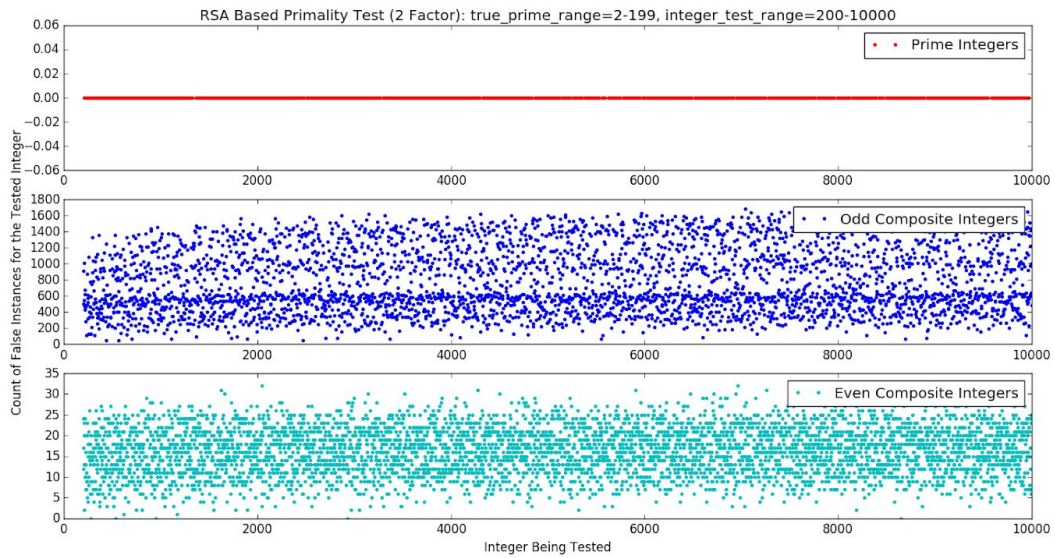
```
        if m1 != m:
            return False
    return True
start = time.clock()
p, t = rsabpt(65537, 20), time.clock() - start
print(p, t)
#Test example, 65537 is a fermat prime
#Also showing the time.clock() function for determining CPU seconds used
```

Results

The analysis of the functionality and capabilities of the RSA Based Primality test involved massive quantities of data analysis, therefore much of the benchmark results obtained are best illustrated in the form of graphs. All graphs were constructed using the Python library matplotlib. All CPU computation time benchmark tests were done on a Intel® Core™2 Duo CPU E4600 @ 2.40GHz × 2, and the operating system used was Kali Linux. All of the computations done for this project were completed across Windows 10, Kali Linux, and Ubuntu Linux machines, as well as online IDE's which ran the code on remote Linux machines. A total of 6 different computers were involved in gathering these results, including one which handled online IDE computations.

Figures 3, 4, 5, and 6 show some results of the 4 different possible conditions under which the RSA Based Primality Test could operate. The two defining aspects are 2 factor, or 3 factor, and whether the test outputs a count of true instances or a count of False instances. Figure 3 is the most promising form of the test, which is a 2 factor count of false instances test.

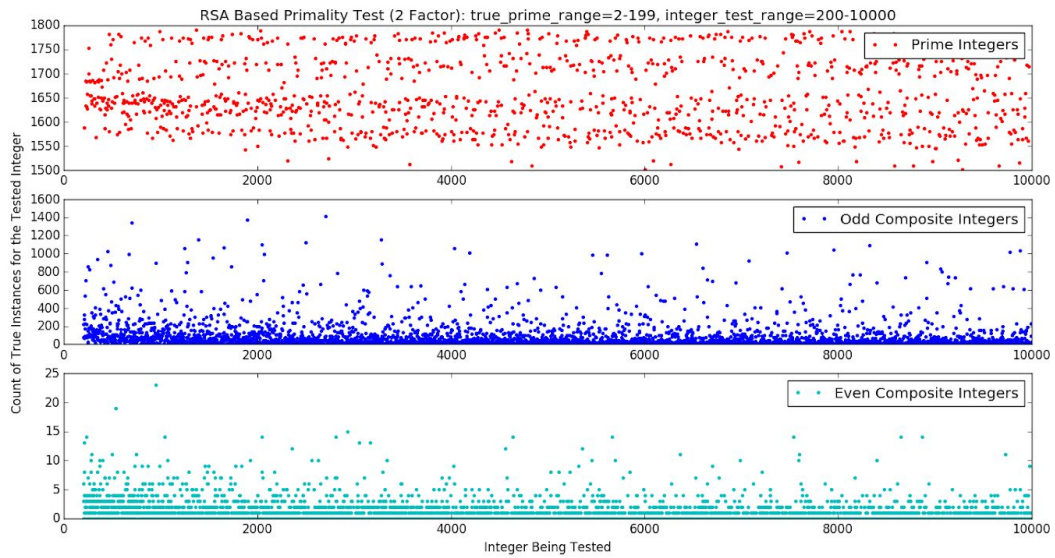
Figure 3



RSA Based Primality Test Minimums and Maximums for 2 Factor Count of False Instances

$(\max(\text{false_prime}), \min(\text{false_prime})) = (0, 0)$
 $(\max(\text{false_odd}), \min(\text{false_odd})) = (1683, 47)$
 $(\max(\text{false_even}), \min(\text{false_even})) = (32, 0)$

Figure 4



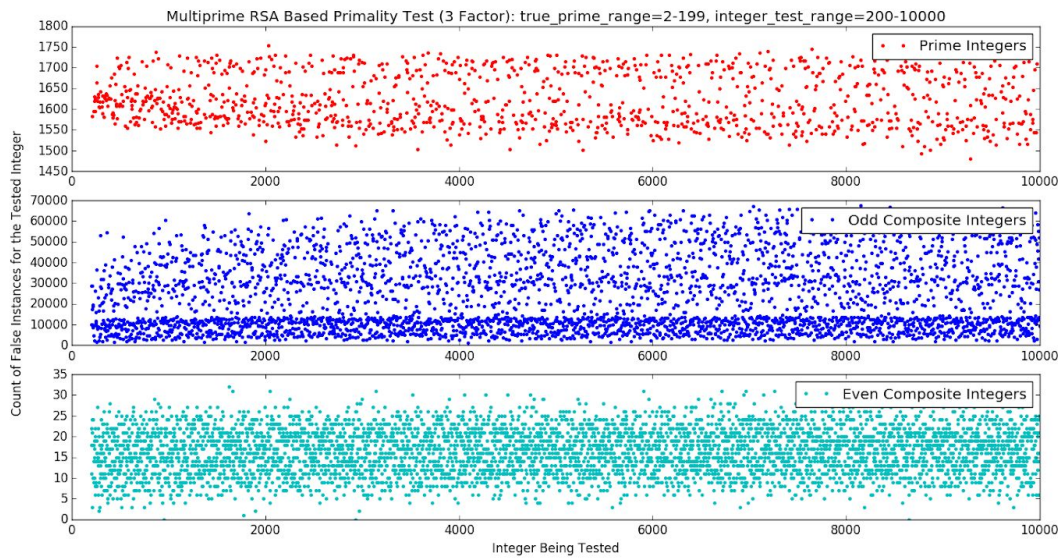
RSA Based Primality Test Minimums and Maximums for 2 Factor Count of True Instances

$(\max(\text{true_prime}), \min(\text{true_prime})) = (1790, 1502)$

$(\max(\text{true_odd}), \min(\text{true_odd})) = (1412, 0)$

$(\max(\text{true_even}), \min(\text{true_even})) = (23, 0)$

Figure 5



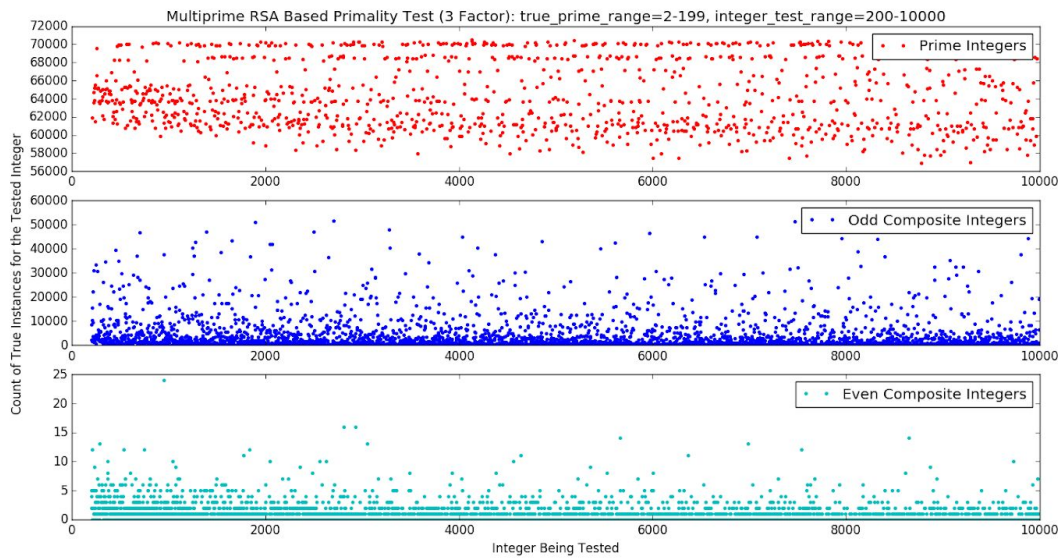
RSA Based Primality Test Minimums and Maximums for 3 Factor Count of False Instances

$(\max(\text{false_prime}), \min(\text{false_prime})) = (1748, 1489)$

$(\max(\text{false_odd}), \min(\text{false_odd})) = (67465, 1299)$

$(\max(\text{false_even}), \min(\text{false_even})) = (32, 0)$

Figure 6



RSA Based Primality Test Minimums and Maximums for 3 Factor Count of True Instances

$$(\max(\text{true_prime}), \min(\text{true_prime})) = (70483, 56913)$$

$$(\max(\text{true_odd}), \min(\text{true_odd})) = (51426, 2)$$

$$(\max(\text{true_even}), \min(\text{true_even})) = (24, 0)$$

Figure 7 below shows the percentage of incorrect instances for integers less than 10000 given an increasing number of known prime integers supplied to the RSA Based Primality Test, using count of false instances, 2 factor. An important note on the figure itself for clarity is that when $x = 9$, $y = 0$.

Meanwhile, Figure 8 shows percentage of incorrect instances for all positive integers. Again, this test was only done for the 2 factor, false count version of the RSA Based Primality Test.

Figure 7

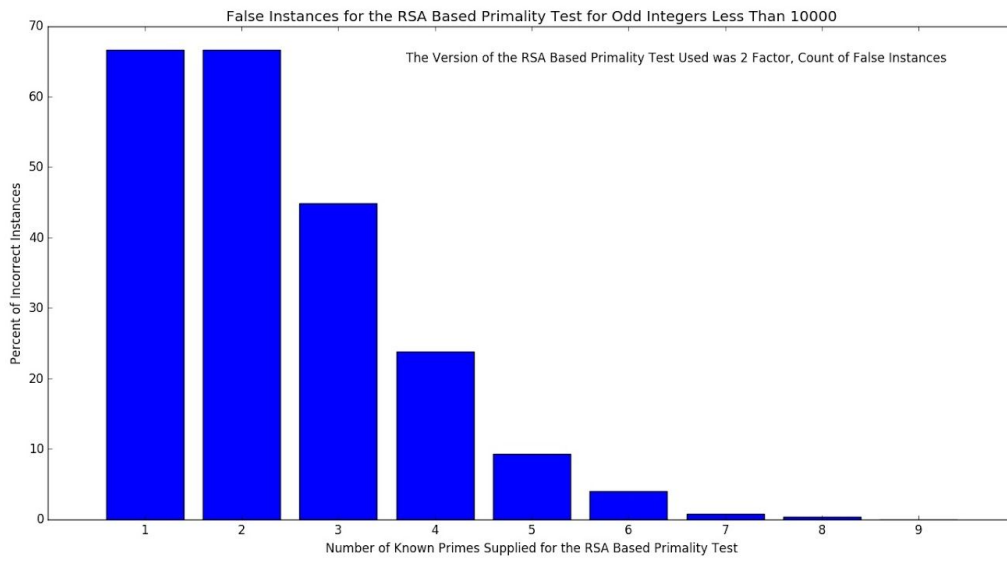


Figure 8

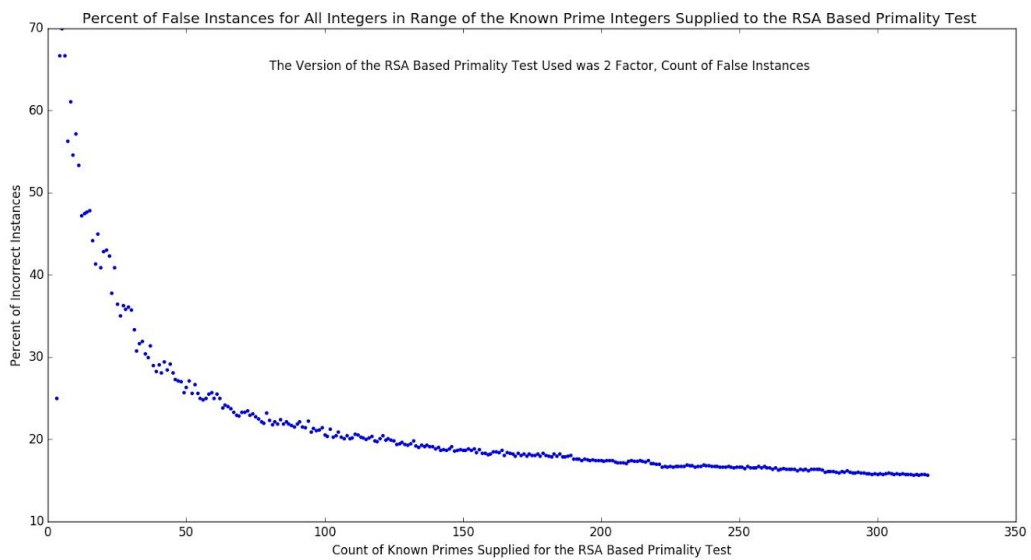
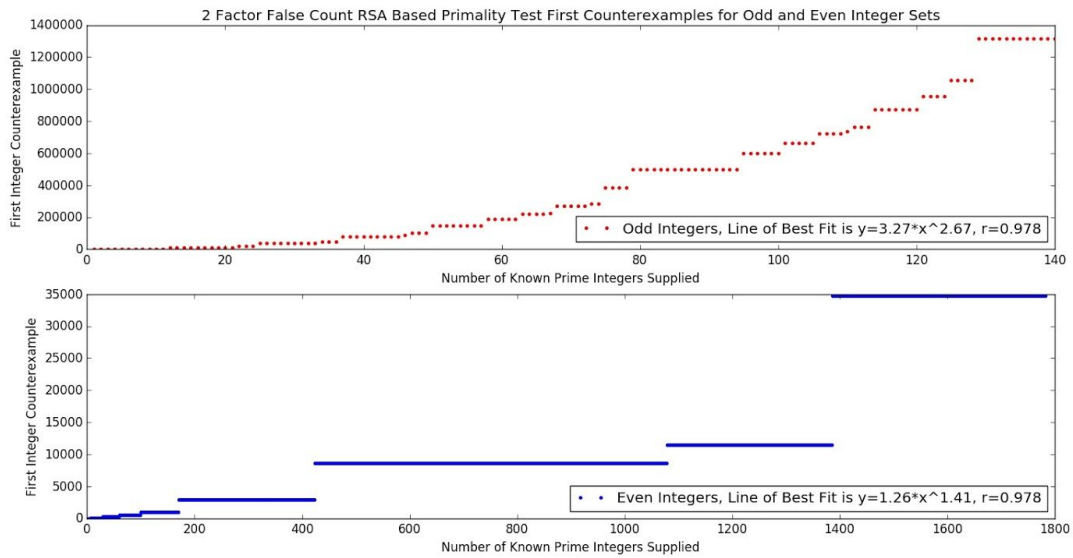


Figure 9 shows the curve that describes where the first counterexamples occur for each increment of known prime integers supplied to the RSA Based Primality Test 2 factor false count version, for both even and odd integers.

Figure 9



Figures 10, 11, and 12, are CPU computation time benchmark tests for a variety of differing bitlength prime integers.

Figure 10

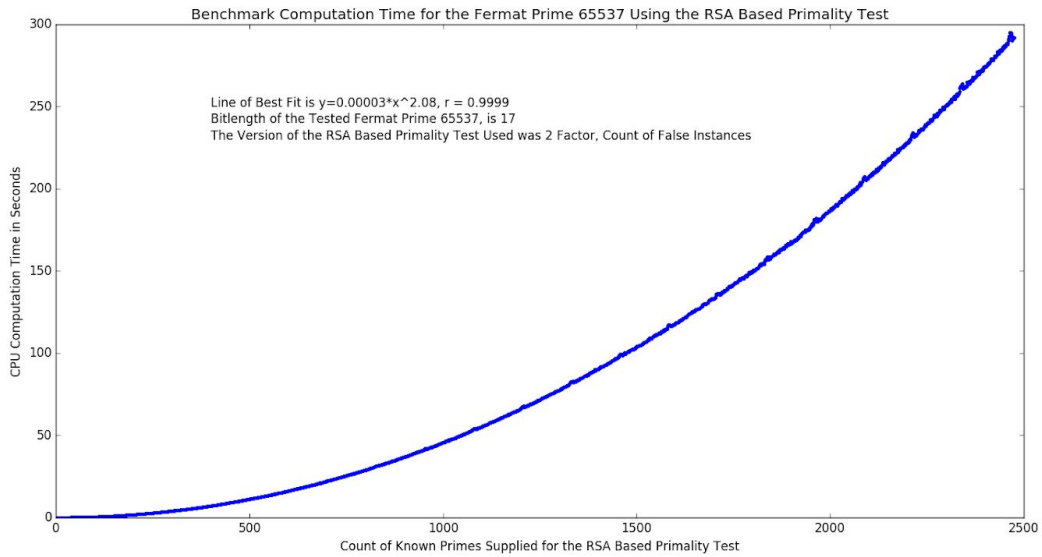


Figure 11

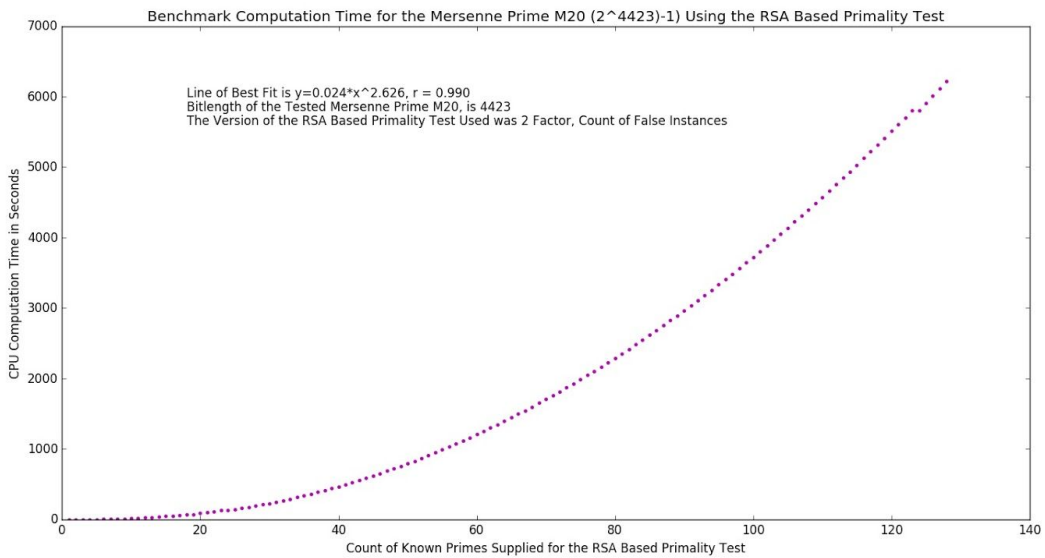
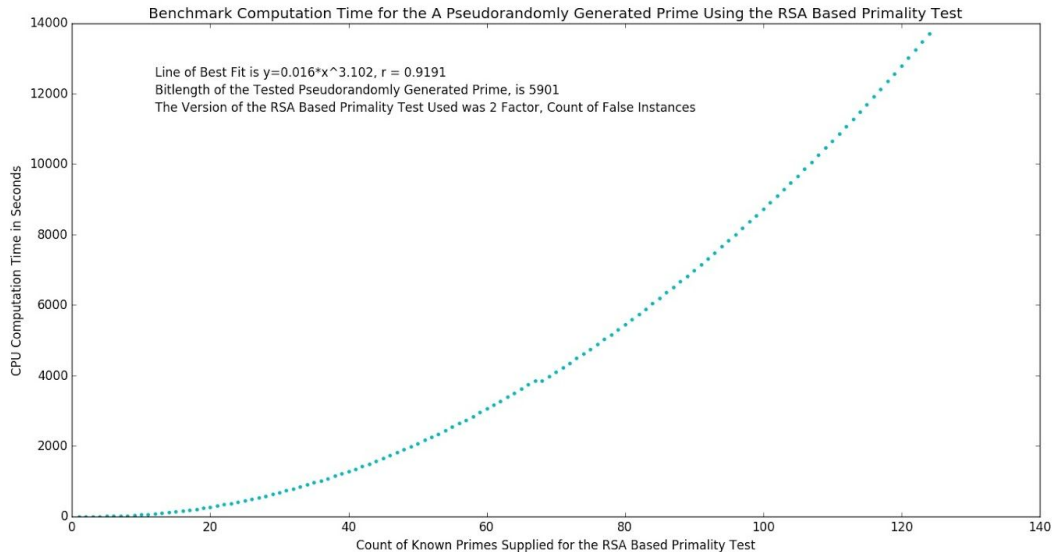


Figure 12



Analysis

Of the 4 types of the RSA Based Primality test, true and false count for 3 factor RSA (Figures 5 and 6), as well as true count of 2 factor (Figure 4) did not show clear differentiation of the bounds (minimums and maximums of the tested data set) for all numbers extending past the tested integer range, which in this case was 200-10000. However, the RSA Based Primality test for 2 factor false count (Figure 3) showed a clear differentiation between composite and prime integers (in this case using the known prime integer set between 2 and 199), making it a highly likely candidate for a primality test. It should be noted that the actual x value for these 4 tests is 46 - i.e. there

are 46 prime integers inclusively between 2 and 199.

The CPU computation time testing for primality was graphed for 3 different bitlength primes. The line of best fit and correlation coefficients for all datasets were very similar in their power lines of best fit, with consistently high correlation coefficients. The number of iterations for the RSA Based Primality Test, also referred to as the number of known prime integers supplied to the test, is limited for primality testing of a given integer N by the count of prime integers that exist between 2 and $N-1$.

The first counterexample for false positives is described by the line $y = 3.3 \cdot x^{2.7}$ with a correlation coefficient of 0.97 for odd integers. The line $y = 1.3 \cdot x^{1.4}$ describes the same conditions described above for even integers with a correlation coefficient of 0.97.

The percentage of false positives from the RSA Based Primality Test decrease dramatically for increased number of known prime integers supplied to the RSA Based Primality Test. However, for larger bitlength prime integers ($\text{Bitlength}(N) > 1024$), the computation time increases as well, making the RSA Based Primality test usable as both a probabilistic and deterministic primality test depending on how much computation time is available. It should be noted that integers within the range of known primes used in the RSA Based Primality Test for the percentage of false positives graph were not tested, at least for Figure 7.

What is significant about figure 8 is that it shows that the test can be used as a low accuracy probabilistic primality test within the range of used primes, and therefore it seems infeasible to use the number of repetitions required to make the test definitive

within that integer range. Additionally, since the test is designed to use only the first N primes, using the RSA Based Primality test in this manner is simply useless - only large bitlength ($\text{Bitlength}(N) > 1024$) prime integers are useful for public key cryptography. An interesting aspect of this particular graph is that it showed that within the pool of used primes, a 2 factor false count RSA Based Primality Test outputs largely false positives, but also occasionally false negatives, which is not the case for integers outside of that set of known primes used in the test itself.

Based on previous RSA Based Primality Test versions and tests, it is possible to use differing pool of primes, besides just the primes that exist between 2 and N. However, this technique is not as efficient, nor does it have the same properties as the tested version used in all of the benchmark tests.

Conclusion

This implementation of a primality test based on the RSA cryptosystem is unique. The versatility of the RSA Based Primality test as being both probabilistic and deterministic based upon the input of known primes used in the code. The accurate primality testing of large integers is critical for accurate encryption and subsequent decryption, as well as secure asymmetric cryptosystems, whose applications include identity verification, and secure message transmission. Two factor RSA Based Primality Test with count of false instances is the most accurate form of the RSA Based Primality test, although the other forms (multiprime RSA and count of true instances) can be used as well, although with a lesser degree of accuracy.

Future experimentation for this project will include coding the RSA Based Primality Test in C++, specifically Nvidia's CUDA, which runs the code on GPUs rather than CPUs, and is consequently much faster. Then additional benchmark computation time tests with large bitlength primes will need to be conducted with this updated version of the code, given that computation time is the only factor that would be affected by coding the test in CUDA.

References

[1] A Method for Obtaining Digital Signatures and Public-Key Cryptosystems.

<https://people.csail.mit.edu/rivest/Rsapaper.pdf>

[2] An Introduction to the Theory of Numbers.

<http://www.fuchs-braun.com/media/532896481f9c1c47ffff8077ffffff0.pdf>

[3] Asymmetric cryptography and practical security. David Pointcheval.

<https://pdfs.semanticscholar.org/f086/fc07648dd8277b0e14702718d3a86c8dc106.pdf>

[f](#)

[4] Diffie-Hellman: Key Exchange and Public Key Cryptosystems. Sivanagaswathi

Kallam. cs.indstate.edu/~skallam/doc.pdf

[5] Eulers Theorem. <http://sites.millersville.edu/bikenaga/number-theory/euler/euler.pdf>

[6] Great Internet Mersenne Prime Search. <https://www.mersenne.org/>

[7] "Primality Test". Wolfram Research inc.

<http://mathworld.wolfram.com/PrimalityTest.html>

[8] Primality Randomization. Shubham Sahai Srivastava.

https://www.cse.iitk.ac.in/users/ssahai/talks/primalty_randomization.pdf

[9] Primes in P. https://www.cse.iitk.ac.in/users/manindra/algebra/primalty_v6.pdf

[10] Python Implementation of Miller Rabin Primality Test.

<https://gist.github.com/Ayryx/5884790>

[11] Python Implementation of Modular Inverse for RSA.

<https://gist.github.com/ofaurax/6103869014c246f962ab30a513fb5b49>

[12] RSA Attacks. Al Rasheed, Abdul Aziz and Fatim.

<https://www.utc.edu/center-information-security-assurance/pdfs/course-paper-5600-rsa.pdf>

[13] The Miller-Rabin Randomized Primality Test.

<http://www.cs.cornell.edu/courses/cs4820/2010sp/handouts/MillerRabin.pdf>

[14] The Miller–Rabin Test.

<http://www.math.uconn.edu/~kconrad/blurbs/ugradnumthy/millerrabin.pdf>

[15] The Rabin-Miller Primality Test.

<http://home.sandiego.edu/~dhoffoss/teaching/cryptography/10-Rabin-Miller.pdf>

Acknowledgements

I would like to thank Alan Didier and the Los Alamos High School Computer Science Club for their help with computing power. I would also like to thank the wonderful online IDE, repl.it.

Infectivity Model

New Mexico Supercomputing Challenge

Final Report

April 4, 2018

Team Number: ATC-3

School Name: Academy for Technology and the Classics

Team Members:

Alex Baten

Ben Thorp

Teddy Gonzales

Teacher:

Jennifer Hooten

Project Mentor:

John Thorp

Executive Summary

Our project is a model of the spread of the Influenza Thread A between several hypothetical neighborhoods. By studying how individual conditions can affect the spread of a disease through several different communities, which were roughly organized using a map found on the City of Albuquerque website¹, we hope to suggest factors that will minimise the spread of Influenza Thread A. We can use our data to show if individual decisions produce a quicker spread through the community. This data can also help us understand, on a local level, how the disease spreads in different conditions. Our model considers countermeasures against disease, such as vaccines and contagion time periods. This model can also be used to compare two or more different variables and their impact on the spread of influenza.

We created a computer model in Python, to demonstrate the spread of Influenza between several communities. The model simulates the movement of agents between day and night congregation points. Each run of the simulation takes a number of parameters that were gathered using Zip Codes located in each district (Appendix A) including the following:

- The total number of humans in each district
- An estimate about the number of houses and rooms in each area
- The times when the disease starts and stops being contagious in communities with or without health care
- Percentage of vaccinated humans in the different communities

Last year we tested the model with a number of simulations that grew in complexity as the model matured. These simulations are presented below. The simulations demonstrate that

¹ *City Of Albuquerque-Find Your Counselor*

the model can predict, on a local level, how the disease spreads through different communities in different conditions. This year we included real-world data taken from the 2015 Community Survey taken by the Census Bureau (Appendix B). This meant that most the data would be fixed in terms of the amount of people and the age groups of those people.

Problem Statement

While discussing potential ideas for our project in our first year, we came across Google Trends. In our Computer Science class we read an article that discussed about how Google Trends could be used as a database for influenza infection rates by viewing the search results for “flu symptoms” over time. The problem is that Google Trends can only present data from the past. To address the factors that cause the Influenza to spread, we wanted a computer model that could simulate future situations and provide useful data. Expanding on the idea of simulating future situations, we decided to add variables, such as vaccines, to see how the initial conditions in a community would affect the spread of influenza. Of course our model is only a basic simulation, however, we have built on last year’s model to increase its complexity and develop a program that makes more accurate predictions by incorporating more real world data than we used last year.

This year we originally had the idea to incorporate many new parameters such as: wealth, population of a real-life city, and a formula that would help us calculate the probability of getting the disease. We originally had decided to use San Diego as the city that we would use as the model. However, as we worked on the program, we decided to use Albuquerque as the city that we would model. This is because Albuquerque is more local and applicable to where we live.

Solution Method

We planned to solve our problem by creating a model that incorporates given variables. We did this by using a basic open source agent-based Python/Mesa model in order to show the spread of the disease between different communities. We compared how the different

demographics affected the severity of the illness in the different communities. We would also like to explore what factors battle the infections (healthcare, vaccinations, and lifestyle).

Our simulation was created with the intent of understanding influenza so that it can be controlled and combatted. By manipulating different variables, we hope to discover which methods of disease control are the most effective.

The first year we struggled to make the program accurately simulate real life. The immense complexity of the world makes it hard to replicate, so we've performed research to create a generalized model. Its accuracy is limited, but it is useful for tracking trends.

This year we were able to create a more realistic model, that would still be able to detect trends. Although, compared to the real world, it is still a fairly basic model. It neglects socio-economic data, the strength of the immune system, a more complex look at interactions between people, mental health, and the environment that the agents are in

Validation and Verification of the Model

- Validation - Are you building the right thing?
- Verification - Are you building it right?

Validation

Influenza is a virus that infects many during the winter and early spring. With research we have found that the disease has a main “type”, Influenza A. This virus usually spreads once a year peaking in February. We have found that 5-20% of the United States will get the virus per year². From these infections about 1-4% of people will die, mainly due to pre-existing health

² webMD

conditions ^[2]. Due to these facts we decided to ignore the lethality of the disease, and focus on its spread (contagion).

The first year we used the nationwide average vaccine rate by age group³. In this group, 59.3% of children 0.5-17 years old have received a vaccine, 36.3% of adults 18-64 have received a vaccine, and 63.4% of people 65+ have received a vaccine. We incorporated this data into our program by giving 59.3% of children 0.5-17 years old a smaller chance of being infected, and by giving 36.3% of adults 18-64 a smaller chance of being infected. We excluded people 65 and up from the program. Since vaccinations play a major role in fighting Influenza, we decided to incorporate vaccines into our model. We have also researched and incorporated the amount of time you can pass the disease to other people for. This is important as it serves as a base for a more complex epidemiology model. From our research we have a hypothesis that people from neighborhoods without health care possibly have a higher chance of infection. This is a hypothesis that still requires more research to back up our inferences.

In the second year we managed to improve upon how the disease is passed from one community to another community. We used a formula created by a group that was also trying to simulate the spread of Influenza⁴. The formula, $P(\text{inf}) = 1 - \prod k \prod j \prod B C \psi_{Ajk} \exp \{-\lambda_v d_{iAjk} y_B\}$, when translated into simpler math it is just, $1 - \exp \{-\lambda_v d_a y_B\}$, with λ_v correlating to a value (Appendix C1) that is dependent on the age of agent whose chance is being calculated, agent A, and of the age of the agent who is coming into contact with agent A, agent B. The variable d_a correlates to the amount of time that agent A was in contact with agent B (Appendix C2). The variable y_b is a binary value, that is 0 if agent B is not sick, and 1 if agent B is sick. This is so that if agent B is not sick, it will not be inputted into the equation that calculates if agent A gets sick. There is one

³ Influenza (flu)

⁴ Effectiveness of Interventions to Reduce Contact Rates during a Simulated Influenza Pandemic

$\exp\{-\lambda_a d_a y_B\}$ for each agent that Agent A comes into contact with which are multiplied all together before being subtracted from one. The application and values that go into the formula for each age group of each district is in Appendix C2. If the agent is infected, it will take 4-16 times steps or 1-4 days after for the agent to display symptoms of Influenza⁵ it should be noted that the agent will be contagious during this time.

Verification

The entered data gave simulations that provided useful graphs that displayed the infection trends for the two communities. Trials 6 and 7 (shown below) provide a clear example. In Trial 6, the communities had about 10 people per room and the disease had a 40% ($L=0.4$) chance to infect others. In Trial 7 the communities also had about 10 people per room but the disease had a lower, 20% ($L=0.2$), chance to infect others.

Since it was easier to catch the disease in Trial 6, the infection sum shows a turning point later than the infection sum turning point in Trial 7 (Trial 6 shows a turning point of infections at about 22 days and Trial 7 shows a turning point of infections at about 45 days). Because the infection probability in Trial 7 was half of the probability of Trial 6, one would assume that the turning point in infectivity would have been later in Trial 6.

This is a useful simulation in that it shows the connection of the contagiousness parameter and the actual rate of spread of the disease in the model.

⁵ CDC

The graph above is the recorded counts of Influenza like Illness (ILI) from the 2015-2016 Flu season in New Mexico. Comparing our graphs to the actually recorded data some similarities should be noticed. One of which is both graphs overall shapes, which resemble a bell curve, even though our model was a hyper-realistic model that only ran for 25 “days”. Another similarity is despite the overall bell curve shape both graph had irregularities on a small scale. This is an improvement compared to last year when our graphs were notably more smooth throughout the whole simulation, with few regularities.

In the first 12 graphs from this year , we tested to see how the presence of a hospital would affect the spread of the disease. In the first six graphs it is shown that all communities regardless of if they had a hospital that they all had the progressive upward curve. What separated the communities with hospitals from the countries without hospitals graphically was that the countries without hospitals took longer to completely recover from the flu outbreak, no matter the population size. We then proceeded to do the opposite, and have all of the communities had a hospital except one community, and see how that impacts the epidemic. The results we received were that the communities with hospitals recovered much faster , as we had predicted. While the community without the hospital took longer to recover from the epidemic. This trend continues through the rest of our tests. This reaffirmed our belief that access to hospitals helps alleviate the epidemic and end it faster than if there was no hospitals or healthcare centers.

Conclusion and Achievements

In the first year we started with a basic model, a model with just rooms and the ability to catch the disease. Then after doing some research, we added a vaccine percentage, or the percentage of the population that is immune to the disease. We consulted our teacher, who suggested that we add a variable to describe access to health care, or hospitals, into the communities and a variable for the agent's desire for a healthy lifestyle. Although our project provides only basic simulations, it is useful for comparing trials that gives potentially challenging outcomes. Data that challenges preconceptions introduces new questions that prompts the maturing of our simulation and the introduction of variables that we may have thought to be unnecessary.

Our most significant achievement is the ability to take several variables and obtain a simulation. Each simulation is different and each provides different insight. The example provided in the "validation" shows that comparing two simulations allows us to decide which variable we should change next. If we find that the rate at which people come in contact is more of a factor than vaccines or contagiousness, then we can improve the conditions of the community by lowering the "room's density" and see if it does in fact improve conditions (lower number of infections).

We have successfully added an additional four communities, as well as an age variable and a public transportation variable. We found out, just as we had originally thought, that public transportation does in fact spread the disease significantly faster than if they humans traveled directly from their house to their school or work and vice versa. Age also played a large role in the simulation.

If we continue this project into the future, we intend to add weekends to the program. This will slow the spread of the disease in the school and work settings and on the bus, but allow it to grow at a higher concentration in the communities. Even though the percentage is small, we can also include mortality rates of the flu.

Selecting a language

We chose Python because we went on the Santa Fe Tech tour in Socorro during the 2016/2017 Supercomputing Computing Challenge and discovered that Python fit our needs. We are using Python 3.6.5 for our project. It creates a nice graph when all of the variables are filled. Python was more complicated than was required, but it still performed well.

Choosing Mesa for agents

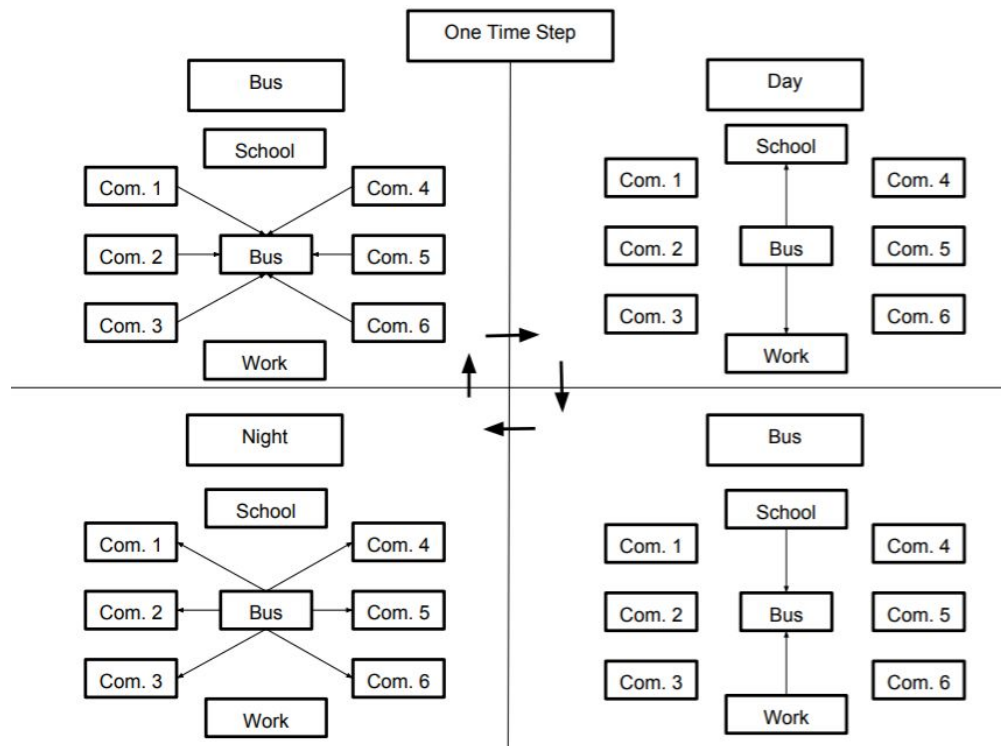
Our program requires us to make a model about the spread of disease. Writing models from scratch in Python is quite difficult, so we decided to look online for premade libraries to make the job easier. What we found was Mesa, an open-source framework⁷.

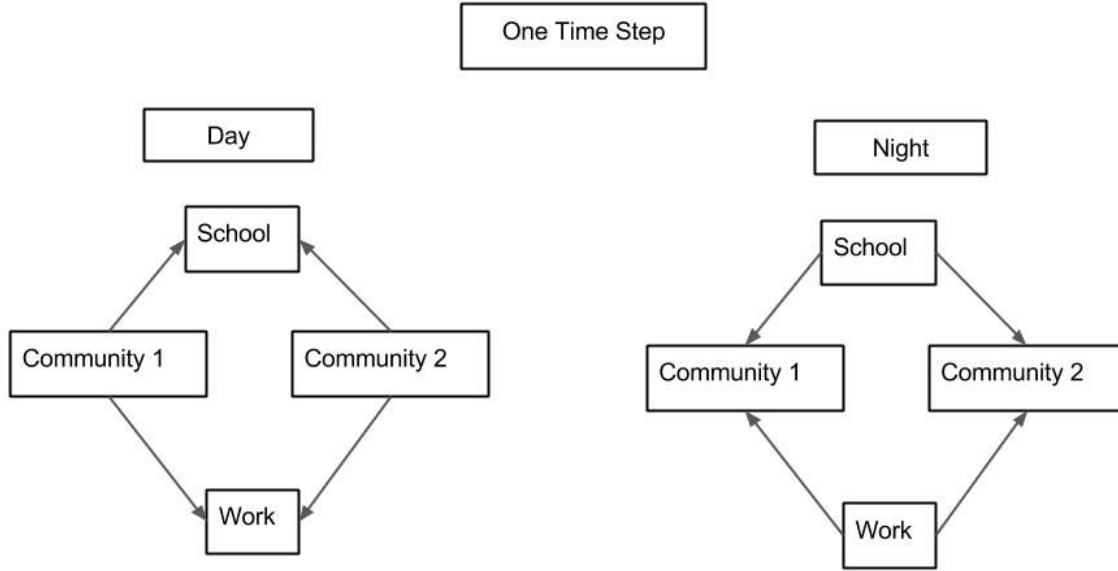
Mesa was first released on June 21, 2015 by the programmers at George Mason University in Virginia. We used a video and a web page to learn how to use it. It is an open-source basic agent-based modeling Framework. Mesa simplifies tasks and allows the user to analyse and visualize their work. Mesa has built-in support for the agents, like spatial grids and agent scheduling.

⁷ http://mesa.readthedocs.io/en/latest/tutorials/intro_tutorial.html

Day and Night for agents

In our model, humans travel between their nighttime communities, their daytime school or work and a bus to and from their community or daytime activity. All humans take the bus to and from their assigned community and their school or work. The communities are divided into households and infected human can infect others within their household. The school and work are divided into rooms, where inhabitants can infect each other. The buses are divided into individual bus routes, when inhabitants can infect each other. A time step is comprised of a day, a bus route, a night, and a second bus route.





The parameters we used are number of humans, number of rooms, infectivity likelihood, contagion start, contagion end, and the percentage of vaccinated humans.

The Big Data Structure

We hold all the agent's placement in a matrix. The height (number of rows) is based on the number of houses or rooms in each community. There are four columns that correspond to the day and night placement of the agents. Each agent has its own concept of its location during the day or night, and will move around the matrix in accordance to the timestep cycle.

| Bus | Com. 1 | Com. 2 | Com. 3 | Com. 4 | Com. 5 | Com. 6 | School | Work |
|---------|---------|---------|---------|---------|---------|---------|-------------|----------|
| Route 1 | House 1 | House 1 | House 1 | House 1 | House 1 | House 1 | Classroom 1 | Office 1 |

| | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|-----------------|----------|
| Route 2 | House 2 | House 2 | House 2 | House 2 | House 2 | House 2 | Classro om 2 | Office 2 |
| Route 3 | House 3 | House 3 | House 3 | House 3 | House 3 | House 3 | Classro om 3 | Office 3 |
| Route 4 | House 4 | House 4 | House 4 | House 4 | House 4 | House 4 | Classro om 4 | Office 4 |
| Route 5 | House 5 | House 5 | House 5 | House 5 | House 5 | House 5 | Classro om 5 | Office 5 |

Code

```
# Project 2017-2018 ATC - Flu Transmission
# Built on Mesa Agent Library from George Mason Univ.
# Written by Ben Thorp
# ATC-3 Ben Thorp, Alex Baten, Teddy Gonzales
# Version 1.1
# Added start and stop of contagiousness
# Version 1.2
# Added HealthCareAccess and HealthLifeStyle
# Version 1.3
# Added community3 and buses
# Version 1.4
# Added community4, community5, community6, and age groups
```

```
from mesa import Agent, Model
from mesa.time import RandomActivation
import random
from mesa.space import MultiGrid
from mesa.datacollection import DataCollector
```

Sources:

[1]

City Of Albuquerque-Find Your Councilor , www.cabq.gov/council/find-your-councilor.

[2]

“WebMD.” *WebMD*, WebMD, www.webmd.com/cold-and-flu/flu-statistics. Accessed 2 Apr. 2017.

[3]

“Influenza (Flu).” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 1 Nov. 2017, www.cdc.gov/flu/fluview/coverage-1516estimates.htm.

[4]

Haber, M. J., Shay, D. K., Davis, X. M., Patel, R., Jin, X., Weintraub, E....Thompson, W. W. (2007). Effectiveness of Interventions to Reduce Contact Rates during a Simulated Influenza Pandemic. *Emerging Infectious Diseases*, 13(4), 581-589. <https://dx.doi.org/10.3201/eid1304.060828>.

[5]

“How Flu Spreads.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 12 Sept. 2013, www.cdc.gov/flu/about/disease/spread.htm. Accessed 2 Apr. 2017.

[6]

CDC GIS FluView. (n.d.). Retrieved April 04, 2018, from <https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>

[7]

Introductory Tutorial¶. (n.d.). Retrieved April 04, 2018, from http://mesa.readthedocs.io/en/latest/tutorials/intro_tutorial.html

[8]

HHS Office of the Secretary, & National Vaccine Program Office. (2017, November 15). Flu Vaccination Claims Rates by Geographic Area. Retrieved April 04, 2018, from

<https://www.hhs.gov/nvpo/about/resources/interactive-mapping-tool-tracking-flu-vaccinations/index.html>

[9]

https://www.census.gov/content/dam/Census/programs-surveys/acs/guidance/training-presentations/20170419_MOE.pdf

Acknowledgements

A large thank you to the Los Alamos National Lab and the judges for making the Challenge possible. We had a great time and spent many hours learning new information. We would also like to thank the George Mason University for creating the mesa library for Python. We are very grateful for the time spent with our teacher and the mentor, Jennifer Hooten and John Thorp, respectively.

A-Maze-Ing

New Mexico

Supercomputing Challenge

Final Report

JMS32

Jackson Middle School

Team Members

Brendan Kuncel

Quentin Dye

Teacher

Karen Glennon

Project Menor

Patty Meyer

Executive Summary

The first ever recorded labyrinth was found in Egypt in the 5th century, and it was discovered by Herodotus.¹ Unfortunately this is not a true maze, it was a labyrinth, as hedge mazes only started to occur in Belgium 13 century. The word “Maze” was also first seen in the 13th century, and it meant the feeling of delusion or bewilderment, and that connects to the idea of “Amazement” or to “Amaze.”¹ So Mazes have bewildered many people tracing back all the way to the 5th century, and surely there are quite a few ways to solve one. So what is the best way to solve one, there are thousands and thousands of algorithms that can solve a maze, but which one can solve them the fastest and with the most success. The question could be applied in a couple of different situations like navigating city streets. Before we got into that question we needed to do a little bit of research, so we first explored the types of mazes that we could potentially solve. After doing that research we found that there are quite a few type of mazes that we could use in our project. One of which being regular two dimensional mazes, that could be drawn out onto a piece of paper, and these types of mazes can be as complex or a simple as you want. The Second type of maze we looked at was labyrinths, and they are defined as a complex network of passages that are difficult to traverse.² Although that sounds a lot like a maze labyrinths and mazes do have a large difference, as mazes usually include complex branching, and labyrinths being more unicursal, only having a single path to follow with one entrance and one exit. The other classes of mazes get even more complicated like three dimensional mazes, as three dimensional mazes have 6 different directions of travel. The other important thing to research is how we are going to solve the mazes, which include things like the wallflower method, the flood method and Dijkstra algorithm, each one having a strength and weakness.

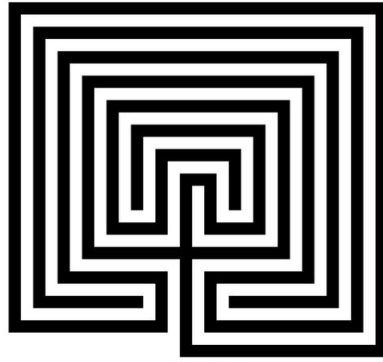
Wallflower being useful for regular two dimensional mazes that don't have any loops, as loops are one of the weaknesses that wallflower method have. Dijkstra method thrives in situations regarding loops, so this Algorithm would work quite well for mazes that include loops. So our project aims to explore the options you have in a situation where you have to navigate your way around a city without any map or any external help.

Statement of the Problem

Mazes are a quite complicated and interesting, as they can be as complicated as the program that runs your phone or as simple as a hello world program built in python. So how do we solve them? Well first we should look at what kinds of mazes we are dealing with or what maze you want to solve, as there are regular two dimensional mazes, labyrinths and three dimensional mazes. Then you need to know how you can solve the maze, your first thought may be the left hand method or a completely random method. The question is what is the most effective method you could have. Our project aims to answer the question of what method is the most effective or what can solve the maze the fastest. We decided that we would stick to the traditional maze and a city block type maze. We decided that we would try and choose 3 to 4 different methods to test which would be the fastest or the most effective. Those being a Flood method, Wallflower method, and the Dijkstra Method.



MAZE



LABYRINTH

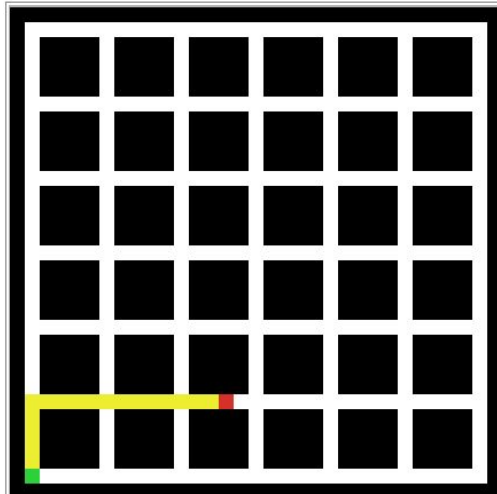


3D maze

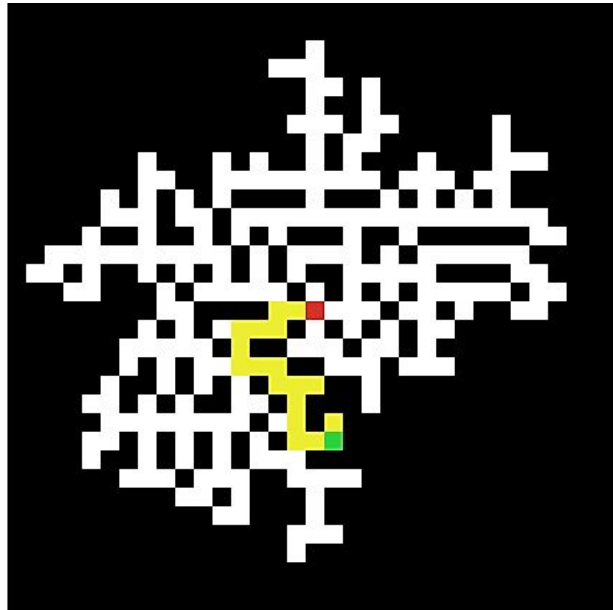
Method

The method we used was to implement a variety of maze solving and maze creating algorithms. We took some of the most common ways to solve a maze, following one wall, or just picking the path closest to the end of the maze, and put them into netlogo with two different maze types, one that was random every time, and a grid, representing a city. We used the netlogo behaviour space to run many repetitions of the different types of solutions with the different types of mazes. We then took the information and sorted it by the fastest average solver (fastest meaning in the lowest number of steps taken to the end of the maze). The steps taken are monitored by the patches themselves, as they have a variable that is assigned whenever an agent steps on them. This helped demonstrate the steps taken that are more than what is necessary for the design, which could be improved in the future by having a way to follow the right or left hand rule while ruling out the places already visited and straight dead ends (the places that you can see that it is a dead end at the start of the passage). The code used for the right hand method just checks to see if there is path next to it from right to left, and the opposite for the left hand

method. The facing strategy was based on facing the exit and turning either clockwise or counter clockwise until a turn is passed. That path is then taken until another turn is presented.



“Grid” maze



“Random” maze

Verified and Validated

Our model is representative of real world situations. It can be used in helping to find the shortest route to a place in a maze or a city. This was verified by looking at the most common ways to solve real-world mazes. This inquiry found following walls, using a map, and just going in the closest direction to the maze. While most would not consider using a map “solving” a maze, it is still a method used, and one of the fastest ways. We included this method by having both a computer solve the maze and have the other methods show the optimal route that they took (which still could not be the best route, if they go all the way around and reach the end, the

“best route” could actually be one of the longest.) As the way a person will solve a maze can be anywhere from knowing for sure to randomly running around, it is good to have a method to follow to help not be lost and confused.



Downtown Denver:

Finding your way through a modern city can sometimes be difficult and finding the best way to navigate can be quite beneficial.

Results

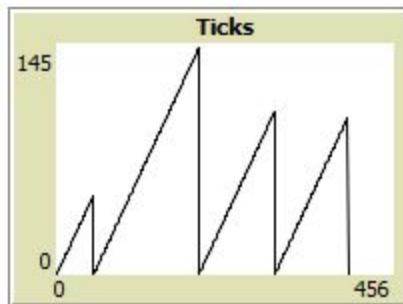
The results that we gathered, were that in a grid formation (aka what would be used in the case of a map) the best way to reach the end quickly would be to follow the method of going in the direction closest to that of the direction to the end of the maze. This did not, however hold true with a randomly generated maze, where facing the end would get there only on certain occasions. On all of the other trials, facing the end would get the agent stuck, going into a dead end, and no trying to get out, but just continually going back to the dead end. In this type of maze, the only method that would solve it 100% of the time was either the right hand or left hand method. These, however, could take very few steps, or almost every possible step, depending on if there were several branches in the best path from start to finish. The best path, however, is

shown in a fashion very similar to the “flood” solution (where the maze is solved by the computer by “flooding” the maze till an end is reached) by taking the path with the least steps and highlighting it, or giving a sort of “map” to the finish

Conclusion

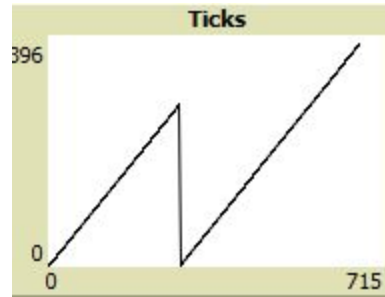
Mazes are complex problems that are just waiting to be solved, but which method is the best, well this is what our project aims to answer. The initial mazes we were looking at was random mazes, and this took us some time to get down, as the program to build a random maze is not that simple, and took some time to program. Then we have to look at what methods we are going to use to solve these mazes, and the first thought we had was the left and right hand methods or what is called the Wallflower method, but we needed to attempt to go a little deeper in terms of the solving algorithms. Then we looked into Dijkstra’s algorithm, this unfortunately did not play out, as we had trouble with moving turtles from node to node, but it works by finding the node with the shortest distance from itself, then looking for the next shortest distance keeping track of what it has already gone too and staying away from those spots. So as of now, we were only able to find that the flood method worked the best, and this makes sense as it does the entire maze and looks at every possibility and decides what was the quickest and shortest path to the finish. Unfortunately we were hoping to find a method that humans could use to find there way to the end of the maze. So we intent to continue this project and implement some more complex and creative mazes algorithms and a link based maze generator, as we want to further explore the Dijkstra Algorithm.

Data



Picture 1:

The number of steps (y-axis) for each method. From left to right, right-hand, left-hand, face the end always going clockwise when unsure, and facing the end always going counter clockwise when unsure.



Picture 2:

The number of steps in a randomly generated maze, this graph only shows the right and left hand methods, respectively. The other two methods are not shown, as they were unable to complete the maze, and made the rest of the data unviewable..

Most Significant Achievement

Brendan Kuncel- My most significant achievement on our project was the type of coding I had to try to understand and write. The whole point of the project is to have several different methods to perform the same action, so it was quite a strain to come up with the code for several ways to solve a maze.

Quentin Dye- My most significant achievement on our project was diving deeper into research, as this was quite an important part of our project. I found a couple of different methods to solving mazes which helped us to look at what method is the most efficient, and we found that

some are better than others but some of that also depends on the type of maze we have. The research done helped to drive our project in more focused direction, and I think I did a great job looking into and analysing what we found for our project.

Acknowledgements

Karen Glennon - Ms.Glennon was a big help in a couple of things throughout our project, as she always read all of our papers, and checked for spelling and grammar. She also looked at our program to make sure that it made sense to her as other people are going to need to understand it other than ourselves. Ms.Glennon was a great help in making sure that our work was presentable, and for that we thank her, and acknowledge her help in our project.

Patty Meyer - Ms.Patty helped in quite a few things as well, the first and most important would be that she would help us talk through an idea, and help us to develop our own Ideas. This was big help to us as our idea was quite unclear in the beginning as first we started with a project about entropy then moved towards something we actually liked, Ms.Patty helped us to come to the idea of mazes and what the most efficient way to solve them is. The second thing would be helping us in our research, as she suggested that we look into a couple of algorithms to test in our simulation. Ms.Patty was always there to help us bounce our ideas off of her, and helped us to develop those ideas, and for that we thank her, and acknowledge her help in our project.

References

Pullen, W. D. (n.d.). Maze Algorithms. Retrieved November 27, 2017, from <http://www.astrolog.org/labyrnth/algrithm.htm>

“Cris' Image Analysis Blog.” *Cris Image Analysis Blog RSS*, www.crisluengo.net/index.php/archives/277.

“Maze Generation.” *Maze Generation - Rosetta Code*, www.rosettacode.org/wiki/Maze_generation

“Programming Theory: Solve a Maze.” *Algorithm - Programming Theory: Solve a Maze - Stack Overflow*, www.stackoverflow.com/questions/3097556/programming-theory-solve-a-maze

“Maze Solving Algorithm.” *Wikipedia*, Wikimedia Foundation, 17 Feb. 2018, en.wikipedia.org/wiki/Maze_solving_algorithm.

Rathor, Shantur, et al. “Maze Solving Algorithms.” pp. 1–15.

[1] Krystek, Lee. “Amazing Mazes”. The UnMuseum. 2001

[2] Oxford Dictionary. “Labyrinth”

Implementing an Artificial Neural Network to Recognize Handwritten Digits

New Mexico
Supercomputing Challenge
Final Report
April 4, 2018

Team Number 201
Los Alamos High School

Team Members:
Sam Crooks

Project Mentor:
Jacob Stinnett

Table of Contents

| | |
|--|-----------|
| Executive Summary..... | 3 |
| 1. Introduction..... | 4 |
| 2. Background Research..... | 4 |
| 3. Problem Description..... | 5 |
| 4. Objective..... | 3 |
| 5. Algorithm Design..... | 6 |
| 5.1 Data Processing..... | 6 |
| 5.2 Neural Network Structure..... | 7 |
| 5.3 Training..... | 8 |
| 6. Performance Analysis and Optimization..... | 10 |
| 6.1 Accuracy and Error Graphs..... | 10 |
| 6.2 Optimization..... | 11 |
| 6.3 Applying the Model to the Real World..... | 12 |
| 7. Conclusion..... | 13 |
| 8. Acknowledgements..... | 13 |
| 9. References..... | 14 |
| 10. Source Code..... | 15 |

Executive Summary

Computer vision is a rapidly developing field of computer science which aims to autonomously identify objects from digital images and videos. One widely used approach in computer vision is an artificial neural network, which was inspired by the structure of biological neurons in the brain. The objective of this project was to implement an artificial neural network algorithm that could learn to accurately classify handwritten digits.

The algorithm was programmed in Anaconda Python and includes the following stages: data processing, defining the neural network structure, training to optimize the weight and bias matrices, and performance analysis. During data processing, each 28 by 28 pixelated digit is stored as a CSV file and is converted to an input vector. Then, 42,000 labeled handwritten digits from the MNIST dataset are divided into 30,000 training examples and 12,000 test examples. Next, the parameters of the neural network (weights and biases) are initialized. During the training stage, the parameters of the neural network update by gradient descent, where the goal is to minimize a cost function.

After 100 training iterations, the best performing structure correctly classified 97.4% of the test dataset. To evaluate the performance, accuracy and cost were graphed as a function of iterations.

Further work includes varying the number of hidden layers in the neural network, implementing more advanced machine learning techniques, and training the algorithm to recognize other symbols and objects. Applications of this technology include reading addresses on postage, processing images of checks, and autonomous vehicles.

1. Introduction

Computer vision is a rapidly developing field of computer science which aims to autonomously identify objects from digital images and videos. When humans look at symbols and objects, their brains effortlessly recognize them. Similar to the way cone cells in human eyes are used to recognize objects, computer vision systems view objects as pixelated images. One specific task within computer vision is to recognize and classify handwritten digits. In this project, a machine learning technique called an artificial neural network is implemented to recognize handwritten digits.

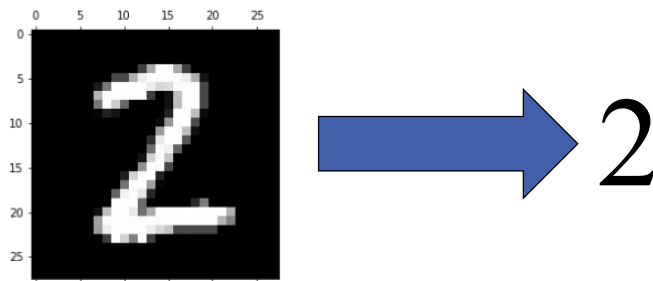


Figure 1: Pixelated image is recognized as the number two

2. Background Research

Simple logical rules cannot accurately classify handwritten digits. Rather than explicitly programming rules, machine learning techniques provide systems with the ability to learn and make accurate predictions on their own. The type of machine learning technique implemented in this project is an artificial neural network. An artificial neural network is a model inspired by the structure of biological neurons in the brain [3]. Neural networks are a highly researched and developed field, and there are a variety of neural network structures.

A feed-forward, fully-connected artificial neural network consists of a number of artificial neurons arranged into layers. The first layer (the input layer) receives input from the real world, and the last layer is the calculated predictions. Each middle (hidden) layer processes the output of the previous layer. Each neuron is connected to each other neuron in neighboring layers by a weight. By adjusting the weights and biases of the network to minimize a cost function, the neural network learns to classify images.

3. Problem Description

One specific type of computer image recognition involves recognizing human handwriting. It can be challenging for a computer algorithm to recognize handwritten digits because not only can the specific pixel values of one digit greatly differ across images of the same digit, but images of different digits may also look very similar. In order to implement a machine learning algorithm for digit recognition, the algorithm must learn from numerous labeled training examples.

4. Objective

The objective of this project is to implement an artificial neural network algorithm that can learn to accurately classify handwritten digits (zero through nine) from the MNIST dataset.

5. Algorithm Design

5.1 Data Processing

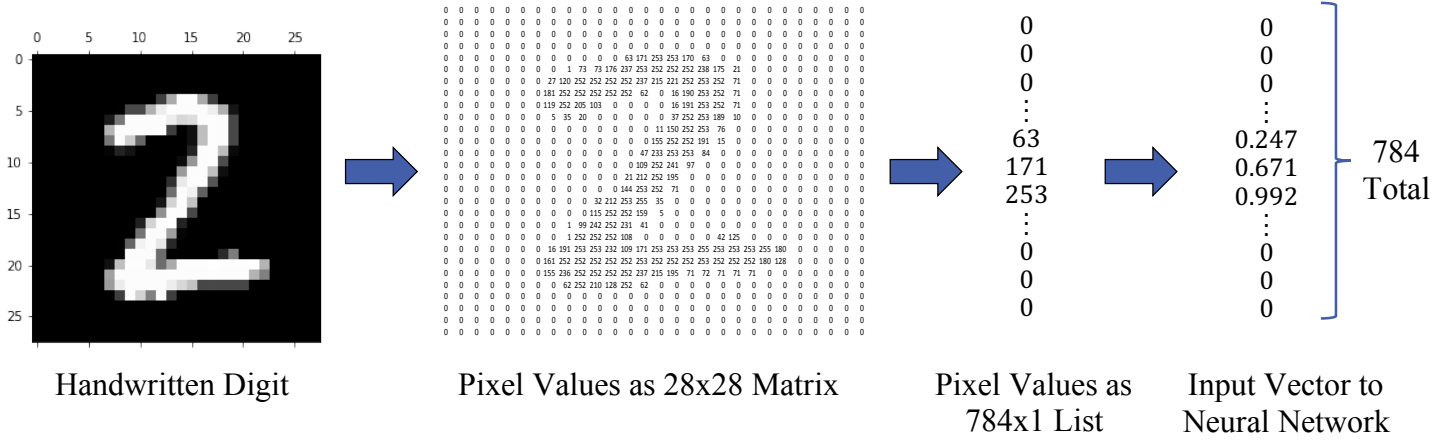


Figure 2: Data Processing Flowchart

A version of the MNIST (Modified National Institute of Standards and Technology) dataset contains 42,000 labeled digits stored as a CSV file [2]. Each digit is 28 by 28 pixels, for a total of 784 pixels. Each row stores each digit's data. The first column contains a label, and the remaining 784 columns are the pixel values corresponding to the digit. Each pixel value ranges from 0 (black) to 255 (white).

In this stage, images are converted into data that the computer can interpret. First, each digit image is converted to a 28 by 28 matrix. Second, all of the pixel values are put into a list. Third, each pixel value is divided by 255 to normalize the list to values between 0 and 1. Figure 2 demonstrates this process.

The 42,000 labeled digits were divided into 30,000 training examples and 12,000 testing examples. The training examples are used to develop the predictive model, and the testing examples are used to evaluate the performance of the model.

5.2 Neural Network Structure

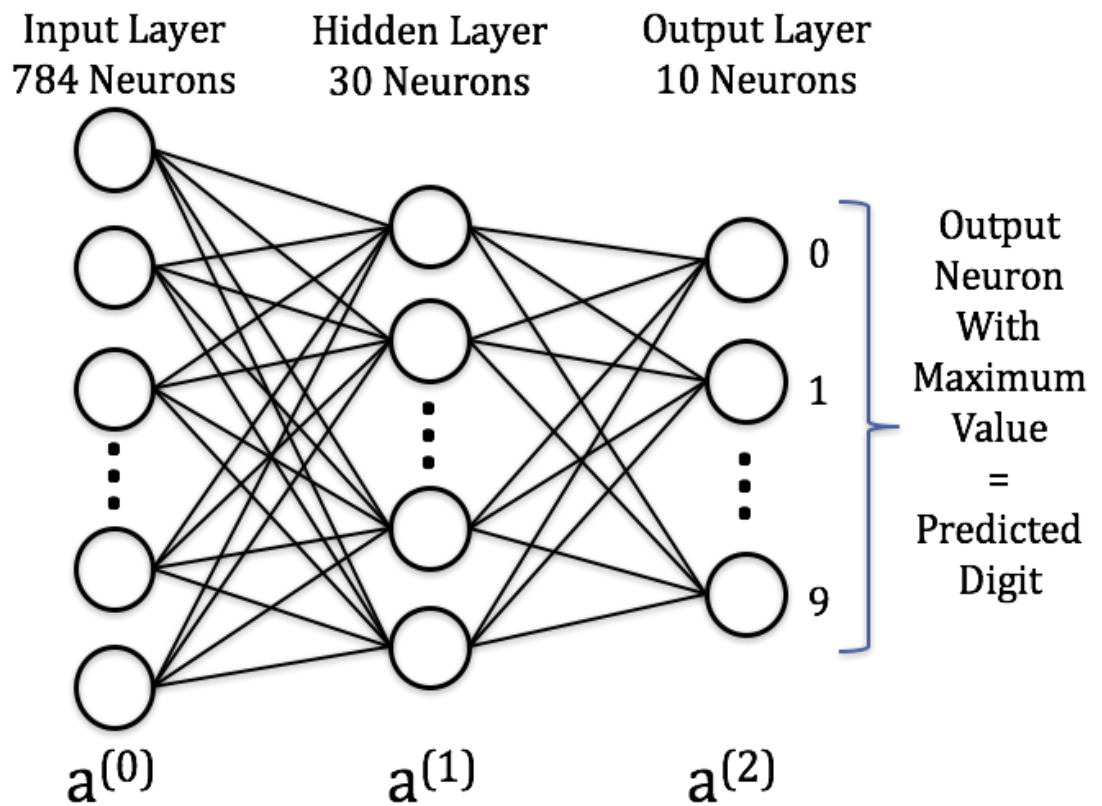


Figure 3: Neural Network Diagram

This neural network structure consists of three layers. Each neuron in the network is a function that outputs a value between 0 and 1 depending on the outputs of all the neurons in the previous layer. The first layer outputs the data of the digits after processing and the last layer's output indicates a confidence that the given image corresponds to each digit 0 through 9. There

are a total of 40 biases and 23,820 weights in this network that can be adjusted. Each weight is initialized to a random number between -0.01 and 0.01. Each bias is initialized to 0.

5.3 Training

A single training iteration exposes the neural network to the entire training dataset, and this consists of forward propagation and backward propagation. Training is done through a process called gradient descent. During forward propagation, each digit input vector in the training dataset is sent through the neural network and predictions are made. During backward propagation, the weights and biases of the neural network adjust to minimize error.

Forward Propagation:

During forward propagation, each of the digit input vectors is sent through the neural network. First, each neuron receives a weighted sum of the output from the previous layer. Then, each neuron activates depending on its input; the sigmoid function outputs a value between 0 and 1. These operations are performed with matrices.

$$\begin{aligned}
 & \mathbf{a}^{(0)} = \text{Input Vector to Neural Network} \\
 & \mathbf{z}^{(1)} = \mathbf{a}^{(0)} \cdot \mathbf{w}^{(1)} + \mathbf{b}^{(1)} \\
 & \mathbf{a}^{(1)} = \text{sigmoid}(\mathbf{z}^{(1)}) \\
 & \mathbf{z}^{(2)} = \mathbf{a}^{(1)} \cdot \mathbf{w}^{(2)} + \mathbf{b}^{(2)} \\
 & \mathbf{a}^{(2)} = \text{sigmoid}(\mathbf{z}^{(2)})
 \end{aligned}
 \quad
 \begin{matrix}
 \text{S} \\
 \text{i} \\
 \text{g} \\
 \text{m} \\
 \text{O} \\
 \text{i} \\
 \text{d}
 \end{matrix}
 \left(
 \begin{bmatrix}
 w_{(0,0)} & w_{(0,1)} & \dots & w_{(0,783)} \\
 w_{(1,0)} & w_{(1,1)} & \dots & w_{(1,783)} \\
 \dots & \dots & \dots & \dots \\
 w_{(269,0)} & w_{(269,1)} & \dots & w_{(269,783)}
 \end{bmatrix}
 \begin{bmatrix}
 a_0^{(0)} \\
 a_1^{(0)} \\
 \dots \\
 a_{783}^{(0)}
 \end{bmatrix}
 +
 \begin{bmatrix}
 b_0 \\
 b_1 \\
 \dots \\
 b_{30}
 \end{bmatrix}
 \right)$$

Backward Propagation:

The goal of backward propagation is to minimize the error function. The error function used is cross-entropy cost [4]:

$$\text{Cost} = -\frac{1}{M} \sum_{i=1}^M \left(y_i^T \cdot \log(a_i^{(2)}) + (1 - y_i)^T \cdot \log(1 - a_i^{(2)}) \right)$$

M is the number of training examples

y is a 10 by 1 target matrix that corresponds to the label of a digit (the expectation as to what the neural network should output)

$a^{(2)}$ is a 10 by 1 matrix that contains the activations of the output layer (confidences of a digit)

To minimize the cost function, the algorithm calculates the derivatives of the cost with respect to each adjustable parameter in the model: w_2, b_2, w_1, b_1 using chain rule.

Ex:

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial a_2} \cdot \frac{\partial a_2}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_2}$$

Then, those parameters are adjusted in the direction to minimize the cost.

Ex:

$$w_2 = w_2 - \alpha \cdot \frac{\partial L}{\partial w_2}$$

w_2 contains adjustable parameters in the model.

α is the learning rate and determines the leap size during gradient descent.

$\frac{\partial L}{\partial w_2}$ is the derivative of the cost function with respect to w_2

6. Performance Analysis and Optimization

6.1 Accuracy and Error Graphs

To evaluate the performance, the accuracy on the training set, the accuracy on the testing dataset, and the cost were graphed as a function of training iterations.

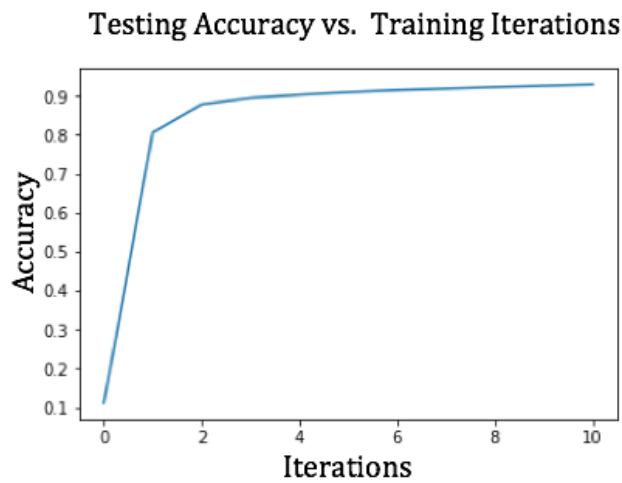


Figure 4: Plot of testing accuracy after 10 training iterations

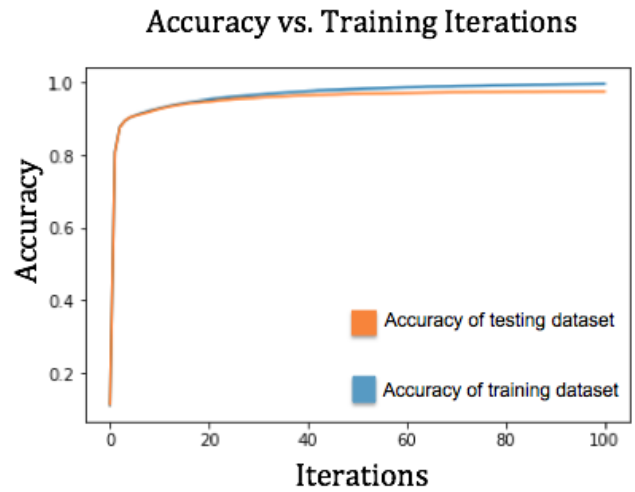


Figure 5: Plot of testing and training accuracy after 100 training iterations

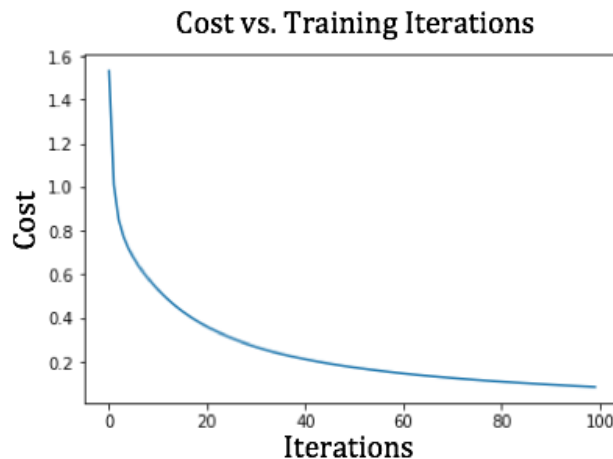


Figure 6: Plot of cost (error) per digit after 100 training iterations

The shape of the graph in Figure 6 decreases quickly during the initial iterations and gradually flattens out. The shape of the graphs in Figures 4 and 5 increase quickly during the

initial iterations and gradually flatten out. As the parameters update to reach a minimum of error, the derivative of the cost function with respect to each parameter was expected to be, on average, flatter. Thus, the neural network was expected to learn at a gradually decreasing rate. These results demonstrate the neural network is effectively being trained on the dataset.

6.2 Optimization

To optimize the neural network, the number of hidden neurons was varied in the neural network.

| Hidden Neurons | 1 | 10 | 30 | 90 | 270 | 810 |
|----------------|-------|-------|-------|-------|-------|-------|
| Accuracy | 11.3% | 90.5% | 95.6% | 97.1% | 97.4% | 97.3% |

Figure 7: Accuracy on test dataset after 100 training iterations for varying number of hidden neurons.

| Method | Test Accuracy |
|---|---------------|
| Mine: 2 layers, 270 hidden neurons | 97.4% |
| Linear classifier (1-layer NN) | 88.0% |
| K-nearest-neighbors, Euclidean (L2) | 95.0% |
| 3-layer NN, 300+100 hidden neurons | 97.0% |
| Committee of 35 conv. net, 1-20-P-40-P-150-10 [elastic distortions] | 99.8% |

Figure 8: Comparison to other published methods [1]

The best performing structure was the neural network with 270 hidden neurons. After 100 training iterations, it correctly classified 97.4% of the test dataset. A possible explanation for why the accuracy of the neural network structure with 810 hidden neurons was lower than the

that of 270 hidden neurons could be due to overfitting. Overfitting occurs when a model learns the patterns and noise in training data to the extent that it negatively impacts accuracy on data it has never seen.

6.3 Applying the Model to the Real World

To test the model on my own handwritten digits, the following steps were conducted. First, a digit was handwritten on a piece of paper using a pen. Second, a PNG image file of the digit was scanned by creating a signature in the desktop application Preview. Third, code was written to load the digit into the program and process from an image file to an input vector. Fourth, the input vector was sent through the neural network (forward propagation) and it produced a prediction. The figures below display the handwritten digit after scanning and the neural network's confidences given that digit.

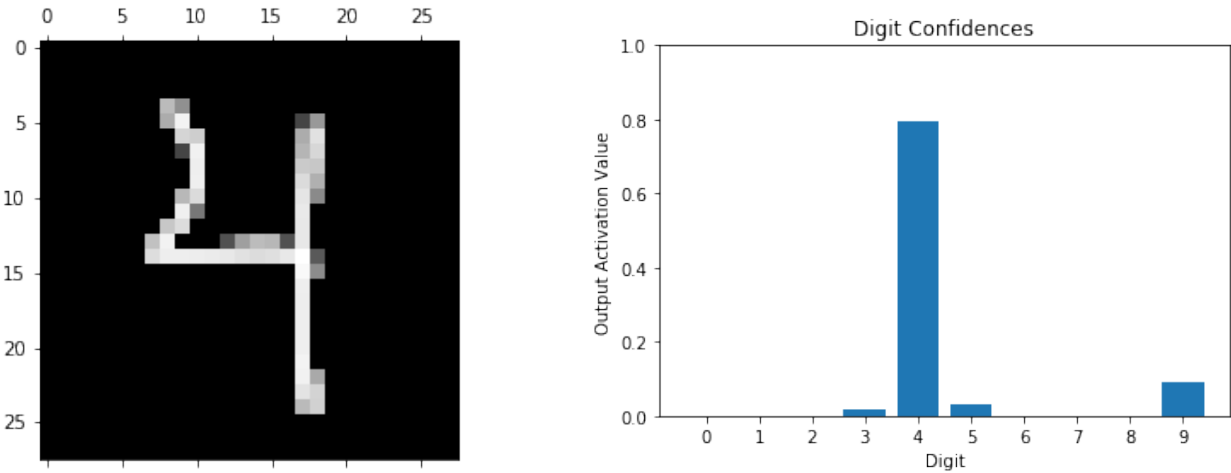


Figure 9: Scanned handwritten digit and output layer activation (digit confidences) bar plot from model

7. Conclusion

During this project, a neural network was implemented that could accurately classify digits from the MNIST dataset. This algorithm was programmed in Anaconda Python without any neural network libraries. Future projects could possibly improve the accuracy of this neural network by varying the number of hidden layers or by implementing more advanced techniques (regularization, deep learning). In addition, future projects could train the model using different or more complex datasets containing letters, symbols, shapes, etc. Even further, applications of this technique include reading addresses on postage, processing images of checks, and autonomous vehicles.

8. Acknowledgements

I appreciate the Los Alamos Mesa Public Library for providing quiet space with fast internet. I thank my mentor, Jacob Stinnett, for offering great guidance, helping me debug code (even late at night), and teaching me calculus concepts that I needed to understand for my project.

9. References

- [1] Y. LeCun, C. Cortes, and C. Burges, “MNIST Handwritten Digit Database,” [Online]. Available at: <http://yann.lecun.com/exdb/mnist/>. [Accessed 14 Jan. 2018].
- [2] “Digit Recognizer,” Kaggle, 2013. [Online]. Available at: <https://www.kaggle.com/c/digit-recognizer>. [Accessed 14 Jan. 2018].
- [3] XeronStack, “Overview of Artificial Neural Networks and its Applications,” Medium, July 16, 2017. [Online]. Available at: <https://medium.com/@xenonstack/overview-of-artificial-neural-networks-and-its-applications-2525c1adff7>. [Accessed 14 Jan. 2018].
- [4] Michael A. Nielsen, “Neural Networks and Deep Learning”, Determination Press, 2015

Spread of Dicamba Through Resistant and Not Resistant Crops

New Mexico
SuperComputing Challenge
Final Report
March 23, 2018

MDS185
Monte del Sol Charter School

Team Members:

Robert Borrego

Lia Fukuda

Stephen Herrera

Teachers:

Rhonda Crespo

Project Mentors:

Susan Gibbs

Paige Prescott

Table of Contents

| | |
|---------------------------------|----|
| 1. Executive Summary | 3 |
| 2. Introduction | 4 |
| 3. Description | 6 |
| a. Figure 1.1..... | 7 |
| b. Figure 1.2..... | 8 |
| c. Figure 1.3..... | 9 |
| 4. Results | 11 |
| a. Table | |
| 1..... | 12 |
| b. Table | |
| 2..... | 13 |
| c. Graph 1..... | 13 |
| 5. Conclusion | 14 |
| 6. Recommendations | 14 |
| 7. Acknowledgments | 14 |
| 8. References | 15 |
| 9. Appendix A: Model Code | 16 |

Executive Summary:

Dicamba is a broad spectrum herbicide that was developed in 1967 to kill stubborn weeds in large industrial farms. It has affected over 4% of the United States soybean population due to the new standard application, the “over the top” application when the herbicide is sprayed on the crops after sprouting to 6 inches in height[1]. In the last three years, Monsanto Company has created genetically-modified resistant crops to combat the destructive nature of this herbicide[2]. Growers using Dicamba have been given specific instructions on the over the top application of Dicamba but failed to follow the recommendations resulting in a large spread across the United States. When applied in high wind and/or before 7:30 a.m. or after 5:30 p.m., Dicamba drifts over multiple acres, causing nearby non resistant crops to be affected.

To begin, the team investigated the effects of the over the top application compared to the previous method of spraying the crops before they sprouted. This led to the conclusion that the over the top method produced more vaporized dicamba and resulted in a larger area affected by the herbicide. Because this method is now the standard way to apply Dicamba, the team chose to follow this method and explore what can be done to prevent Dicamba drift to non-resistant crops.

When searching for ways to reduce the diffusion of Dicamba through non-resistant crops, the team came across a diagram of wind speed reductions due to tree breaks[5]. This prompted us to create an experiment modeling the difference in crops infected with and without tree breaks.

Our model showed significant difference when tree breaks are used to slow the spread of Dicamba. The tree breaks diminished the number of infected trees from an average of 348 crops without tree breaks to 80 crops with tree breaks. We collected data from 10 different concentrations of Dicamba and every one had a decrease in infected crops with tree breaks.

Spread of Dicamba Through Resistant and Not Resistant Crops

Introduction:

First registered as a broad spectrum herbicide in 1967, dicamba has been used throughout the years in various weed killers, it's currently found in about 1,100 herbicide products [4]. Dicamba only attacks broad-leafed plants and kills them by mimicking the plant's growth hormones, causing growth abnormalities and eventual death [4].

Recently growers have been facing a war on weeds due to many becoming naturally resistant to the common herbicide called glyphosate (a popular herbicide used by planters and the active ingredient in Roundup) [4]. These "super weeds" have infested over 90 million acres of farmland, leaving many agricultural businesses in a race to find a new herbicide. Companies including Monsanto, BASF, and Dow-Dupont developed genetically modified seeds for soybeans and cotton that possessed a dicamba tolerance [1]. Starting in 2015, the Monsanto

Company, responsible for much of the crop industry, came out with new genetically modified cotton and soybean plant. Under the name Xtend, these plants are resistant to the killing effects of dicamba [1].

Due to dicamba becoming easily vaporized into the air, causing it to become susceptible to drift, it was only permitted to be applied to the soil prior to planting to kill off weeds [2]. A new type of dicamba with the promises of being drift resistant wasn't yet approved by the EPA when Monsanto released the new genetically modified crop to growers. Despite Monsanto stating that Dicamba shouldn't be applied to the crops in an "over the top method", meaning spraying the herbicide after the crops have sprouted, many growers ignored the manufacturer. A big mistake that caused a massive widespread of the herbicide on non-GMO resistant crops.

Within two years, over two dozen states across the midwest region from Arkansas to North Carolina have been polluted with dicamba with 3.6 million acres of non-genetically modified soybean crops affected [7]. 2,708 complaints have been filed by non-GMO planters with many suing Monsanto. The situation has gotten so bad that a murder case was connected to a dispute over using the herbicide [1]. In addition to soybeans, crops like tomatoes, watermelon, cantaloupe, and pumpkins have been affected by the herbicide. Dicamba has devastating results once it comes in contact with non-genetically modified crops, causing leaves to pucker, pods to buckle, and stunted growth resulting in eventual death [2].

The federal crop insurance won't cover damages from the herbicide due to most policies only covering natural disasters[2]. This widespread issue affects not just planters, with their crops being their life earnings, but also the consumers where they'll end up having to pay more for all goods affected.

Arkansas and Missouri had issued temporary bans on the use and sale of Dicamba[2]. Well after farmers in the Midwest had pulled the last soy and corn plants from their fields and were deep into preparation for this year's planting season, the EPA approved three new formulations of dicamba. Monsanto hoped that its new formula under the name XtendiMax along with the EPA's prediction that carefully applying the new herbicide in certain environmental conditions will reduce the risk of drift [2]. Despite these new, supposedly less volatile formulations, the problem seems to be getting worse with more dicamba based complaints for the 2017 season.

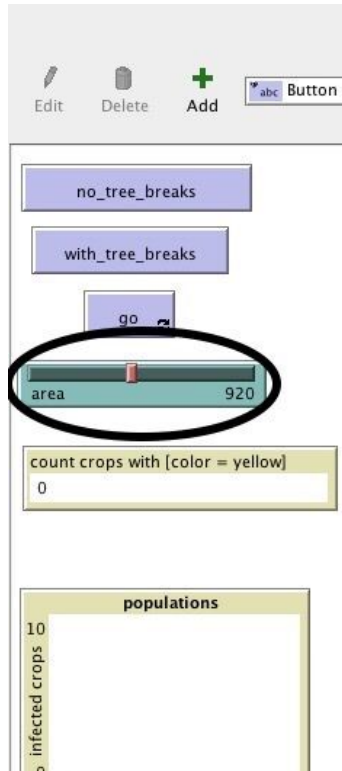
For the 2018 season, much research and testing was put in to reduce dicamba drift. On October 13th the EPA announced enhancements to Monsanto's branded XtendiMax with VaporGrip Technology in hopes of having a better growing season. Along with the new enhancements, new updates on how the herbicide should be applied were voluntarily proposed by Monsanto and supported by the EPA. With these new updates and regulations proposed, Monsanto's Global Regulatory lead hopes that, "Based on the science behind our low-volatility dicamba product and learnings from the 2017 season, we are confident the required training and record keeping can address the main causes of off-target movement" [10].

Despite new improvements made to the herbicide and the genetically modified crops we believe it's important to find a way to help prepare growers with the certain non-GMO crop if another spread were to occur. With an issue to this magnitude, it is important to find the necessary solutions to help prevent further and any additional widespread contamination of crops.

Description:

We solved this problem computationally with the NetLogo agent-based modeling program. Our model depicts crops that neighbor the genetically modified dicamba resistant crop where a planter has ignored Monsanto's recommendations and sprayed the herbicide. We chose an area in the upper northeastern part of New Mexico, next to Oklahoma due to 19 dicamba related cases coming from Oklahoma [8]. Based on wind maps and predicted weather forecasts we were able to depict the wind direction and speeds around the growing season [9]. As a possible solution, we studied the difference between crops with and without tree breaks.

A variable that we manipulated to find what extent tree breaks helped decrease the spread of the herbicide was the concentration of the herbicide. By depicting the land area the herbicide is applied to we're able to manipulate the concentration of dicamba used. In theory, the larger the area, the greater concentration of dicamba.



create-herbicides area / 40

Figure 1.1

We then had the slider value divided over 40. This value depicts the standard amount of gallons of water mixed with 16 ounces of dicamba per acre that is recommended by Monsanto.

In the interface tab, we sprouted multiple agents based on the patch color under the name crops. We created two setup buttons under the name `no_tree_breaks` and `with_tree_breaks`. The `no_tree_breaks` setup button sets up 30 black patches that are positioned in lines that are specially set up in columns.

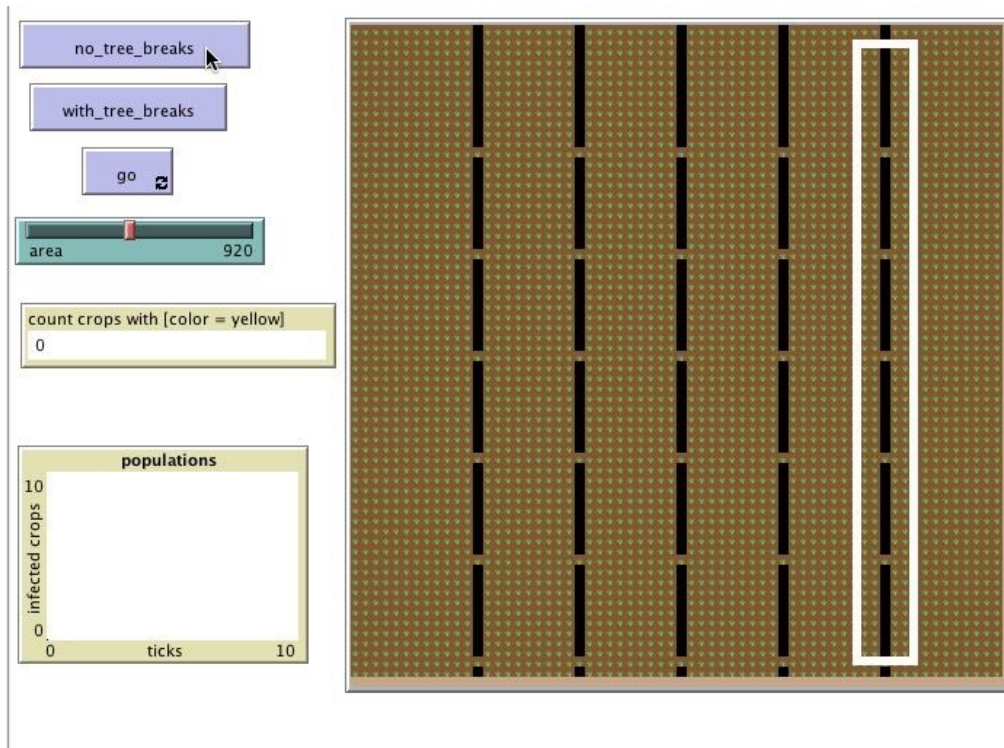


Figure 1.2

These black patches simulate roads. When our agent, under the name herbicide, hits this black patch it decreases its potency by 10%.

The `with_tree_breaks` setup button introduces the second agent named trees. This agent takes over all areas where the black patches were placed.

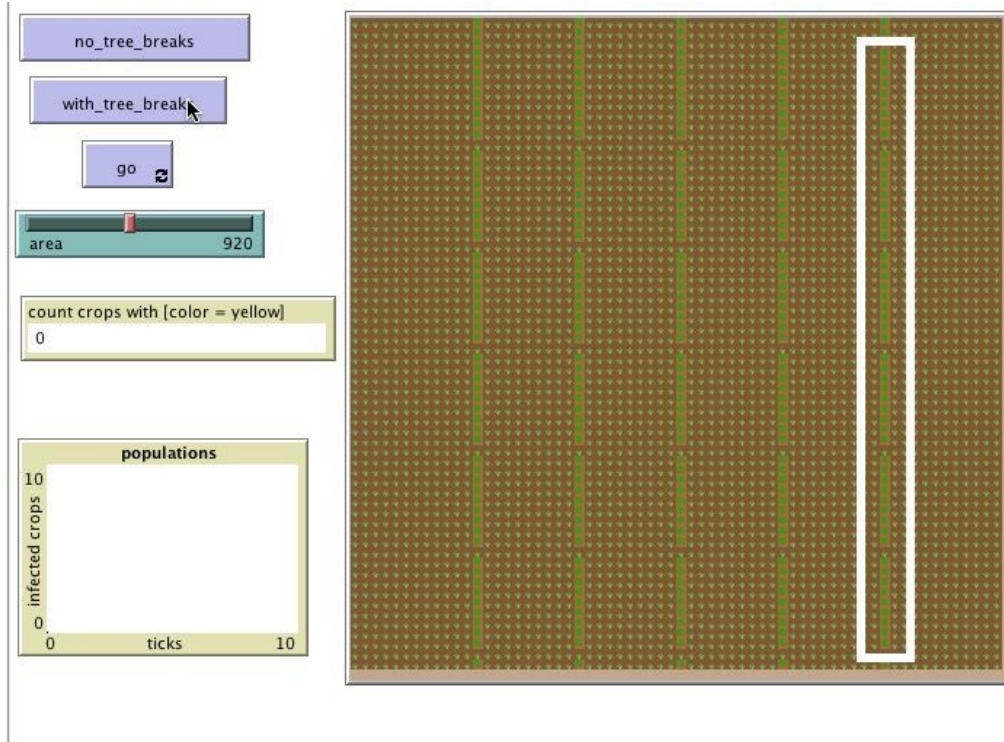


Figure 1.3

In order to simulate the effect of tree breaks on wind, the herbicide slows down by going forward from 1.0 (60 mph) down to .3 (18 mph). Also, the potency goes down by 70% and does a wiggle walk to simulate the air disruption from the trees.

To depict the herbicide diffusing through air, the herbicide follows a set of potency values that's randomly generated in the setup procedure as shown below:

```
set potency random 200 ;Able to adjust the potency of the herbicide based on the slider in the interface tab.
```

```
]
```

```
ask trees
```

```
[
```

```
  set potency random 100
```

```
]
```

We chose this potency value as a baseline for how much dicamba is mixed in the solution that's sprayed onto the plants. If we were to convert this over to the real world data that random 200 would be equivalent to 12-16 oz of dicamba. Overtime the herbicide potency values will slowly decrease based on what patches or agents it interacts with. Once the potency level reaches zero the herbicide dies out.

For the go button we have the herbicide agent do a wiggle walk to simulate the herbicide going through the wind. The agents follow the sub set code below titled “to move”:

```
to move
ask herbicides
[
  let colorOfPatchAhead pcolor
  ask patch-ahead 1 ;defines colorOfPatchAhead
  [
    set colorOfPatchAhead pcolor ;sets colorOfPatchAhead to be pcolor
  ]

  if (colorOfPatchAhead = brown) ;; conditional statement that tells the herbicide to wigglewalk
  [
    ;pen-down

    right 1
    left 2
    fd .3
    set potency (potency * .3)
  ]
```

Using multiple conditional statements with defining the “colorOfPatchAhead” statement that tells the agent to look ahead at the patch in front of them. Certain patches cause the herbicide to move differently or lose a certain amount of potency. We also have the herbicide interact with the other agents. When it hits the crops it turns the crops yellow visualizing that they are affected

by the herbicide. This example of the interaction of the agents is under the sub set of code within the go button titled “to infect”:

to infect ;;makes non-resistant crops yellow and makes the potency go down for the herbicide

```
ask crops
[
if color = green + 1
[
if any? herbicides-here
[set color yellow]
]
]
```

To simulate the effects of dicamba the crop turns yellow and grows at a slower rate. To simulate the growth of the plant we followed three agent shapes “plant small” this represents the plant in the early phases of the growing season. “Plant medium” this simulates the plants growth during the middle of the growing season. Lastly there is “plant” this simulates a full grown plant at the end of the growing season and beginning of the harvesting season. After a certain amount of ticks the crops that are yellow die out, leaving a trail where the herbicide drifted through.

We finally used the behavior space tool in the NetLogo program to quickly and neatly organized our results. We ran a total of a hundred trails to test out varying concentrations based on the land size that dicamba is applied to.

Results:

To test our model, we first ran a hundred trials with both setups then averaged out the maximum amount of crops infected based on the variable sizes of the area controlled by a slider. For the first setup we ran the model with no tree breaks. We started off with 200 acres then increased by 200 acres until it hits the maximum of 2000 acres.

When the area increased in size we noticed that the amount of crops that were turning yellow (meaning they were infected) would increase. Starting at 200 acres, the average infection rate was 137 crops. When it reach the maximum area of 2000 acres the average infection rate was 473 crops. Rather interestingly, the number of affected crops increased at variable rate. For instance, the areas where we saw the number greatest number of crops infected was between 200 and 400 acres and where we saw the least amount of infected was between 1600 and 1800 acres.

We then ran the same test with the same increasing variables of acreage but with tree breaks included, we saw a dramatic difference between the model with and without tree breaks. The number of crops infected dramatically decreased with the largest amount infected reaching a peak at 107 crops. Looking through the averages of the variable acreages of land we see the same trend that we saw with no tree breaks. As the acreage increased the number of infected crops increased as well until it reached 2000 acres where the number of infected crops went down from 107 to 102 crops.

This is where we started to see a difference in trends between with and without tree breaks. That theory I stated in the description of our model proved correct, as the acreage went up, the amount of crops infected also went up. This is most likely due to an increasing concentration of the herbicide used in order to cover more land.

Results from an area with no tree breaks

| size in acres | average amount of crops infected after ten trials |
|---------------|---|
| 200 | 137 |
| 400 | 212 |
| 600 | 274 |
| 800 | 324 |
| 1000 | 357 |
| 1200 | 385 |
| 1400 | 412 |
| 1600 | 448 |
| 1800 | 459 |
| 2000 | 473 |

Table 1

Results from an area with tree breaks

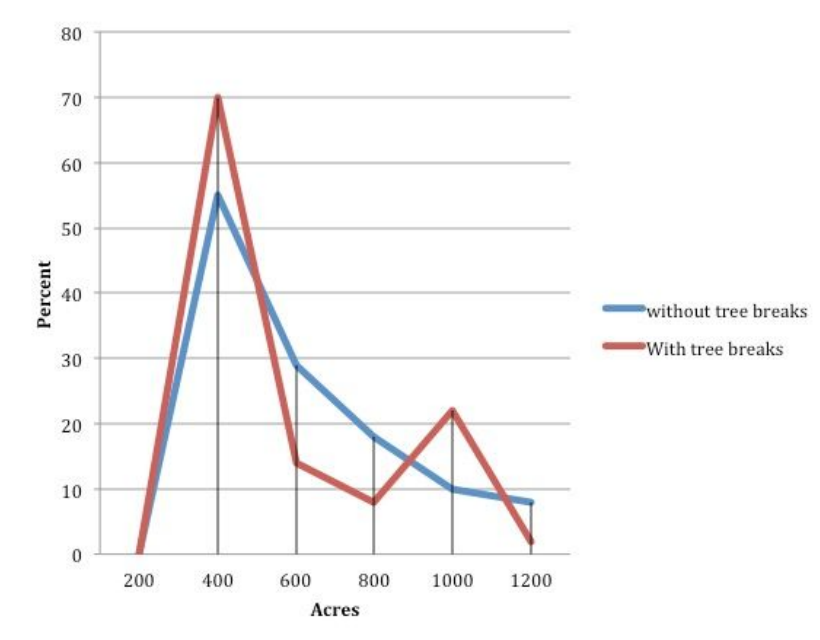
| Size in acres | average amount of crops infected after ten trials |
|---------------|---|
| 200 | 33 |
| 400 | 56 |
| 600 | 64 |
| 800 | 69 |
| 1000 | 84 |
| 1200 | 86 |
| 1400 | 97 |
| 1600 | 102 |
| 1800 | 107 |

2000

102

Table 2

Percent Increase of Infection Between Increasing Areas



Graph 1

Conclusion:

Our data supported our hypothesis that tree breaks would reduce the infection rate of the crops. On average, the Dicamba infected 348 crops without tree breaks and 80 crops with tree breaks. At every interval, tree breaks reduced the amount of infected crops to less than 25% of the amount before the tree breaks. By running ten trials of each model with and without tree breaks we are able to pinpoint where we would see the largest difference if we were to implement tree breaks.

In lower concentrations of Dicamba, the difference between tree breaks and no tree breaks is very apparent. At a rate of 400 acres sprayed, the Dicamba shows the greatest

difference between tree breaks and no tree breaks at 26%. At 1,200 acres sprayed, the difference was the least at 21%.

Recommendations:

In order to make our model more realistic we first had wanted to track different methods of application of the Dicamba. We narrowed it down to one solid model showing its travel and spread through the air. We had wanted to show it's spread through irrigation as well but we had trouble finding an algorithm to properly represent that spreading. Another idea we explored was the implication of wind turbines. This would increase the lengths of our growing seasons by warming up the ground and preventing frost. In conclusion we decided to have our model show the difference between having tree breaks and not having tree breaks.

Acknowledgements:

We want to thank our project mentors Susan Gibbs and Paige Prescott. They have been a huge help in establishing and refining our project ideas in order to further express them onto paper and Netlogo. We also want to give a huge thanks to our teacher Rhonda Crespo. She is the main reason why we joined the Supercomputing Challenge and has always kept us on track and ensured our project is going smoothly.

References

- [1]. Arkin, Daniel. "Dicamba Drift': Monsanto Defends Herbicide as Farmers Say It Harms Crops." NBCNews.com, NBCUniversal News Group, 28 Oct. 2017, www.nbcnews.com/news/us-news/amp/dicamba-drift-monsanto-defends-herbicide-farmers-say-it-harms-crops-n814716.
- [2]. "EPA-Approved Dicamba Is an Airborne Menace and Some States Are Banning It." Modern Farmer, 11 July 2017, modernfarmer.com/2017/07/dicamba-drift/.
- [3]. Farm Press Staff 1 | Dec 05, 2017. "2017 Mid-South Weed Control: Heavy with Dicamba Concerns." Delta Farm Press, 5 Dec. 2017, www.deltafarmpress.com/weeds/2017-mid-south-weed-control-heavy-dicamba-concerns.
- [4]. Hamilton, K. C., and H. F. Arle. "Response of Cotton (*Gossypium Hirsutum*) to Dicamba | Weed Science." Cambridge Core, Cambridge University Press, 12 June 2017, www.cambridge.org/core/journals/weed-science/article/response-of-cotton-gossypium-hirsutum-to-dicamba/A78A2261CBBA913AA586B8C154952F46.
- [5]. "How to Plant a Windbreak to Conserve Energy". Arbor Day Foundation. 17 March, 2018, <https://www.arborday.org/trees/climatechange/windbreak.cfm>

[6.] Ishii-Eiteman, Marcia. "Dicamba Drift? This Is on You, Monsanto." HuffPost - Breaking News, U.S. and World News, HuffPost, 8 Sept. 2017, m.huffpost.com/us/entry/us_59b2e283e4b0bef3378cdfda/amp.

[7]. Lipton, Eric. "Crops in 25 States Damaged by Unintended Drift of Weed Killer" The New York Times. 1 November, 2017, <https://www.nytimes.com/2017/11/01/business/soybeans-pesticide.html>.

[8]. "These Maps Show How Much Farmland Has Been Damaged by Dicamba Drift." Modern Farmer, 1 Nov. 2017, modernfarmer.com/2017/11/maps-show-much-farmland-damaged-dicamba-drift/.

[9]. "Windy as Forecasted." Windy: Wind Map & Weather Forecast (Also Known as Windyty), www.windy.com/?36.001%2C-106.132%2C5.

[10]. "Dicamba and the Treadmill of Herbicide Resistance." Integrated Weed Management Resource Center, integratedweedmanagement.org/index.php/2017/11/14/dicamba-and-the-treadmill-of-herbicide-resistance/.

Efficient Evacuation of At-Risk Populations

New Mexico Supercomputing Challenge

Final Report

April 4th, 2017

Team Number: SFHS 162

Santa Fe High School

Team Members:

Rowan Cahill

Ramona Park

Luke Shankin

Teacher Sponsor:

Brian Smith

Teacher Sponsor and Mentor:

Hope Cahill

Table of Contents

| | |
|--------------------------------|----|
| Table of Figures | 2 |
| Executive Summary | 3 |
| Introduction | 4 |
| Purpose..... | 4 |
| Significance | 4 |
| Background | 7 |
| Description..... | 10 |
| Scope | 10 |
| Setup and Scaling..... | 11 |
| Methods..... | 12 |
| Challenges | 14 |
| Conclusions..... | 15 |
| Results..... | 15 |
| Analysis and Future Plans..... | 16 |
| References..... | 18 |
| Acknowledgements | 19 |
| Team Biographies | 20 |
| Appendix A: Data Tables | 21 |
| Appendix B: Code | 23 |

Table of Figures

| | |
|---|----|
| Figure 1: Death toll rises as flood subsides in Texas..... | 5 |
| Figure 2: Traffic jams on highways in Miami, Florida Source: zmescience | 6 |
| Figure 3: 9/11 al Qaeda Attack on World Trade Center leaves lasting effects on NYPD in present day..... | 7 |
| Figure 4: City of Kenner, Louisiana central flow map hurricane information..... | 8 |
| Figure 5: Screenshot of expanded grid layout | 11 |
| Figure 6: Screenshot of evacuate-to and evacuate-from in process | 13 |
| Figure 7: Screenshot of segment grid visualization | 13 |
| Figure 8: Methods of Evacuate-to Time Comparison Graph | 15 |
| Figure 9: Methods of Evacuate-from Time Comparison Graph..... | 16 |

Executive Summary

In light of the numerous natural and man-made disasters affecting the United States, if and how a population evacuates have become pertinent questions. For instance, during Hurricane Harvey, the city of Houston questioned and chose not to issue mandatory evacuation orders due to their experience with Hurricane Rita. On the other hand, in Florida and Puerto Rico such orders were given out and, in the case of the former, an organized method was used. The goal of our simulation was to determine which real-world method of evacuation should be utilized via observation of the time taken for complete egress. These methods include contraflow, chaotic (or unplanned), and segmented evacuation.

Over the course of this project, research concerning real world evacuation plans and statistics from past evacuations was conducted. A controlled two-dimensional Netlogo model was developed from this research. The model does not include unpredictable or complex human behaviors and is therefore limited in its applicability and accuracy. However, it provides a baseline for a more in-depth simulation. Preliminary data from the simulation indicates there is little time difference between contraflow and chaotic evacuation. On the other hand, segmented evacuation results in a greater time duration for complete egress. The length of time recorded for this was roughly proportional to the number of segments evacuated in relation to chaotic evacuation.

Introduction

Purpose

Given recent events, including natural disasters and terrorist attacks within populated areas, the questions of what the most effective method of evacuation is and whether there is a need to evacuate have become increasingly relevant. In response, many cities and areas commonly affected by threats requiring evacuation and even those less threatened are in the process of developing egress strategies. Unfortunately, as is the case with Hurricane Rita, history has demonstrated that the methods proposed don't always work and can at times cause more of a risk than staying put.

To determine which method results in the fastest evacuation of a region our code simulates several of these methods within a grid-like road system. These include contraflow, chaotic, and segmented evacuation. In developing this rudimentary model, multiple algorithms have been tested to provide a baseline for future applications. With modification, the model created could be utilized by city planners to test potential methods of evacuation before a disaster occurs.

Significance

Over the past several years, sudden threats to communities have become an imminent problem both locally and worldwide. These threats, such as hurricanes and wildfires, often require a planned response, including threat assessment and evacuation.

Many recent headlines highlight this issue. For instance, when Houston, Texas, the nation's fourth largest city, was struck by Hurricane Harvey in 2017 government officials chose not to evacuate 2.3 million residents based on flawed procedures during Hurricane Rita in September 2005. Their likely reasoning, "of the more than 100 people

who died during Hurricane Rita, at least 60 of those deaths were in connection to the evacuation [1].” Moreover, as reported by the BBC, flawed instructions caused “millions of Texans to flee their homes and led to 20-hour traffic jams, stranding drivers across hundreds of miles [2].” 108 casualties have since been attributed to Hurricane Harvey. It is impossible to say whether these deaths could have been avoided by evacuation.



Figure 1: Death toll rises as flood subsides in Texas
Source: *MSNBC/NBC News*

Similarly, Hurricane Irma, which traveled across the Caribbean Islands, Cuba, and Florida, resulted in the evacuation of upwards of 6.3 million people—one of the largest evacuations seen in U.S. history [3].

According to the Washington Post, however, Florida's two main roads, Interstates 95 and

75, became "parking lots." The roads reportedly had no exits for "miles at a time" and families were faced with the option of ignoring government orders and staying in their homes, or risk running out of gas before being able to get off the highway.



Figure 2: Traffic jams on highways in Miami, Florida
Source: *zmescience*

More recently, in Southern California, fire officials ordered evacuations in response to the Thomas Fire, resulting in tens of thousands of residents fleeing their homes. Despite casualties being less significant than those seen in the aforementioned situations, the devastation to California was undeniably brutal. More than 259,000 acres were burned and over 1,000 structures were destroyed with about 18,000 more homes at risk of being burned [4]. It is difficult to imagine, in this instance, that the evacuation orders issued didn't save lives.

This study is significant on a local level as areas in New Mexico, such as the Sandia and Los Alamos National laboratories, are at risk of nuclear leakage or terrorist and/or military attacks. Additionally, the potential of forest fires, like those in California, is undeniable for New Mexicans, as evidenced by the 2017 wild-fire season. Any of these hypothetical situations could require evacuation of the local populace and fires have.

Background

Evacuations fall into two overarching categories, small-scale and regional evacuation. By definition, a small-scale evacuation is localized; they are often the result of a sudden, violent occurrence [5]. The evacuations implemented after the attacks on the World Trade Center and the Boston Marathon Bombing are an example of this. In these scenarios, while it was necessary to evacuate neighborhoods, it was unnecessary to evacuate the entire city or region.



*Figure 3: 9/11 al Qaeda Attack on World Trade Center leaves lasting effects on NYPD in present day
Source: APA Photo/NYPD via ABC New*

In contrast, a regional evacuation is instigated by a large-scale threat. Natural disasters, including hurricanes or tsunamis, that have the potential to cause wide scale damage and pose a large casualty risk qualify as such. Evacuation of all areas affected may be deemed necessary during these events [5].

Models of evacuation are further subdivided into maximum flow (MFM) and minimum-cost maximum flow (MC-MFM) models. In both cases, the most efficient mode of egress, that which takes the least amount of time to evacuate the maximum number of subjects, is desired. A key difference is that the MC-MFM factors in the possible risks and costs associated with evacuation as its intent is to combine efficiency with minimal loss of life and costs [5]. The latter are not considered in the development of MFMs.

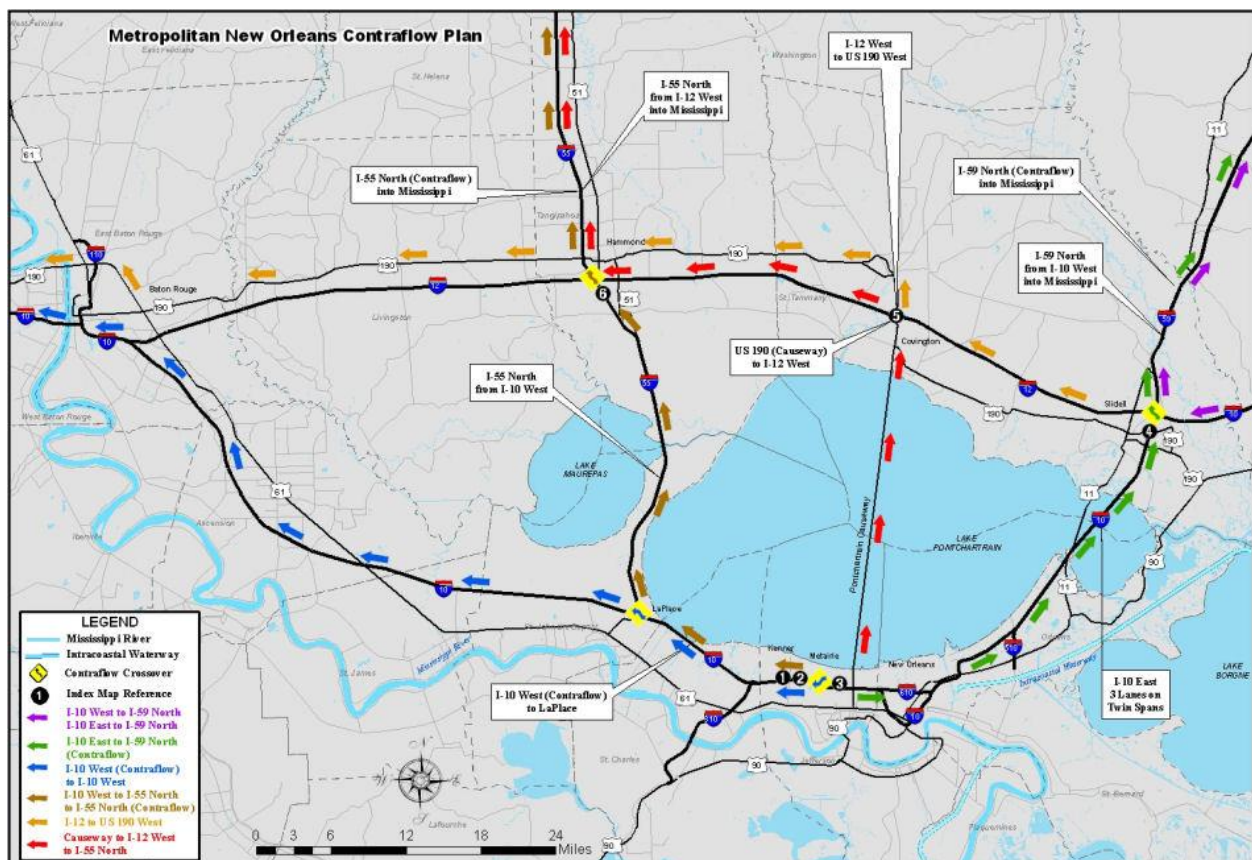


Figure 4: City of Kenner, Louisiana central flow map hurricane information.
Source: Louisiana Citizen Awareness and Disaster Evacuation Guide

The potential risks and costs of an evacuation are numerous and can change depending on the location. The unpredictable and complex nature of humans, however,

is a constant. Lack of preparation, inability or refusal to follow directions, road rage, and accidents can cause confusion on the roadways, leading to congestion and longer evacuation times. Accordingly, evacuation related deaths have been linked to accidents and health issues such as heat stroke. Some evacuation methods (e.g. segmented and contraflow) take these factors into account and attempt to limit the number of people on the road or increase the carrying capacity of the roads. The costs that are associated with evacuations must also be considered, with the most worrisome being the loss of lives. Hurricane Rita, after which many casualties were attributed to the evacuation, exemplifies the potential risks and costs of evacuation. Currently, our model does not take these issues into account, as it is a maximum flow model.

The two methods simulated beyond chaotic evacuation, segmented and contraflow, are based on real-world methods. The first, segmented evacuation, is used most notably in Florida, where the state is divided into several regions. In case of an emergency, the regions are given specific dates and times to evacuate to ease the burden on the roadways [6]. This is most relevant during large-scale evacuations. To reflect this, the road system in our Netlogo model, may be divided into sections via a slider. The second method, contraflow evacuation, involves converting major roads and those parallel to them into one-way streets, repurposing the other direction to increase traffic flow[7].

Description

Scope

Our two-dimensional Netlogo model simulates the evacuation of a population from an at-risk area or to an area of safety through a simulated road system. This enables us to compare the time taken for complete egress via several modes of evacuation. The road system is created using a combination of agents and links, representing nodes (or intersection) and roads respectively. Nodes have variables that allow them to track their distance from the source to aid in finding routes. Links, on the other hand, have a mile per hour (mph) variable that represents the speed limit along that road.

Currently, two real-world methods of structured evacuation have been implemented, as well as a chaotic evacuation. The first structured evacuation model, based on the method used in Florida, is segmented. Recall, this involves the designation of regions or sections which are evacuated separately based on risk and proximity to the threat. The second structured method is contraflow evacuation. This mode of egress requires the conversion of major roads and those that run parallel to them into one-way streets running in the direction of the desired exit. Within our program, the chaotic method of evacuation provides the foundation for both models. In chaotic evacuation, the entire area exits via the most efficient path as determined by a flood-fill algorithm. All modes of egress will be explained in greater detail in the *Methods* section below.

Setup and Scaling

To control the experiment, both the layout of the intersections and the speed limits of the roads were standardized. The nodes are distributed in a square grid with the distance between them established by the user before running the model. This allows for expansion or contraction of the area to be simulated for more versatile data collection and testing. Standardization of the speed limits made use of their location in the Netlogo world. This provided a distribution of values to demonstrate the functionality of the model.

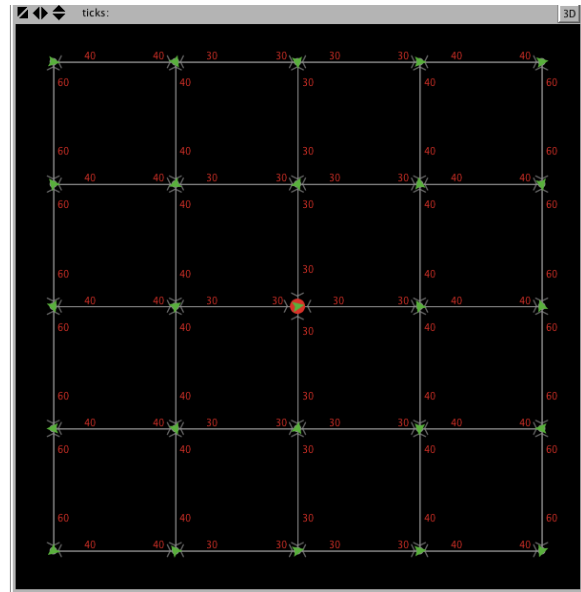


Figure 5: Screenshot of expanded grid layout

To analyze the results two assumption were made on the scale of the model. Each “step” or cycle the model ran through was decided to be 1/1000 of an hour. This means each car in one cycle traveled the speed limit of the link they were on divided by 1000. To further simplify this math and retain clean numbers, each patch was set as 1 square mile measuring one mile horizontally and vertically. From these numbers the amount of cycles taken to complete an evacuation by the last car to reach the goal, or the furthest point, can be converted into a time in minutes by dividing by 1000 and then dividing by 60.

Methods

As stated above, the simulation models several methods of evacuation. However, before any of these can be run, the program completes a flood-fill to determine the distance of every node from the source node which represents either a destination or danger. The flood-fill algorithm is a recursive function originally called by the source and then run by all linked nodes. First every node sets its distance from the source to an arbitrarily high value greater than any possible distance. This step is necessary for the algorithm to function. The source then sets its variable which records the distance from source to zero. Finally, it asks all connected nodes with a value for their distance from source greater than its own, to add their distance from itself to its distance from source and then run the same program. By having only those with a higher value run the program, it ensures that the lower values, indicating distances already calculated, are not overwritten. When these next nodes run the same program, they have their connected nodes add their distance from the asking nodes to that nodes distance source. With each successive node that runs the program, the values for their distances are filled in a flood outward from the source. Once there are no more nodes with values higher than the previous nodes, the program completes and the methods for evacuating can be run.

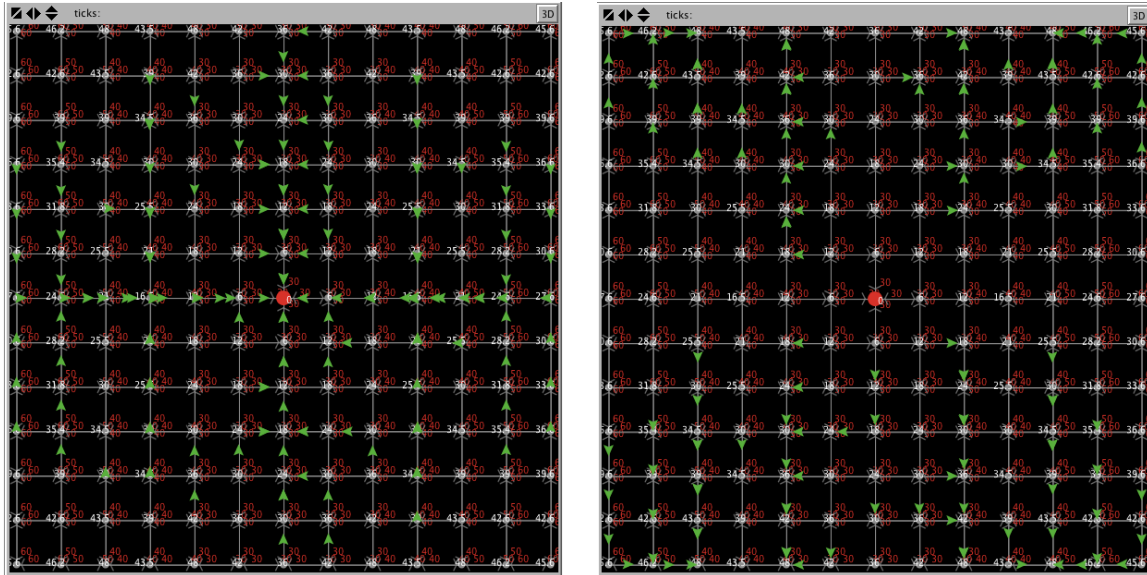


Figure 6: Screenshot of evacuate-to and evacuate-from in process

The two modes of structured evacuation, segmented and contraflow, both rely on the systems of chaotic evacuation, evacuate-to and evacuate-from. In these systems, the cars select their next target intersection, or goal, from those nodes connected to their current node. For evacuate-to, the node with the lowest value for distance from source is chosen, as choosing a lower value will eventually lead it to zero (0), the value held by the source. Conversely, in evacuate-from the node with the highest value is chosen. In segmented evacuation, the road system is divided into several segments as determined by a slider. Then the cars within each region, progressing from bottom to top and left to right, run either evacuate-to or evacuate-from. Once a region is fully evacuated,

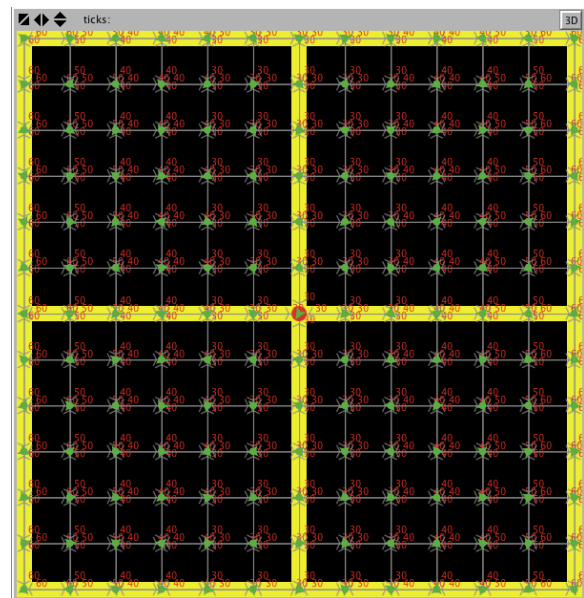


Figure 7: Screenshot of segment grid visualization

indicated by an absence of cars, the vehicles in the next segment run the same method and so forth and so on until no more need to be evacuated. For contraflow evacuation, the change is made before the flood-fill algorithm is run. The links with a heading approximately equal to the number representing one of the cardinal directions (e.g. 0 for north and 90 for east) are leaving only the link pointing in the opposite direction. Then the flood-fill and chaotic functions are run. It should be noted, that it is possible to perform both methods simultaneously, but for comparison they were only run separately.

Challenges

Throughout the development of the code, several challenges had to be addressed and the code's scope changed accordingly. The first and frequently recurring problem was the use of unfamiliar Netlogo functionalities, specifically links. Due to this, the code is not as fully developed as planned; self-teaching and communication with mentors to gain the necessary knowledge delayed the coding process. This delay prohibited the inclusion of more complicated variables and behaviors such as human actions, inhibiting the model's real-world applicability.

The most significant coding issue occurred during the development of the algorithm for contraflow evacuations. To establish one-way roads, directed links were used which only respond and interact in one direction. Once those links were implemented, however, the flood-fill program and chaotic evacuation methods stopped functioning. This was due to an incompatibility between directed and undirected links that required a light, but expansive, reworking of the entire code to ensure that all parts were able to function together.

Conclusions

Results

To gather consistent data, several governing decisions were made. First, the distance between intersections was set at 3, providing a large road system. Second, only four segments were used for segmented evacuation; up to 36 are possible. Finally, only one of the four potential contraflow evacuation methods were used as they produce nearly identical results. With these parameters, the chaotic evacuation and contraflow evacuation for evacuate-to are nearly identical in the time taken for complete egress, averaging 78.140 minutes and 78.540 minutes, respectively. Evacuate-to segmented evacuation was drastically greater at 244.488 minutes. This number is roughly proportional to the number of sections evacuated as compared to chaotic evacuation.

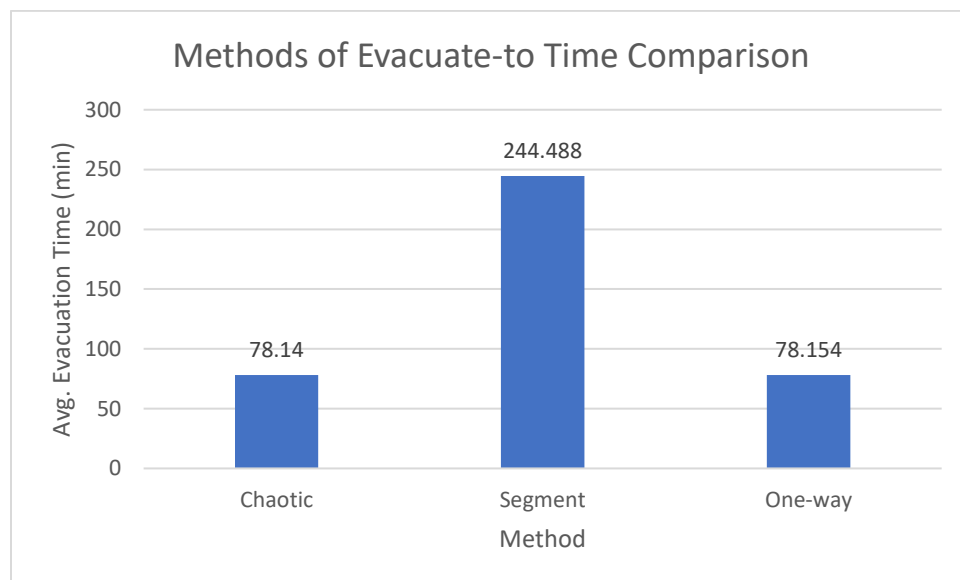


Figure 8: Methods of Evacuate-to Time Comparison Graph

The evacuate-from data includes only the chaotic and segmented modes which are similarly far apart. The former recorded 89.707 minutes and the latter 268.800

minutes. Contraflow was not modeled as it is a method of evacuating to a point outside the city.

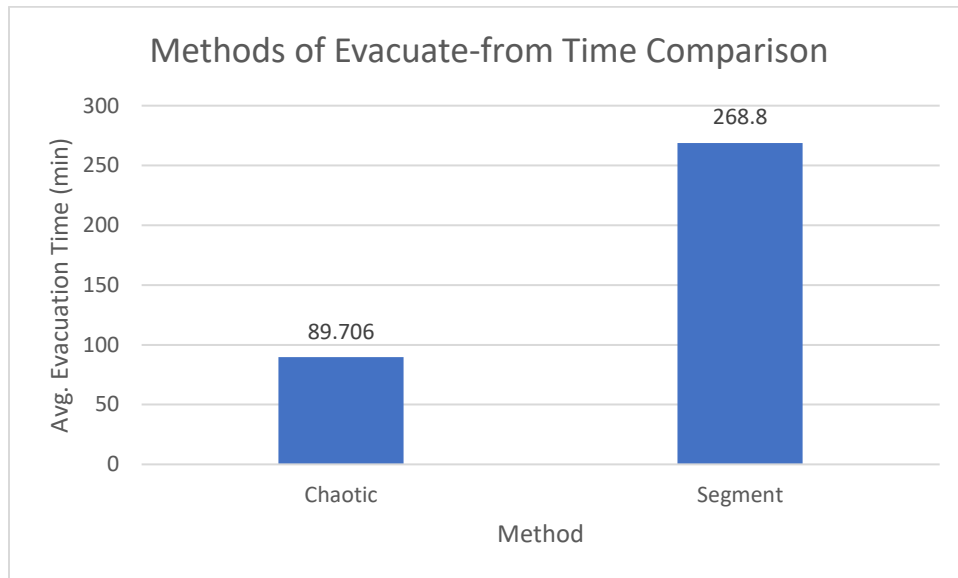


Figure 9: Methods of Evacuate-from Time Comparison Graph

Analysis and Future Plans

The data gathered from our model provides a basic understanding of the comparative times associated with each method. However, as the model does not account for human behavior and other complex variables the results it produces are not applicable to the real-world. This is further illustrated by the narrow range observed in the results. Due to the nature of the methods at this time, the results produced are computable and standard. To expand, as the cars simulated always choose the best routes and are unimpeded, the maximum time is roughly equivalent to the furthest agent. For the model to represent the real world, human behavior and its consequences, such as traffic and accidents, need to be implemented. These have the potential to

drastically affect the time taken to evacuate in all cases. In the meantime, the model as it currently stands, could be utilized as a foundation for more complicated iterations.

References

- [1] “Evacuation Planning in Texas: Before and After Hurricane Rita.” *Interim News*, 14 Feb. 2006, www.hro.house.state.tx.us/interim/int79-2.pdf.
- [2] “Why Was Houston Not Evacuated before Hurricane Harvey?” BBC News, BBC, 29 Aug. 2017, www.bbc.com/news/world-us-canada-41085438
- [3] “Hurricane Irma: Time Runs out for Florida Evacuation.” BBC News, BBC, 10 Sept. 2017, www.bbc.com/news/world-latin-america-41214622.
- [4] Schneider, Avie. “Southern California Fire Forces New Evacuations, Threatens Coastal Communities.” NPR, NPR, 17 Dec. 2017, www.npr.org/sections/thetwo-way/2017/12/16/571380748/southern-california-fire-forces-new-evacuations-threatens-coastal-communities.
- [5] Li, et al. “Optimization and Planning of Emergency Evacuation Routes Considering Traffic Control.” *The Scientific World Journal*, Hindawi, 29 May 2014, www.hindawi.com/journals/tswj/2014/164031/.
- [6] “Emergency Information.” FloridaDisaster.org, Florida Division of Emergency Management, www.floridadisaster.org/info/.
- [7] MCEMA. “Contraflow.” *Mobile County EMA > Evacuations Procedures > Contraflow*, 2018, www.mcema.net/Evacuations-Procedures/Contraflow.
- [8] Wang, Jing-Hong. “Principal Aspects Regarding to the Emergency Evacuation of Large-Scale Crowds: A Brief Review of Literatures Until 2010.” *Procedia Engineering*, Elsevier, 16 May 2014, www.sciencedirect.com/science/article/pii/S1877705814004196.
- [9] Griggs, Brandon. “Harvey’s Devastating Impact by the Numbers.” *CNN*, Cable News Network, 1 Sept. 2017.
- [10] Levin, Matt. “How Hurricane Rita Anxiety Led to the Worst Gridlock in Houston History.” *Houston Chronicle*, 22 Sept. 2015.
- [11] Modesti, Kevin. “Assessing the Scope: Southern California Fires Add up to Disaster on a Jaw-Dropping Scale.” *Daily News*, Daily News, 9 Dec. 2017.
- [12] Shaw, John R. “QuickFill: An Efficient Flood Fill Algorithm.” *QuickFill: An Efficient Flood Fill Algorithm*. - CodeProject.
- [13] Wang, Amy B, et al. “As Harvey Submerges Houston, Local Officials Defend Their Calls Not to Evacuate.” *The Washington Post*, WP Company, 28 Aug. 2017.

Acknowledgements

Thank you to all those below...

Ms. Hope Cahill, a teacher at El Dorado Community School, provided regular brainstorming, organizational, and editing support throughout the entire project. She made sure that the team kept focus during our meetings and helped us keep on track. This project would not be completed without her.

Mr. Brian Smith was extremely helpful in the initial organization of the team and keeping it on track to finish. He made sure that we had the funds to register for the Supercomputing Challenge and participate in the Kickoff.

Mr. Stephen Guerin, who taught us how to use links and nodes in Netlogo, a system we were unfamiliar with, to the point of never having heard of it. And who helped us get the code off the ground by guiding us towards using a flood-fill algorithm to solve our pathfinding dilemma. He also pointed us in the direction of libraries we wouldn't have accessed otherwise.

Team Biographies

Rowan Cahill is a senior attending Santa Fe High School and has an interest in all things mathematical. He will be attending Colorado School of Mines to study Computer Science and Mechanical Engineering. This year, Rowan was primarily responsible for the development of the code and the methods used to simulate the evacuations. He also did extensive research and segments of the paper and final edit.

email: rcahill505@gmail.com

Ramona Park is a senior at Santa Fe High School who holds interests in graphic design and the production of creative content. She is aimed at pursuing a degree in computer science with an emphasis on communication and visual arts in order to implement better flow of information in the digital age. During this project, she played a central role in organizing our resources and contributed to both the interim and final reports.

email: yoharamona@gmail.com

Luke Shankin is a senior at Santa Fe High School who will begin attending New Mexico Tech in the fall of 2018 in order to pursue a degree related to Computer Science. Over the course of the Challenge, Luke played a role in leadership, conducted research related to historical evacuations, worked with the code, and contributed to both the interim and final reports.

email: lukeshankin@gmail.com

Modeling Straw Mulch in Taos Agriculture

NM Supercomputing Challenge

Team: THS-31

Taos High School

April 4, 2018

Team Members:

Rowan Kinney
Eben Bellas

Sponsor:

Tracy Galligan

Mentor:

Gregory Rael

Executive Summary:

The goal of this project was to find the optimal amount of straw mulch to put down to minimize water related costs. To solve this optimum we modeled the evapotranspiration based on weather data collected in NM. Each year we here about the water resources such as the Ogallala Aquifer being drained. Furthermore, many farmers in the US are struggling financially. By finding a way to decrease water consumption, we can potentially save farmers thousands of dollars, while protecting water resources.

We built the model in MATLAB following a guide to computing evapotranspiration (ET). Our model seemed to work at first, however after working with the data we found that there were some errors. After going through the code again we found that there was a ratio that could not exceed 1.0, however the guide we had been using had omitted this detail. We found the ratio information on the official website of the FAO-56 method for the Penman-Monteith equation.

We found the error because our code was calculating negative ET. Furthermore, we found that the optimum was a negative value that indicated to us that there was probably some kind of error.

Introduction and Background Research:

Mulching reduces soil loss primarily by decreasing raindrop impact and runoff velocity. Mulching is also considered an effective method of manipulating crop growing environment to increase yield and improve the quality of the product through weed growth, the reduction of soil temperature, the conservation of soil moisture, reduction of soil erosion, and the improvement of soil structure [4]. This works whether the mulch applications are surface applied or surface incorporated, and depends on multiple other factors including the erosivity of the rainfall, the type of soil, the soil condition, the length and angle of the desired slope, and the type of mulch application used. As for choosing a mulch application most effective for soil erosion control, the most realistic selections are naturally occurring sources, and depend greatly upon cheap availability and land use. Another important factor in the determination of a particular mulch application is ease of management, particularly when the applied mulch is for annual cropping. Rate of application is considered to be the most important parameter of mulching, and previous research has shown that there is a correspondence between increased mulching rates and decreased soil erosion up to an optimum mulching rate. The effectiveness of erosion control varies depending on the slope. The more gentle the slope, the more effective the erosion control will be up to an optimum slope. As the slope of the cultivated increases, mulching requirement for effective erosion control increases.

Although the majority of published research in the field, (global effectiveness of using crop residues as mulch in soil erosion control), involved heavier mulch applications, research has shown that mulch applications of only 0.56 t/ha or lower could be effective on slopes as steep as 15% during drastic rainfall [2]. This is considerably beneficial considering most published research involved mulch applications of about five tons per hectare (t/ha).

Mulching improves biotic activity, adding nutrients to the soil, and therefore increasing soil fertility through decomposition [4]. The type of mulch application used determines the impact on physical and chemical properties of soil, as well as crop yield. The quality of different mulching applications is determined by the nutrient value, texture, rate of decomposition, availability, cost, growth rate and vegetative matter turn over. The residue quality of the mulch applications determines the nutritional effects that they have on the plants. Higher quality materials improve the nutrition of the plant by releasing excess nutrients, lower quality residues

have a comparatively weak nutritional effect on the plants. The chemical composition of plants differ, therefore changing the rate decomposition and suitability as a viable mulch application. To test this, experiments were conducted on an Oxic Tropudalf (an environment containing sandy loam soil), to study the effects of Chromolaena and Tithonia mulches on soil chemical properties, the nutrient composition of leaves, and growth and tuber yield of white yam [4]. As for the site where the experiment occurred, the surface and subsoil layers of the site were sandy loam in texture, with an increase in clay content as you got to the subsoil layers of the site [4]. Before adding the Chromolaena and Tithonia into the environment in 2006, there was a lack of organic matter, which was later attributed to the high bulk density of the soil. The content of organic matter and nutrients was higher at the surface levels of the Oxic Tropudalf, and decreased regularly the deeper into the subsoil layers that was tested [4]. In the studies conducted, both Chromolaena and Tithonia both reduced soil bulk density and temperature. Both mulches also increased the content of Nitrogen, Phosphorus, Potassium, Calcium, and Magnesium in the soil as well as the leaves [4]. When compared with the control, the mulches also increased the growth and yield of yam. The findings showed that Chromolaena and Tithonia mulches increased the soil moisture content, and reduced the bulk density of the soil as well as the temperature. Both the higher moisture content and the lower temperature could be attributed to the reduction of evaporation losses [4]. The reduction of the soil bulk density observed in both the Chromolaena and Tithonia mulch applications compared with the unmulched control plots could be ascribed to the lower concentration of organic materials in the subsoil layers, as well as less aggregation, less root penetration, and compaction caused by the weight of overlying layers [4]. Organic matter reduces soil bulk density, improves soil structure, improves aeration, and enhances water infiltration and retention [4]. Using mulch applications increased activities of soil fauna in the decomposition of organic matter, leading to the enhancement of soil porosity and reduction of the soil bulk density. The mulch also stabilized the soil structure against raindrop impact, preventing any soil erosion, compaction or crusting.

Although there are many advantages for the agricultural sector by using plastic mulch, the widespread utilization of plastic mulch has begun to detriment the environment more than benefit it. The development of completely biodegradable plastic mulch applications was started by Chinese plastic companies to establish more sustainable and “greener” cultivation [5]. Some of the advantages to using plastic mulch applications are the conservation of water during crop production, as well as the suppression of weeds. Statistics showed that twenty million hectares of land in China used polyethylene film for cultivation, and the annual consumption of mulch film is increasing at an average annual rate of 10%. Economically speaking, mulch film is extremely beneficial and does yield more crop production than that of regular mulch applications, but there is also substantial ecological damage from using mulch film [5]. The accumulation of plastic mulch on arable land causes both ecological and environmental pollution, as well as the obstruction of crops’ water and nutrient intake. This obstruction leads to as much as a 20% decrease in a crop’s yield. There are two ways to alleviate pollution problems caused by mulch film, integration of biodegradable alternatives, and recycling [5]. Fully biodegradable plastic materials remain stable during use due to their special molecular structure, also allowing for complete decomposition after its service life. Full biodegradable plastic mulches are an ideal substitution for mulch films, and alleviate the impact that film residuals had on the environment.

There are many advantages to using completely biodegradable plastic mulch applications, starting with economic benefits. There are no labor costs for the reclamation of the film as it degrades after it completing its job. There is absolutely no risk of secondary pollution as it

completely metabolizes in water, and finally the use of biodegradable mulch applications allows for complete control over its covering time, in turn allowing the product to be tailored specifically to the region in which it is used [5].

Project Description:

Model Parameters:

| | | |
|------------------|-----------------|--|
| day | 91-214 J | Day of the year in Julians |
| T _{max} | Degrees Celsius | Maximum temperature of each given day |
| T _{min} | Degrees Celsius | Minimum temperature of each given day |
| RHmax | % | Maximum relative humidity |
| RHmin | % | Minimum relative humidity |
| p | inHG | Pressure, later converted to kPa |
| wind | mph | Average wind speed in miles per hour, later converted to m/s |

Parameters Explanation:

The weather parameters were based weather data collected at the Taos airport since 1993. Our final year of data was 2017. Data was broken up into years. Each year consisted of data starting on the first of April of the given year, and ending on the thirty first of October. The data was retrieved from the website of Weather Underground [1].

The rest of the parameters are based on available data gathered from various websites. All of the values were based on either Taos New Mexico, or in the event that the data was unavailable, we used values for New Mexico.

Method and Equations:

For our research we used the Penman-Monteith equation for Evapotranspiration, solved with the FAO-56 Method. The equation solves for Reference Evapotranspiration (ET₀). The equation listed below is based on the variables in Appendix 1.

$$ET_0 = \frac{.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + .34u_2)}$$

The first part of our code reads excel spreadsheets and puts them into arrays for each weather data so that we can calculate day by day. Each call is for a single year.

```

day = xlsread('Weather Data 1993-2018.xlsx', year(j), 'A2:A215');
RHmax= xlsread('Weather Data 1993-2018.xlsx', year(j), 'D2:D215');
RHmin= xlsread('Weather Data 1993-2018.xlsx', year(j), 'E2:E215');
p = xlsread('Weather Data 1993-2018.xlsx', year(j), 'F2:F215');
Wind = xlsread('Weather Data 1993-2018.xlsx', year(j), 'G2:G215');
Tmin = xlsread('Weather Data 1993-2018.xlsx', year(j), 'C2:C215');
Tmax = xlsread('Weather Data 1993-2018.xlsx', year(j), 'B2:B215');

```

The University of Florida (UF) department of Agricultural and Biological Engineering produced a report in 2010, edited in 2015, that described the step by step methodology for calculating ET₀. In the project we followed their guide. The first step is to calculate the mean daily temperature. In their guide they explain that if only average daily temperatures are available it can be calculated, however it might result in some underestimation of ET due to the saturation vapor pressure being non-linear [3]. As a result, they advise the use of maximum and minimum daily temperature.

```
Tmean = (Tmax + Tmin)/2;
```

Once the mean daily temperature is calculated the guide instructs us to calculate the mean daily solar radiation. Because the weather station we used did not monitor solar radiation, we used monthly averages for New Mexico. This code works by making an array with the daily value based on the month, where Ri is the daily average based on the month.

```
Ri = [9.6, 10.6, 10.9, 9.9, 9.2, 8.3, 7.0];
Rmda = [];
for i = 1:length(day)
    if 1 <= day(i) && day(i) <= 30
        Rnettemp = 3.6 * Ri(1);
        Rmda = [Rmda, Rnettemp];
    end
    if 31 <= day(i) && day(i) <= 61
        Rnettemp = 3.6 * Ri(2);
        Rmda = [Rmda, Rnettemp];
    end
    if 62 <= day(i) && day(i) <= 91
        Rnettemp = 3.6 * Ri(3);
        Rmda = [Rmda, Rnettemp];
    end
    if 92 <= day(i) && day(i) <= 122
        Rnettemp = 3.6 * Ri(4);
        Rmda = [Rmda, Rnettemp];
    end
    if 123 <= day(i) && day(i) <= 153
        Rnettemp = 3.6 * Ri(5);
        Rmda = [Rmda, Rnettemp];
    end
    if 154 <= day(i) && day(i) <= 183
        Rnettemp = 3.6 * Ri(6);
        Rmda = [Rmda, Rnettemp];
    end
    if 184 <= day(i) && day(i) <= 214
        Rnettemp = 3.6 * Ri(7);
        Rmda = [Rmda, Rnettemp];
    end
    if 215 <= day(i) && day(i) <= 365
        Rnettemp = 3.6 * Ri(7);
        Rmda = [Rmda, Rnettemp];
    end
end
Rmd = Rmda';
```

Next calculate average daily wind speed at 2m above the surface. Our weather data includes the average wind speed, however as it was measured at about 9m, we have to run a conversion. We got the conversion formula from the UF document.

```

windms = Wind*0.44704;
windconverter = (4.87/(log(67.8*8.3536-5.42)));
Wind2 = windms*windconverter;

```

Using the mean daily temperature data, we previously calculated we calculate the slope of the vapor saturation pressure curve. Again, we use the UF paper for the formula.

```

slope_vapor_pressure_curve = (4098*(.6108*exp((17.27*Tmean)/(Tmean+237.3)))/((Tmean + 237.3).^2);

```

As we have mean daily pressure we do not use their standardized pressure calculation, we do however convert the inches Hg that our weather data is in to kPa that their guide uses. Once pressure has been calculated, use it to calculate the Psychrometric constant.

```

Pressure = p*3.38639;
Psychrometric_constant = .000665*Pressure;

```

Next, we solve the Delta Term, a calculation for the radiation related ET. TO solve we use the slope of the saturation vapor pressure curve, the psychrometric constant, and the wind speed at two meters.

```

Delta_term = slope_vapor_pressure_curve/(slope_vapor_pressure_curve + Psychrometric_constant.*(1+.34*Wind2));

```

After that, we calculate the Psi Term (PT), and the Temperature Term (TT), both are for the Wind related ET calculation. These are based on the psychrometric constant, slope of the saturation vapor pressure curve, wind speed at 2m, and mean daily temperature.

```

PT = Psychrometric_constant/(slope_vapor_pressure_curve + Psychrometric_constant.*(1+0.34*Wind2));
TT = (900./(Tmean + 273)).*Wind2;

```

Next, we calculated the mean saturation vapor pressure. We do this by calculating the mean with max temperature and with min temperature, then averaging the two values.

```

Emin = .6108*exp((17.27*Tmin)/(Tmin+237.3));
Emax = .6108*exp((17.27*Tmax)/(Tmax+237.3));
Es = (Emax+Emin)/2;

```

Next we calculate the actual vapor pressure (AVP) using RHmax, RHmin, Emax, and Emin.

```

AVP = Emin.*(RHmax/100)+Emax.*(RHmin/100))/2;

```

Next we calculated the inverse relative distance between the Earth and the Sun, the solar declination, the latitude in radians, and the sunset hour angle. To do this we use the latitude in degrees and the day of the year.

```

Dr = 1+.033*cos(((2*pi)/365)*day);
declination = 0.409*sin(((2*pi)/365)*day-1.39);
lat_dec_deg = 36.75;
lat_rad = pi/180*lat_dec_deg;
sunset_hr_angle = acos(-tan(lat_rad)*tan(declination));

```

Next we calculated for extraterrestrial, clear sky, and albedo value radiation based on the solar constant, the inverse relative distance (calculated above), the sunset hour angle, the latitude in radians, elevation, and solar declination. To calculate the albedo value we calculated for it based on the value of straw, the value of soil, and the percent surface covered by each material.

```

Rextra = 1440/pi*.082*Dr.*(sunset_hr_angle*sin(lat_rad).*sin(declination))+(cos(lat_rad)*cos(declination).*sin(sunset_hr_angle));

```



```

Rclearsky = (.75+(2*(10^(-5)))*2163)*Rextra;
Bare = xlsread('imagedata.xlsx','Sheet1','B2:B14');
Straw = xlsread('imagedata.xlsx','Sheet1','C2:C14');
if Bare(m) == 0
    Albedo = .28;
end
if Straw(m)==0
    Albedo = .15;
end
if Bare(m)~= 0 && Straw(m)~= 0
    Albedo = .15*Bare(m) + .28*Straw(m);
end

```

Next, we calculated net shortwave radiation, net longwave radiation, and net radiation. We calculated these with the albedo value, mean daily radiation, Stefan Boltzmann constant, maximum daily temperature in kelvins, minimum daily temperature in kelvins, actual vapor pressure, and clear sky radiation.

```

Rnetshort = (1-Albedo)*Rmd;
if Rclearsky > Rmd
    Rnetlong = (4.903*10^-9)*(((Tmax+273.16).^4+(Tmin+273.16).^4)/2).*(.34-.14*sqrt(AVP)).*((1.35*Rmd./Rclearsky)-.35);
else
    Rnetlong = (4.903*10^-9)*(((Tmax+273.16).^4+(Tmin+273.16).^4)/2).*(.34-.14*sqrt(AVP)).*(1.35-.35);
end
Rnet = .408*(Rnetshort-Rnetlong);

```

Finally, we calculated the daily total ET. We did this by first calculating the solar radiation related ET, then the wind related ET. To make the calculations we used the delta term (DT), net radiation, psi Term (PT), temperature term (TT), actual vapor pressure (AVP), and mean saturation vapor pressure derived from air temperature.

```

ETrad = Delta_term.*Rnet;
ETwind = PT.*TT.*(Es-AVP);
ET=ETrad+ETwind;

```

Results:

Our model output each depth of straw into a separate excel file, each year was output on a separate sheet. We then solved the average daily ET for each year at each depth and put it into a spreadsheet (Appendix 2, the full outputs would have been 1,600 pages and as a result we did not include them in the report). Using this data we made a scatterplot with each dept and year, with the dept as the independent variable, and the daily ET as the y (scatterplot with line of best fit in Appendix 3). After finding the line of best fit for the scatterplot, we then solved for the minimum price possible considering \$.342 per inch per square meter, as the cost of the straw. Our initial run of the model did not work due to an error in the model. However after re running the model we solve it:

$$C_w = 0.17129452291(.2549x^2 - .8228x + 5.8991)$$

$$C_m = .03x$$

$$C = C_w + C_m$$

$$C = 0.17129452291(.2549x^2 - .8228x + 5.8991) + .03x$$

$$C' = \frac{d}{dx} [0.17129452291(.2549x^2 - .8228x + 5.8991) + .03x]$$

$$C' = 0$$

$$0 = \frac{d}{dx} [0.17129452291(.2549x^2 - .8228x + 5.8991) + .03x]$$

$$x = 1.27$$

At this optimum we estimate that the mulch could save \$.012 per meter. Considering that the average farm in the US is 442 acres, this mulch could save more than 20,000 dollars each year.

Validation of Results:

We used weather data and followed exactly a guide to calculating ET, based on an equation that is the standard way to estimate ET. This standard was developed by a subset of the American Society of Civil Engineering. Furthermore, we validated albedo values by measuring the straw with color summarizers and a solar panel, where we could read the radiation that the surface was exposed to and the radiation reflected from the surface. As an albedo value is the reflected light divided by the exposed light we could validate the calculations.

Error Propagation:

There is some slight error in the albedo values starting at two inches. This is because at that point the color summarizer no longer detected the green in the paper backing, even though we could still see it in the pictures, however the change appears to be minimal. Furthermore, some years were missing a few days, however we still used them. This could have off put some of the declination related calculations. Additionally, our time frame was a little larger than the standard growing season (although some research suggests that organic mulch can increase the season), and our model did not account for mulch decay through the year. Further errors could exist in the wind related ET because of surface resistance changes.

Conclusions:

Based on our model and calculations the optimal depth of straw mulch in Taos New Mexico is 1.27 inches or 3.23 centimeters. At this mulch level farms would save 1.2 cents per square meter (including the cost of the mulch). This may seem like a small amount, however the average farm in the US is 442 acres and this would amount to about \$49 per acre. This adds up to an estimated savings of 21,500 each year. This may be a little smaller due to the expense of mulching, however in addition to saving money in water, as we explained earlier, mulch can increase crop yield. Based on our research we believe that it is worth it for farmers to mulch their fields.

Recommendations:

Future research should be conducted taking wind into account. Furthermore, a more thorough model should be developed for the rates, including better data on the albedo values of the straw at any given depth, and the change in surface resistance. Further models should be made on mulches impact on rainwater retention. Additionally real world testing should be conducted on a similar system to see the total impact.

Acknowledgements:

Ms. Galligan helped us as our sponsoring teacher. She made her time and class room available to us as we needed. She also encouraged us regularly and made sure that we completed

the project. Further thanks to Mr. Rael who helped with the modeling and the science side of the project.

References:

1. Weather History for KSKX - April, 1993. (n.d.). Retrieved April 02, 2018, from https://www.wunderground.com/history/airportfrompws/KSKX/1993/4/1/CustomHistory.html?dayend=31&monthend=10&yearend=1993&req_city=&req_state=&req_statename=&reqdb.zip=&reqdb.magic=&reqdb.wmo=
2. Edwards, L., Burney, J., & DeHaan, R. (1995). Researching the effects of mulching on cool-period soil erosion control in prince Edward Island, Canada. *Journal of Soil and Water Conservation*, 50(2), 184+. Retrieved from http://go.galegroup.com/ps/i.do?p=AONE&sw=w&u=nm_p_elportal&v=2.1&it=r&id=G ALE%7CA116791745&asid=09796408ad388d8ad4177d1d0e6ba9a1
3. Zotarelli, Lincoln, et al. “Step by Step Calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method).” *University of Florida IFAS Extension*, University of Florida, Feb. 2010, edis.ifas.ufl.edu/pdffiles/ae/ae45900.pdf.
4. Agbede, T. M., Adekiya, A. O., & Ogeh, J. S. (2013). Effects of Chromolaena and Tithonia mulches on soil properties, leaf nutrient composition, growth and yam yield. *West African Journal of Applied Ecology*, 21(1), 15+. Retrieved from http://go.galegroup.com/ps/i.do?p=AONE&sw=w&u=nm_p_elportal&v=2.1&it=r&id=G ALE%7CA362966062&asid=daa1088c47a54d890b0a3a451f5ae439
5. Green cultivation with biodegradable plastic mulch: fully biodegradable agricultural mulch films offer special advantages over conventional plastic films. (2014, May). *Plastics Engineering*, 70(5), 24+. Retrieved from http://go.galegroup.com/ps/i.do?p=AONE&sw=w&u=nm_p_elportal&v=2.1&it=r&id=G ALE%7CA368379546&asid=edf8d75ffd9a634b582b4d524b375bbc
6. Agele, S. O., Olaore, J. B., & Akinbode, F. A. (2010). Effect of some mulch materials on soil physical properties, growth and yield of sunflower (*Helianthus Annuus*, L). *Advances in Environmental Biology*, 368+. Retrieved from http://go.galegroup.com/ps/i.do?p=AONE&sw=w&u=nm_p_elportal&v=2.1&it=r&id=G ALE%7CA252944767&asid=7cbb18b7f2821f4fa4882556b5632043

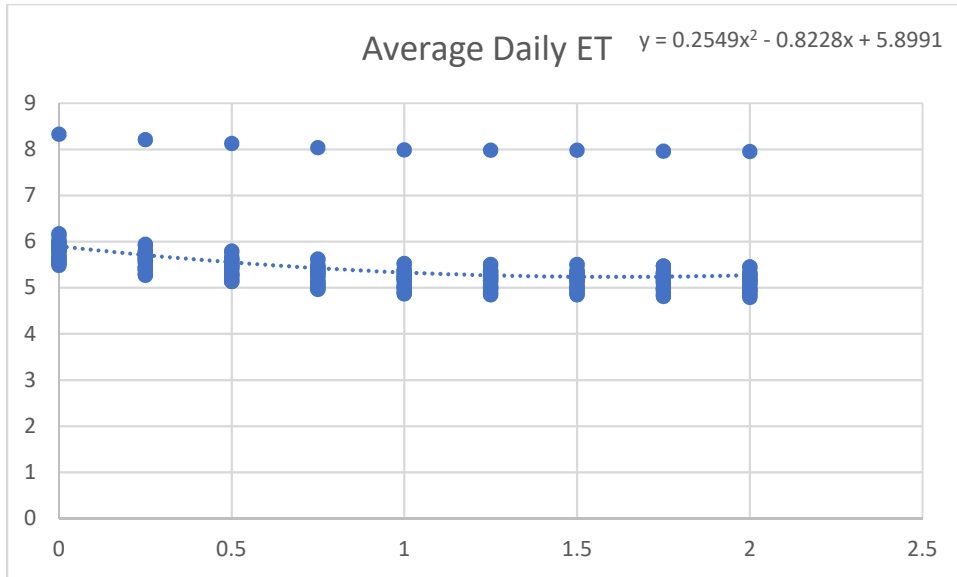
Appendix 1

| | |
|----------|-----------------------------------|
| Δ | Slope of the vapor pressure curve |
| R_n | Net radiation |
| G | Soil heat flux density |
| γ | Psychometric constant |
| T | Mean daily temperature |
| u_2 | Wind speed 2m above the ground |
| e_s | Saturation vapor pressure |
| e_a | Actual vapor pressure |

Appendix 2

4.964146 4.943115 4.943115 4.943115 4.943115 4.943115
5.201636 5.180849 5.180849 5.180849 5.180849 5.180849
5.188505 5.168057 5.168057 5.168057 5.168057 5.168057

Appendix 3



MitoSensor:

A lightweight and wearable sensor that can detect cell signaling, galvanic cell response and mitochondrial toxicity/dysfunction at the cellular level and relay the information back in a timely manner.

New Mexico
Supercomputing Challenge
Final Report
April 23-24, 2018

Team 159
Taos High School

Team Members:

Sierra Ferguson
Delaney Galligan

Team Sponsor:

Tracy Galligan

Table of Contents

1. Executive Summary (2)
2. Research Fast Facts (3-5)
3. In Depth Research (6-32)
4. Experimentation (32-60)
5. Problem Definition (60)
6. Problem Solution (60-61)
7. Software and Code (62-77)
8. Data and End Results (78)
9. Verification and Validation of Code (78)
10. Conclusions (78-79)
11. Achievements (79)
12. Acknowledgements (79)
13. Bibliography (80-82)

Executive Summary

Mitochondrial dysfunction is caused by toxins disrupting the ATP and reproduction processes of mitochondria, the powerhouse of the cells. When mitochondrial homeostasis is disrupted, illness prevails. Once mitochondria enter hyperactive, or stunted glycolysis, the cell cycle becomes unbalanced, causing non-homeostatic programmed cell death. Mitochondria are the key to cellular function and theoretically should be the primary indicator of the presence and progression of diseases. Human epidermal cells have stem cell properties, but can be obtained with non-invasive means for culturing purposes. These cells produce microvolts from the ATP production cycle and are an indicator of deviation from homeostasis. Documenting this mV change can produce a database of cellular conductivity in the presence of numerous stressors and stimuli to gauge mitochondrial function and dysfunction. This, in theory, can be used to build an early indication system of disease.

The first phase of this project was dedicated to building a database. The second phase of this project was dedicated to seeing if the conductivity of cells that are stressed by toxins will increase due to an immune response and the viable cells' attempt to work harder against the toxin and to build a sensor that could read these changes in conductivity. Both phases were completed through laboratory experimentation over several years. At this point, or in the third phase, we are solely working on the model that supports the theory behind this project from both a visual and data processing aspect. The model was completed using NetLogo and will continue to be added to as further research is conducted and data is gathered.

Background Research-Fast Facts

Mitochondria: A double membrane-bound organelle found in all eukaryotic organisms.

Mitochondria generate most of the cell's supply of adenosine triphosphate (ATP), used as a source of chemical energy

Mitochondrial Toxicity: Condition in which the mitochondria of a body's cells become damaged or decline significantly in number; it commonly occurs as a side effect of certain antiretroviral drugs.

Cellular Respiration: A set of metabolic reactions and processes that take place in the cells of organisms to convert biochemical energy from nutrients into adenosine triphosphate (ATP), and then release waste products such as carbon dioxide.

Epidermal Cells: The cells that are on the surface of the skin, the outer organ that encases the rest of the bodies' organs.

Galvanic Skin Response/Electrodermal Activity: It is a property of the human body that causes a change in the electrical resistance of the skin caused by some sort of stress.

Cell Signaling/Cell Voltage Change: A chemical response method in which one cell sends information to another cell through potassium positive and sodium positive pathways.

Cell Cycle: Series of events that take place in a cell leading to its division and duplication of its DNA to produce two daughter cells.

Programmed Cell Death: There are three types of cell death including, apoptosis, autophagisis, and necrosis. Each has a varying degree of "idleness."

Multimeter: Instrument designed to measure electric current, voltage, and resistance.

Janus Green B: Dye that detects the health of the mitochondria by changing different shades of green. When the mitochondria is working anaerobically the dye is pink.

Rhodamine: Dye that tests cell viability

Ciprofloxacin: Very commonly prescribed to fight against most strains of bacterial pathogens responsible for respiratory, urinary tract, gastrointestinal, and abdominal infections.

Penicillin: Used only for the treatment of mild to moderate infections, and not for severe or deep-seated infections.

Streptomycin: Used against infective endocarditis caused by enterococcus when the organism is not sensitive to Gentamicin, Tuberculosis, Plague and Tularemia infections.

Tetracycline: Commonly used against skin conditions. It is useful against the majority of high level illnesses, infections and parasites

Vancomycin: Used in infections where antibiotic resistance to other antibiotics are prevalent.

Osha: Used for sore throat, bronchitis, cough, common cold, influenza, swine flu, and pneumonia.

Rosemary: Used for digestion problems, including heartburn, intestinal gas (flatulence), liver and gallbladder complaints, and loss of appetite. It is also used for gout, cough, headache, and high blood pressure.

Cayenne Pepper: Consumption dilates the blood vessels and speeds the metabolism due to the high amounts of capsaicin. With consumption the amount of heat the human body puts off is influenced.

Cota: Used to cure ailments ranging from kidney problems, to digestive problems including stomach cramps, and to purify the blood.

Ginger: Used for nausea and arthritis pain.

Garlic: Used for many conditions related to the heart and blood system. These conditions include high blood pressure, high cholesterol, coronary heart disease, heart attack, and “hardening of the arteries.”

Copper: It is a soft, malleable and ductile metal with very high thermal and electrical conductivity.

Volts: The volt is the SI measurement of electric potential, electric potential difference, and electromotive force.

Silver: It possesses the highest electrical conductivity, thermal conductivity, and reflectivity of any metal.

Conductivity/Resistivity: An intrinsic property that quantifies how strongly a given material opposes the flow of electric current. A low resistivity indicates a material that readily allows the flow of electric current.

Kanthal: The trademark for a family of iron-chromium- aluminium alloys used in a wide range of resistance and high-temperature applications.

Nichrome: Nichrome refers to any alloy of nickel, chromium, and often iron and/or other elements or substances. Nichrome alloys are typically used in resistance wire.

In Depth Background Research

Mitochondria

The mitochondrion (plural mitochondria) is a double membrane-bound organelle found in all eukaryotic organisms. Some cells in some multicellular organisms may lack them (mature mammalian red blood cells). A number of unicellular organisms, such as microsporidia, parabasalids, and diplomonads, have also reduced or transformed their mitochondria into other structures. Mitochondria generate most of the cell's supply of adenosine triphosphate (ATP), used as a source of chemical energy. Mitochondria are commonly between 0.75 and 3 μm in diameter but vary considerably in size and structure. Unless specifically stained, they are not visible. In addition to supplying cellular energy, mitochondria are involved in other tasks, such as signaling, cellular differentiation, and cell death, as well as maintaining control of the cell cycle and cell growth. Mitochondrial biogenesis is in turn temporally coordinated with these cellular processes. Mitochondria have been implicated in several human diseases, including mitochondrial disorders, cardiac dysfunction, heart failure and autism. The number of mitochondria in a cell can vary widely by organism, tissue, and cell type. For instance, red blood cells have no mitochondria, where liver cells can have more than 2000. The organelle is composed of compartments that carry out specialized functions. These compartments or regions include the outer membrane, the intermembrane space, the inner membrane, and the cristae and matrix. Although most of a cell's DNA is contained in the cell nucleus, the mitochondrion has its own independent genome that shows substantial similarity to bacterial genomes. Mitochondrial proteins (proteins transcribed from mitochondrial DNA) vary depending on the tissue and the species. In humans, 615 distinct types of protein have been identified from cardiac mitochondria, whereas in rats, 940 proteins have been reported. The mitochondrial proteome is thought to be dynamically regulated (Alberts, Johnson, Lewis, et. al., 2002).

Mitochondrial Toxicity

Condition in which the mitochondria of a body's cells become damaged or decline significantly in number; it commonly occurs as a side effect of certain antiretroviral drugs used to treat human immunodeficiency virus, or HIV, but has also been found in certain other drugs that target proteins, amino acid production, RNA and DNA functions within the body as part of their function, such as some high impact, broad spectrum antibiotics (Alberts, et. al., 2002).

Mitochondrial toxicity leads to disease, illness and potentially death. It is also an emerging factor in drug resistance. The disruption of cell function that accompanies the condition can cause both mild and severe problems in people suffering from mitochondrial toxicity. The most commonly observed symptom is muscle weakness, or myopathy and lactic acidosis, though other ailments including ulcers, lupus and many 'mystery diseases' can be eventually attributed to this syndrome. Other symptoms include peripheral neuropathy (numbness in the fingers and toes) and pancreatitis (inflammation of the pancreas), with the most severe being lactic acidosis, in which a build-up of lactic acid in the tissues of the body leads to loss of energy, organ failure, and eventually death (Alberts, et. al., 2002).

Epidermal Cells

Human epidermal cells are the cells that are on the surface of the skin, the outer organ that encases the rest of the bodies' organs. Together with the dermis these cells form the cutis or skin. The epidermis is a stratified squamous epithelium, composed of proliferating basal and differentiated supra-basal keratinocytes. The epidermis is ectodermal (primary germ cell) in origin. The epidermis is avascular (not directly supplied by blood cells) nourished by diffusion from the dermis. Blood capillaries are found beneath the epidermis. The epidermis serves as a barrier to protect the body against microbial pathogens, oxidant stress and chemical compounds

and provides mechanical resistance. Most of that function is played by the stratum corneum (Alberts, et. al., 2002).

Galvanic skin response/Electrodermal activity

Electrodermal activity (EDA), is the property of the human body that causes continuous variation in the electrical characteristics of the skin. Historically, EDA has also been known as skin conductance, galvanic skin response (GSR), electrodermal response (EDR), psychogalvanic reflex (PGR), skin conductance response (SCR), sympathetic skin response (SSR) and skin conductance level (SCL) (Boucsein, 2012).

The traditional theory of EDA indicates that skin resistance varies with the state of sweat glands in the skin. Sweating is controlled by the sympathetic nervous system, and skin conductance is an indication of psychological or physiological arousal. If the sympathetic branch of the autonomic nervous system is highly aroused, then sweat gland activity also increases, which in turn increases skin conductance. In this way, skin conductance can be a measure of emotional and sympathetic responses. The study of EDA has led to such important and vital tools as the electrocardiograph (ECG or EKG) and the electroencephalograph (Boucsein, 2012).

Human extremities, including fingers, palms, and soles of feet display different bio-electrical phenomena. They can be detected with an EDA meter, a device that displays the change electrical conductance between two points over time. The two current paths are along the surface of the skin and through the body. Active measuring involves sending a small amount of current through the body. Some studies include the human skin's response to alternating current, including recently deceased bodies. There is a relationship between emotional arousal and sympathetic activity, although the electrical change alone does not identify which specific emotion is being elicited. These autonomic sympathetic changes alter sweat and blood flow,

which in turn affects GSR and GSP. The response of the skin and muscle tissue to external and internal stimuli can cause the conductance to vary by several microsiemens or minutely microvolts. A correctly calibrated device can record and display the subtle changes (Boucsein, 2012).

The combined changes between electrodermal resistance and electrodermal potential make up electrodermal activity. Galvanic skin resistance (GSR) is an older term that refers to the recorded electrical resistance between two electrodes when a very weak current is steadily passed between them. The electrodes are normally placed about an inch apart, and the resistance recorded varies according to the emotional state of the subject. Galvanic skin potential (GSP) refers to the voltage measured between two electrodes without any externally applied current. It is measured by connecting the electrodes to a voltage amplifier. This voltage also varies with the emotional state of the subject. A painful stimulus such as a pinprick elicits a sympathetic response by the sweat glands, increasing secretion. Although this increase is generally very small, sweat contains water and electrolytes, which increase electrical conductivity, thus lowering the electrical resistance of the skin. These changes in turn affect GSR. Another common manifestation is the vasodilation (dilation) of blood vessels in the face, referred to as blushing, as well as increased sweating that occurs when one is embarrassed. EDA is highly responsive to emotions in some people. Fear, anger, startled response, orienting response, and sexual feelings are among the reactions that may be reflected in EDA. These responses are utilized as part of the polygraph or lie detector test. EDA in regular subjects differs according to feelings of being treated fairly or unfairly, but psychopaths have been shown to manifest no such differences. This indicates that the EDA record of a polygraph may be deceptive in a criminal investigation (Boucsein, 2012).

EDA is a common measure of autonomic nervous system activity, with a long history of being used in psychological research. Many biofeedback therapy devices utilize EDA as an indicator of the user's stress response with the goal of helping the user to control anxiety. Oftentimes, EDA monitoring is combined with the recording of heart rate, respiratory rate, and blood pressure, because they are all autonomically dependent variables. EDA measurement is one component of modern polygraph devices, which are often used as lie detectors. EDA measurement is also becoming more popular in hypnotherapy and psychotherapy practices for detecting depth of hypnotic trance prior to the commencement of suggestion therapy. When traumatic experiences are recalled by the client, immediate changes in sweat rate can indicate that the client is experiencing emotional arousal (Boucsein, 2012).

External factors such as temperature and humidity affect GSR measurements, which can lead to inconsistent results. Internal factors such as medications can also change GSR measurements, demonstrating inconsistency with the same stimulus level. Lastly, galvanic skin responses are delayed 1–3 seconds (Boucsein, 2012).

Cell swelling is an indicator of sub-lethal cell damage, where the cell's membranes swell outward as certain substances enter the cell through osmosis, and cause an imbalance of the cell structure. The cell can be recovered, or it can enter apoptosis from this stage. The swelling is an indicator of the accumulation of electrolytes or inducers within the cell beyond the cell's normal capacity or functioning point. Apoptosis is generally signaled by cell shrinkage or cell swelling (NCI, 2011).

Cell Signaling/Cell voltage change

Cell Signaling is a chemical response method in which one cell sends information to another cell through K^+ Na^+ pathways. This is also how cells communicate with each other,

particularly when they are going through cell cycle changes or environmental stimuli changes. The mitochondria in their ATP production process communicate to other cells through cell membrane potential, a chemically reactive positive or negative charge that goes through one membrane to another. This also produces the noticeable cell voltage, through energy production from the electrolytes produced with ATP production. A higher level of resistance indicates homeostasis within the cell pathways and the stability of glycolysis, while a higher rate of conductivity shows instability in the cell cycle and pathways, therefore an interruption in the conductivity, through an excess of Na^+ processing (Alberts, Johnson, Lewis, et. al., 2002).

Cell Cycle

The human body relies on DNA within the nucleus of a cell to send out instructions for body functions and body energy usage. A huge amount of information resides within the DNA and through cellular reproduction processes can separate genetic material and share it via cell division. These methods are called mitosis and meiosis. Mitosis is a process that produces two cells, each of which is identical to the original parent cell. Mitosis is preceded by replication of the cell's DNA so each 'daughter cell' will have a full amount of genetic material. Traditionally in animals, mitosis is used for growth and repair of somatic body cells. This generally leads to asexual reproduction. Meiosis produces four cells from an original parent cell that is not identical to the parent cell, and only has half the parent cell's genetic matter. This is traditionally known as sexual reproduction. (Alters and Alters, 2008)

The cell cycle is the life cycle of the cell. These include prophase, pro-metaphase metaphase, anaphase telophase, interphase and cytokinesis. The cell cycle spends the most amount of time in interphase. This is the part of the cell cycle that includes cell growth, replication of cell organelles, replication of DNA, assembly of the parts of mitosis and the

condensation of DNA. This phase is subdivided into stages G₁, S and G₂. The G₁ stage is the time gap between the last cell division and the start of DNA replication, during which time the cell is growing. This growth period occupies the major portion of the cell's lifespan, where the cell doubles its size and carries out its normal life functions. S (Synthesis) stage produces a complete replica of the cell's DNA for cell division. By the end of this stage, the cell contains two complete, identical copies of the hereditary information. The G₂ stage signifies the time gap between the end of DNA replication and the beginning of cell division. The coils of DNA condense into tightly compacted masses that become visible chromosomes during mitosis. Each chromosome contains two copies of hereditary information in sister chromatids, connected by a centromere. Mitosis is a continuous sequence of events that occurs just after interphase, resulting in the separation of the sister chromatids. (Alters and Alters, 2008)

The first phase of mitosis, prophase is when the chromosomes have condensed. As prophase continues, the chromosomes continue to shorten and thicken. The nucleolus disappears because the cell is no longer capable of producing ribosomal RNA (rRNA). The microtubules (thin tubes of protein structures) are formed. The centrosome is the area in which these microtubules are organized. The centrosomes of animal cells begin to move away from each other in the beginning and by the end, each member of the pair has moved to an opposite end or pole of the cell. As the spindle fibers formulate, the nuclear membrane breaks down and the spindle fibers build a bridge between centrosomes from one pole to another. The spindle fibers then attach to kinetochores. This connection is critical to the separation and movement of sister chromatids during later stages of mitosis. The next stage of animal cell mitosis is metaphase. This occurs when the pairs of sister chromatids align in one plane at the center of the cell. This will indicate where the future plane of the cell division will be. After metaphase comes anaphase.

In this phase, the sister chromatids are pulled in two different directions simultaneously. The chromatids separate into chromosomes and move rapidly toward opposite poles of the cell. They are pulled by the kinetochore by attached and shortening microtubules. This separation produces duplicate sets of hereditary material. The final phase, telophase is the preparation of the cell for cytokinesis. The spindle fibers are chemically disassembled and disappear. The nuclear envelope reforms around each set of chromosomes, which begin to uncoil, and the nucleolus reappears as rRNA. This phase is like prophase in reverse order. (Alters and Alters, 2008)

Mitosis is complete after telophase. The process of cell division is not. The portion of the cell outside the nucleus; cytoplasm is divided starting in telophase and completed in cytokinesis, where the separation of one cell into two takes place. In animal cells, this occurs by pinching the cell in two by a belt of microfilaments encircling the cell at the metaphase stage. As the microfilaments contract, a cleavage furrow appears around the circumference of the cell. As the contraction proceeds the furrow deepens until the opposing edges of the cell membrane make contact with one another. The membranes fuse, producing the cell separation. (Alters and Alters, 2008)

Programmed Cell Death

Cancer and cellular toxicity causes mitochondria to be ‘reprogrammed’, and in the case of tumorigenic activity, the cells are encapsulated and oxygen is cut off, therefore forcing the mitochondria to work in anaerobic conditions, forcing glycolysis. Glycolysis is a series of biochemical reactions from which one molecule of glucose is oxidized to two molecules of pyruvic acid and a small amount of ATP. Increased activity in the glycolytic pathway is an indicator of disease in humans. Malignant, rapidly-growing tumor cells have glycolytic rates that are up to 200 times higher than those of their normal tissues of origin. (NCI, 2012)

Mitochondria are also suspected to play a role in the aging process. Once the mitochondria are forced into glycolysis, they continue to feed cancerous mutations. If these reprogrammed mitochondria are shut off or eliminated, then glycolysis can be shut down, and the surrounding cells can either escape the cancerous mutation cycle, or can be remediated. The mitochondria can also be re-programmed to return to the production of ATP as a primary energy source, which cancerous material is not as prone to reproduce under. Cancerous material feeds off of sugar based chemicals for their energy source and relies heavily on carbohydrates, to reproduce biochemically. Certain chemicals interrupt these carbohydrate and anaerobic cycles, forcing oxygen into the reaction, and balancing off any free radical reactions (American Cancer Society, 2012).

Genetically programmed cell death (apoptosis, PCD), is one of the many concepts that is used to treat cancer. Some cells of the human body self-destruct after a limited lifespan, while others are programmed to last the lifetime of the organism. Those that self-destruct are replaced by the body with new cells produced from cell division of the survivors. Cancer is an indicator of a disruption of the cell's life cycle, where cells that should die do not and develop from failure of programmed cell death and divide uncontrollably. Apoptosis is a type of PCD in which cell suicide is pre-programmed; the cell membrane remains intact as the cell dies so that it does not release its contents and trigger a local inflammatory reaction. The dying cell splits into small membrane-bound bodies that are engulfed and digested by white blood cells. Many compounds have been tested as apoptosis agents (Alters and Alters, 2008).

Programmed Cell Death or PCD has had three main pathways identified: Type I: Apoptosis, Type II: Autophagisis and Type III: Necrosis. Apoptosis, as described above, is cell death produced by biochemical changes that lead to morphology changes and death. These

morphology changes include blebbing, (irregular bulge in the plasma membrane caused by localized decoupling of the cytoskeleton from the plasma membrane) membrane cell shrinkage, nuclear fragmentation, chromatin condensation, and chromosomal DNA fragmentation. Apoptosis occurs naturally in the cell cycle, including in the development of body parts. Apoptosis is responsible for appendages such as fingers and toes to be separated, and can either lead to mutagenic qualities or death of the cell. This process produces cell fragments (apoptotic bodies) that phagocytic cells engulf and remove before the contents of the cell can spill out and cause damage in surrounding cells. The process of apoptosis can be controlled by a large and diverse range of cell signals that can originate extracellularly (extrinsic inducers) which include toxins, hormones, growth factors, nitric oxide or cytokines or intracellularly (intrinsic inducers). These inducers must either cross the plasma membrane or transducer to produce a response; the positive or negative triggering of apoptosis. The cell will initiate intracellular apoptotic signaling in response to stresses, which can potentially lead to a 'cell suicide'. The cell death is precipitated by enzymes, produced by the apoptotic signals that cause regulatory proteins to initiate apoptosis, which allows the selection of apoptosis in certain cells, or can stop the process should the cell no longer need to die. These two main methods of regulation are targeting mitochondria functionality, or transducing the signal through adaptor proteins to apoptotic mechanisms. Apoptosis occurs very quickly (NCI, 2011).

Viral proteins can cause intracellular stress which induces apoptosis. This response to internal stimuli causes a caspase cascade which in turn triggers a series of morphological changes within the cell. Once the caspase cascade has started, it is an irreversible process resulting in certain cell death. Apoptosis in HeLa cells is inhibited by proteins produced by the cell. Inhibitory proteins target retinoblastoma tumor suppressing proteins. These tumor

suppressing proteins regulate the cell cycle, but are rendered inactive when bound to an inhibitory protein (Oyagbemi, et. al., 2010).

Janus Green B

Janus Green B $C_{30}H_{31}ClN_6$ is a basic dye and vital stain used in histology. It is also used to stain mitochondria supravivally. The indicator Janus Green B changes color according to the amount of oxygen present. When oxygen is present, the indicator oxidizes to a blue color. In the absence of oxygen, the indicator is reduced and changes to a pink color. Janus Green B is reduced to diethyl safranin and to leucosafranin by the lactic dehydrogenase and the glucose dehydrogenase enzyme systems. Reduced flavoprotein is the immediate reactant which carries out this reduction and therefore any DPN-flavoprotein enzyme system should be capable of carrying out this reduction (Chemspider, 2016).

Rhodamine B

Rhodamine B is a red or pink fluorescent dye that is commonly used as a tracer dye within water to determine the rate and direction of flow and transport. These dyes fluoresce and can be detected easily and inexpensively. These dyes are used extensively in biotechnology applications including fluorescence microscopy, and other visual highlighting applications. Rhodamine B is used in biology as a staining dye for cells and bacterium. Its luminescence quantum yield is 0.65 in basic ethanol. The fluorescence yield is temperature dependent. Rhodamine B is also used as a biomarker for rabies in animals, and as an indicator of where pesticides have been used. Rhodamine also attaches to matter that processes oxygen and is still considered living, but will also attach to a free radical or free radical affected matter, making it darker and staining it a red color. It will not chemically impact or interact with other chemicals,

being a neutral chemical, and being a tracer, so it will not have a chemical impact on the cancer cell layer or tumor. (NCI. 2011)

The Fluorescence Absorption Test for Cell Viability works on the concept that a live cell will still undergo the different cell stages, and will take in certain nutrients from the environment and absorb them for use via the cell wall. Fluorescent dyes have a small enough particulate structure as to be able to absorb the dye in the cell if it is living or viable, and will simply show up on the outside of the cell wall if the cell is dead. Fluorescent dyes will also attach to the inner ‘dark matter’ of the cancer cell making it more apparent under the microscope, but will also highlight the transparency of a clear or healthy cell. The Fluorescence Absorption Test for Viability is being used in some exploratory surgeries to highlight cancerous cells while still in the body to be able to differentiate from the healthy cells, these are also used as tracers to attempt to determine which cells are diseased and which are healthy in endoscopic procedures. (Estevez, 2009)

| |
|---|
| <u>Ciprofloxacin</u> |
| Antibiotic part of a group of drugs called fluoroquinolones. (2 nd generation) |
| Very commonly prescribed. |
| Is commonly used against most strains of bacterial pathogens responsible for respiratory, urinary tract, gastrointestinal, and abdominal infections (including E. coli, S. epidermis, anthrax) |
| Is a broad spectrum antibiotic with good tissue penetration. |
| Is commonly used in situations where the infection agent is not identified. |
| Works on both gram positive and gram-negative bacteria. |
| 35 th most commonly prescribed drug, 5 th most prescribed antibacterial. |
| Was recently granted a ‘black box’ warning for the number of serious side effects produced, including mitochondrial toxicity and related symptoms. |
| Side effects include: increased risk of tendon damage and/or rupture and for exacerbation of muscle weakness in patients with the neurological disorder myasthenia gravis, and other muscular related ailments, potential seizures, nausea, vomiting, diarrhea and dizziness. |
| Chemical name: 1-cyclopropyl-6-fluoro-4-oxo-7-(piperazin-1-yl)-quinoline-3-carboxylic acid: C17H18FN3O3 |

| |
|---|
| Can develop high resistance with many different bacteria. |
| Reports to have a half-life of 4 hours, but high level traces can be found months after oxidation. |
| Are not supposed to be recommended for respiratory infections. |
| Ciprofloxacin should not be taken with antacids containing magnesium or aluminum, highly buffered drugs, or with supplements containing calcium, iron, or zinc. |
| Inhibits the drug-metabolizing enzyme CYP1A2 and thereby can reduce the clearance of drugs metabolized by that enzyme. |
| Functions by inhibiting DNA gyrase, (a type II topoisomerase) and topoisomerase IV, enzymes necessary to separate bacterial DNA, inhibiting cell division. |
| Was developed to replace other antibiotics that had already become active in antibiotic resistance. |
| Pathogens, including enterococci, Streptococcus pyogenes and Klebsiella pneumoniae (quinolone-resistant) now exhibit resistance. Some Burkholderia cepacia, Clostridium innocuum and Enterococcus faecium strains have developed resistance to ciprofloxacin in varying degrees |
| Can also be considered a potential treatment for breast cancer and other reproductive system cancers. |
| Induces apoptotic activity in cells and works primarily from deconstructing the outer cell walls of the invading bacterial structure. |

| |
|---|
| <u>Penicillin</u> |
| Phenoxyethylpenicillin, commonly known as penicillin V. (oral use) |
| Exerts a bactericidal action against penicillin-sensitive microorganisms during the stage of active multiplication. |
| Acts by inhibiting the biosynthesis of cell-wall peptidoglycan |
| Not active against beta-lactamase-producing bacteria, which include many strains of Staphylococci |
| 3,3-Dimethyl-7-oxo-6-(2-phenoxyacetamido)-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid. Formula: C ₁₆ H ₁₈ N ₂ O ₅ S |
| Primarily used against gram positive bacteria, but is minorly effective against gram negative bacteria. |
| Used only for the treatment of mild to moderate infections, and not for severe or deep-seated infections since absorption can be unpredictable. Primarily used for: Infections caused by Streptococcus pyogenes (incl. Tonsillitis, Pharyngitis, Skin infections), Anthrax (mild uncomplicated infections), Lyme disease (early stage in pregnant women or young children), Rheumatic fever (primary and secondary prophylaxis), Streptococcal skin infections, Spleen disorders (pneumococcal infection prophylaxis), Initial treatment for Dental Abscesses, Moderate-to-severe gingivitis, Avulsion injuries of teeth, Sickle Cell Disease of the blood, odontogenic infections. |
| High level of protein binding. |
| Has a short half-life. |
| Is mostly filtered before excretion, has a higher level of metabolism. |
| Is not as effective as stronger antibiotics. |

May occasionally cause transient nausea, vomiting, epigastric distress, diarrhea, and black hairy tongue.

Streptomycin

Antimycobacterial = type of drug used to treat Mycobacteria infections.

Aminoglycoside, derived from the actinobacterium *Streptomyces griseus*. Bactericide.

The first effective treatment for tuberculosis.

Adverse effects of this medicine are ototoxicity, nephrotoxicity, fetal auditory toxicity, and neuromuscular paralysis. Fever, rashes, tinnitus, vertigo and ataxia, nephrotoxicity and kidney malfunction are amongst the side effects of this drug.

5-(2,4-diguanidino-3,5,6-trihydroxy-cyclohexoxy)- 4-[4,5-dihydroxy-6-(hydroxymethyl)-3-methylamino-tetrahydropyran-2-yl] oxy-3-hydroxy-2-methyl-tetrahydrofuran-3-carbaldehyde Formula: C₂₁H₃₉N₇O₁₂

Used against: Infective endocarditis caused by enterococcus when the organism is not sensitive to Gentamicin, Tuberculosis, Plague (*Yersinia pestis*) & Tularemia infections.

Can easily develop resistance.

20% may not be metabolized and may be excreted whole.

Is used as a pesticide, to combat the growth of bacteria, fungi, and algae.

in combination with penicillin, is used in a standard antibiotic cocktail to prevent bacterial infection in cell culture.

Can cause mitochondrial toxicity.

Is a protein synthesis inhibitor. It binds to the small 16S rRNA of the 30S subunit of the bacterial ribosome, interfering with the binding of formyl-methionyl-tRNA to the 30S subunit.

Inhibits gram positive and gram negative bacteria.

Tetracycline

Broad-spectrum polyketide antibiotic (class of secondary metabolites produced by certain living organisms in order to impart to them some survival advantage, highly active).

Produced by the *Streptomyces* genus of Actinobacteria.

Protein synthesis inhibitor.

Commonly used to treat skin conditions. But was also used to treat cholera. It is first-line therapy for rocky mountain spotted fever (*Rickettsia*), Lyme disease (*B. burgdorferi*), Q fever (*Coxiella*), psittacosis and lymphogranuloma venereum (*Chlamydia*), *Mycoplasma pneumoniae* and to eradicate nasal carriage of meningococci, acne. Also used against the plague. Used as a biomarker. Shuts off ulcers and some cancers, against mosquito repopulation.

C₂₂H₂₄N₂O₈

Gram positive and gram negative.

Is useful against the majority of most high level illnesses, infections and parasites.

| |
|---|
| Bacteria acquire resistance to tetracycline from horizontal transfer of a gene that either encodes an efflux pump or a ribosomal protection protein. |
| Side effects include: Stain developing teeth (even from mother to child), Discolor permanent teeth (yellow-gray-brown), from infancy and childhood to eight years old, Be inactivated by Ca ²⁺ ion, so are not to be taken with milk, yogurt, and other dairy products, Be inactivated by aluminium, iron and zinc, not to be taken at the same time as indigestion remedies, cause skin photosensitivity, cause drug-induced lupus, and hepatitis, cause micro-vesicular fatty liver, tinnitus, interfere with methotrexate by displacing it from the various protein binding sites, cause breathing complications, anaphylactic shock, affect bone growth of the fetus, Fanconi syndrome, Stevens–Johnson syndrome, toxic epidermal necrolysis and erythema multiform associated with doxycycline use. |
| Inhibits protein synthesis by blocking the attachment of charged aminoacyl-tRNA to the A site on the ribosome. Tetracycline binds to the 30S subunit of microbial ribosomes. It prevents introduction of new amino acids to the nascent peptide chain. |

| |
|--|
| <u>Vancomycin</u> |
| Glycopeptide antibiotic. |
| (1S,2R,18R,19R,22S,25R,28R,40S)- 48- {[[(2S,3R,4S,5S,6R)- 3- {[[(2S,4S,5S,6S)- 4- amino- 5- hydroxy- 4,6- dimethyloxan- 2- yl]oxy}- 4,5- dihydroxy- 6- (hydroxymethyl)oxan- 2- yl]oxy}- 22- (carbamoylmethyl)- 5,15- dichloro- 2,18,32,35,37- pentahydroxy- 19- [(2R)- 4- methyl- 2- (methylamino)pentanamido]- 20,23,26,42,44- pentaoxo- 7,13- dioxo- 21,24,27,41,43- pentaazaocyclo[26.14.2.23,6.214,17.18,12.129,33.010,25.034,39]pentaconta- 3,5,8(48),9,11,14,16,29(45),30,32,34,36,38,46,49- pentadecaene- 40- carboxylic acid Formula: C ₆₆ H ₇₅ Cl ₂ N ₉ O ₂₄ |
| Effective mostly against Gram-positive bacteria |
| Naturally occurring antibiotic made by the soil bacterium <i>Amycolatopsis orientalis</i> |
| Comparatively rare halo-organic compound, containing 2 covalently bonded chlorine atoms. |
| Bacteria were relatively slow to acquire vancomycin resistance |
| Used in infections where antibiotic resistance to other antibiotics were prevalent. |
| First-line treatment for complicated skin infections, bloodstream infections, endocarditis, bone and joint infections, intestinal infections, and meningitis caused by methicillin-resistant <i>S. aureus</i> . |
| Has developed major resistances to many bacteria. |
| Excreted unchanged, does not metabolize well. |
| Side effects include: pain, thrombophlebitis, kidney and hearing damage, anaphylaxis, toxic epidermal necrolysis, erythema multiforme, red man syndrome, superinfection, thrombocytopenia, neutropenia, leukopenia, florid petechial hemorrhages, ecchymoses, wet purpura tinnitus, and dizziness and/or ototoxicity. |
| Vancomycin biosynthesis occurs via different nonribosomal protein synthases (NRPSs). |
| Vancomycin acts by inhibiting proper cell wall synthesis in gram-positive bacteria. |

Osha

Osha is a plant. Historically, the root has been used as medicine by Native American and Hispanic cultures. Today, osha is used for sore throat, bronchitis, cough, common cold, influenza, swine flu, and pneumonia. It is also used to treat other viral infections including herpes and AIDS/HIV. Some people use it for indigestion. Some people apply osha directly to the skin to keep wounds from getting infected.

Osha can be confused with the very poisonous plant hemlock. The leaves of the two plants are very similar. Osha must be identified by the root, which people say has an unpleasant celery-like odor. Osha grows at higher elevations in the western US and is difficult to cultivate. The popularity of osha has led to over-harvesting of the wild plant. As a result, osha has been designated an endangered plant by conservationists. Osha contains chemicals that might help fight bacterial and viral infections (Marshner, 2012)

Rosemary

Rosemary is an herb. Oil is extracted from the leaf and used to make medicine. Rosemary is used for digestion problems, including heartburn, intestinal gas (flatulence), liver and gallbladder complaints, and loss of appetite. It is also used for gout, cough, headache, and high blood pressure. Some women use rosemary for increasing menstrual flow and causing abortions (Marshner, 2012)

Rosemary is used topically (applied to the skin) for preventing and treating baldness; and treating circulation problems, toothache, a skin condition called eczema, and joint or muscle pain such as myalgia, sciatica, and intercostal neuralgia. It is also used for wound healing, in bath

therapy (balneotherapy), and as an insect repellent. In foods, rosemary is used as a spice. The leaf and oil are used in foods, and the oil is used in beverages. In manufacturing, rosemary oil is used as a fragrant component in soaps and perfumes. Although it's not clear how rosemary works for hair loss, applying it to the scalp irritates the skin and increases blood circulation (Marshner, 2012).

Cayenne pepper

Cayenne pepper, is a cultivar of *Capsicum annuum*, which is related to bell peppers, jalapeños, paprika, and others. The *Capsicum* genus is in the nightshade family (Solanaceae). It is a hot chili pepper used to flavor dishes. The fruits are generally dried and ground, or pulped and baked into cakes, which are then ground and sifted to make the powdered spice of the same name. Cayenne is used in cooking spicy dishes, as a powder or in its whole form (such as in Korean, Sichuan, and other Asian cuisine), or in a thin, vinegar-based sauce. It is generally rated at 30,000 to 50,000 Scoville units. It is also used as an herbal supplement. Cayenne pepper, by weight, is high in vitamin A. It also contains vitamin B6, vitamin E, vitamin C, riboflavin, potassium, and manganese. However, given the very small amount of cayenne pepper typically consumed in a serving, it makes a negligible contribution to overall dietary intake of these nutrients (Westerterp-Plantenga, et. al, 2009).

Cayenne pepper consumption dilates the blood vessels and speeds the metabolism due to the high amounts of capsaicin. With the consumption of cayenne peppers, the amount of heat the human body puts off is influenced. In animal studies, capsaicin has the ability to boost metabolism, which in turn causes weight loss. This increases circulation and blood flow to all major organs, facilitating oxygen and nutrient delivery. Capsaicin may support a healthy energy balance while suppressing appetite. Capsaicin has been shown to increase energy expenditure, so

acts as a metabolism booster and is beneficial in long-term weight loss. A correlation has been shown between substrate oxidation and capsaicin. Capsaicin treatment sustained fat oxidation during weight maintenance, but did not affect weight regain after modest weight loss. Cayenne pepper is also claimed to be an aphrodisiac because it contains capsaicin. It has also been shown to aid in the oxidation of adipose tissue, regulate high blood pressure, promote healthy liver function and tissue production, help regulate the digestive system, and promote healthy mucus production in the membranes that line internal organs (Westerterp-Plantenga et. al., 2009.)

Cota

Cota is the Spanish word for Hopi or Navajo tea, which is also called greenthread due to the fact the stems look like “little green threads”. Cota is a species of flowering plant in the aster or *asteraceae* family, relating it to daisies and sunflowers. It is native to most parts of the Americas, including the central and western United States, where it grows in many different types of habitat. The herb grows over much of the plains and mountain states, reaching up to Wyoming, Montana, Nebraska, and South Dakota. It is centralized to the southwest, particularly the Hopi and Navajo nations, primarily through New Mexico, Colorado and Arizona. The extended range of habitat is made possible by the fact that several species exist and the growth conditions between the different species vary from high altitude to the plains. Another term to reference this plant includes ‘roadside tea’ due to the fact that the plant can be found near the road side. The herb known as Navajo tea is derived mainly from *T. megapotamicum*, *T. simplicifolium*, *T. filifolium*, and *T. subnudum*. Its taxonomy is as follows: Magnoliophyta (angiosperm), Magnoliopsida (dicot), Asteraceae (Compositae). This perennial is also known as *T. gracile* (Kane, 2011).

Cota is primarily a medicinal herb, which has been used to cure ailments ranging from kidney problems, to digestive problems (including stomach cramps), and to purify the blood. Cota is considered useful for the Kidneys, used against gonorrhea, and as a nervous stimulant. One of the more common uses of Cota is as a toothache remedy and against mouth sores (PDR, 2007). It has been used as a cure against breast cancer at the genetic level, and has the potential for restoring cell cycle damage in other parts of the body against various cancers (Figueroa, 2012).

What makes Cota so effective as a medicinal remedy are the phytochemicals and antioxidants within the plant, which has not been greatly analyzed but has been concluded to contain flavonoids including luteolin, the plant also contains chalcone glycosides, notably marein (Ateya, et. al., 1982). Cota is also used as a fabric dye, particularly in rugs to give the thread a particularly bright yellow hue. The yellow of the flowers is primarily due to the flavonoids within the flower (Lovejoy, et. al., 2012).

Ginger

Zingiber officinale is a flowering plant, in the family *Zingiberaceae* whose rhizome, ginger root or simply ginger, is widely used as a spice or a folk medicine. It is a herbaceous perennial which grows annual stems about a meter tall bearing narrow green leaves and yellow flowers. Ginger produces clusters of white and pink flower buds that bloom into yellow flowers. Because of its aesthetic appeal and the adaptation of the plant to warm climates, ginger is often used as landscaping around subtropical homes. It is a perennial reed-like plant with annual leafy stems, about a meter tall. Traditionally, the rhizome is gathered when the stalk withers; it is

immediately scalded, or washed and scraped, to kill it and prevent sprouting. Ginger produces a hot, fragrant kitchen spice.

Oral or topical uses of ginger to treat various disorders, such as nausea or arthritis pain, are under research, but there are no conclusions possible from these studies about its effectiveness or safety in long-term use. In limited studies, ginger was found to be more effective than placebo for treating nausea caused by seasickness, morning sickness, and chemotherapy, although it was not found superior to placebo for treating postoperative nausea. Some studies indicate taking ginger during pregnancy may cause harm to the fetus or increase the mother's risk of bleeding. Allergic reactions to ginger generally result in a rash. Although generally recognized as safe, ginger can cause heartburn and other side effects, particularly if taken in powdered form. Un-chewed fresh ginger may result in intestinal blockage, and individuals who have had ulcers, inflammatory bowel disease, or blocked intestines may react badly to large quantities of fresh ginger. It can also adversely affect individuals with gallstones. There are suggestions that ginger may affect blood pressure, clotting, and heart rhythms. Studies are inconclusive about the effects of using ginger for nausea or pain associated with various ailments. Side effects of consuming powdered ginger are gas, bloating, heartburn and nausea. One traditional medical form of ginger historically was called "Jamaica ginger"; it was classified as a stimulant and carminative and used frequently for dyspepsia, gastroparesis, slow motility symptoms, constipation or colic. It was also frequently employed to disguise the taste of medicines. Ginger is commonly used to treat various types of "stomach problems," including motion sickness, morning sickness, colic, upset stomach, gas, diarrhea, nausea caused by cancer treatment, nausea and vomiting after surgery, as well as loss of appetite. Other uses include pain relief from arthritis or muscle soreness, menstrual pain, upper respiratory tract infections, cough, and bronchitis. Ginger is also

sometimes used for chest pain, low back pain, and stomach pain. Some people pour the fresh juice on their skin to treat burns. The oil made from ginger is sometimes applied to the skin to relieve pain. (WebMD, 2015).

Garlic

Allium sativum is an herb. Garlic is a seasoning and a food, and has been used as a medicine to prevent or treat a wide range of diseases and conditions. The fresh clove or supplements made from the clove are used for medicine. Garlic is easy to grow and can be grown year-round in mild climates. In the typical serving size of 1-3 cloves (3-9 g), garlic provides no significant nutritional value with the content of all essential nutrients below 10% of the Daily Value (DV). When expressed per 100 grams, garlic contains several nutrients in rich amounts (> 20% DV), including vitamins B6 and C, and the dietary minerals, manganese and phosphorus. Per 100 gram serving, garlic is also a good source (10-19% DV) of certain B vitamins including thiamine (Vitamin B1), and pantothenic acid (Vitamin B5), as well as certain dietary minerals including calcium, iron, and zinc (WebMD, 2015).

Garlic is used for many conditions related to the heart and blood system. These conditions include high blood pressure, high cholesterol, coronary heart disease, heart attack, and “hardening of the arteries” (atherosclerosis). Garlic actually may be effective in slowing the development of atherosclerosis and seems to be able to modestly reduce blood pressure. It also has indications for use in the prevention and treatment of cancers. Garlic has been used to treat enlarged prostate (benign prostatic hyperplasia; BPH), diabetes, osteoarthritis, hayfever (allergic rhinitis), traveler's diarrhea, high blood pressure late in pregnancy (pre-eclampsia), cold and flu. It is also used for building the immune system, preventing tick bites, and preventing and treating

bacterial and fungal infections. Other uses include treatment of fever, coughs, headache, stomach ache, sinus congestion, gout, rheumatism, hemorrhoids, asthma, bronchitis, shortness of breath, low blood pressure, low blood sugar, high blood sugar, and snakebites. It is also used for fighting stress and fatigue, and maintaining healthy liver function. Garlic is an antibacterial, an antifungal, and has some indications to be antiviral (WebMD, 2015).

Copper

Chemical element with symbol Cu, atomic number 29. It is a soft, malleable and ductile metal with very high thermal and electrical conductivity. A freshly exposed surface of pure copper has a reddish-orange color. It is most commonly used as a conductor of heat and electricity, a building material, and as a constituent of various metal alloys. Copper is found as a pure metal in nature, and was the source of the first metal to be used by humans, around 8,000 BC. Its compounds are commonly encountered as copper(II) salts, which often produce blue or green colors to minerals such as azurite, malachite and turquoise and have been widely used historically as pigments. The corrosion state is also used in architecture and art. The corrosion also protects the metal rather than completely deteriorating it. Copper not only corrodes in oxygen, but also will undergo galvanic corrosion when exposed to other metals. Copper tarnishes when exposed to sulfides, which react with it to form various copper sulfides. Copper is recyclable (Hammond, 2004).

Copper is essential to all living organisms as a trace dietary mineral because it is a key constituent of the mitochondrial respiratory enzyme complex cytochrome c oxidase. In mollusks and crustaceans copper is a constituent of the blood pigment hemocyanin, which is replaced by the iron-complexed hemoglobin in fish and other vertebrates. The main areas where copper is found in humans are the liver, muscle and bone. Copper compounds are used as

bacteriostatic/antimicrobial substances, fungicides, and wood preservatives. Certain isotopes of copper are used in radiation based emission topography. Compounds that contain a carbon-copper bond are known as organocopper compounds. They are very reactive towards oxygen to form copper(I) oxide. The human body contains copper at a level of about 1.4 to 2.1 mg per kg of body mass. Required Daily Amount for copper in normal healthy adults is estimated to be 0.97 mg/day and as high as 3.0 mg/day.[122] Copper is absorbed in the gut, then transported to the liver bound to albumin. After processing in the liver, copper is distributed to other tissues in a second phase. Copper transport in the tissues involves the protein ceruloplasmin, which carries the majority of copper in blood. Ceruloplasmin also carries copper that is excreted in milk, and is particularly well-absorbed as a copper source. A lack of copper can cause anemia-type symptoms and cellular dysfunction. An overdose of copper can cause liver failure, DNA damage, and the potential enhancement of neurological diseases such as Alzheimer's (Hammond, 2004).

Volts

The volt is the SI measurement of electric potential, electric potential difference, and electromotive force. The volt is named in honor of the Italian physicist Alessandro Volta. It is defined as the difference in electric potential across a wire or metal conductor when an electric current of one ampere dissipates one watt of power. It is also equal to the potential difference between two parallel, infinite planes spaced 1 meter apart that create an electric field of 1 newton per coulomb (C, SI unit of electric charge). It is also the potential difference between two points that will impart one joule of energy per coulomb of charge that passes through it.

Silver

Silver is the metallic element with the atomic number 47. Its symbol is Ag. A soft, white, lustrous transition metal, it possesses the highest electrical conductivity, thermal

conductivity, and reflectivity of any metal. The metal occurs naturally in its pure, free form (native silver), as an alloy with gold and other metals, and in minerals such as argentite and chlorargyrite. Most silver is produced as a byproduct of copper, gold, lead, and zinc refining. The electrical conductivity of silver is the highest of all metals, even higher than copper, but it is not widely used for electrical purposes due to its much higher cost. An exception to this is in radio-frequency engineering, particularly at VHF and higher frequencies, where silver plating is employed to improve electrical conductivity of parts and wires (at high frequencies current tends to flow on the surface of conductors, not their interior, hence silver plating greatly improves overall conductivity). Silver also has the lowest contact resistance of any metal (WebElements, 2012).

Conductivity/Resistivity

Electrical resistivity (also known as resistivity, specific electrical resistance, or volume resistivity) is an intrinsic property that quantifies how strongly a given material opposes the flow of electric current. A low resistivity indicates a material that readily allows the flow of electric current. Electrical conductivity or specific conductance is the reciprocal of electrical resistivity, and measures a material's ability to conduct an electric current. Many resistors and conductors have a uniform cross section with a uniform flow of electric current, and are made of one material.

Electrons in an atom do not take on arbitrary energy values. Rather, electrons only occupy certain discrete energy levels in an atom or crystal; energies between these levels are impossible. When a large number of such allowed energy levels are spaced close together (in energy-space)—i.e. have similar (minutely differing) energies—these energy levels together are

an "energy band". There can be many such energy bands in a material, depending on the atomic number {number of electrons (if the atom is neutral)} and their distribution (besides external factors like environmental modification of the energy bands). The material's electrons seek to minimize the total energy in the material by going to low energy states; however, the Pauli exclusion principle means that they cannot all go to the lowest state. The electrons instead "fill up" the band structure starting from the bottom. The characteristic energy level up to which the electrons have filled is called the Fermi level. The position of the Fermi level with respect to the band structure is very important for electrical conduction: only electrons in energy levels near the Fermi level are free to move around, since the electrons can easily jump among the partially occupied states in that region. In contrast, the low energy states are rigidly filled with a fixed number of electrons at all times, and the high energy states are empty of electrons at all times (Giancoli, 1995).

In metals there are many energy levels near the Fermi level, meaning that there are many electrons available to move. This is what causes the high electronic conductivity of metals. An important part of band theory is that there may be forbidden bands in energy: energy intervals that contain no energy levels. In insulators and semiconductors, the number of electrons happens to be just the right amount to fill a certain integer number of low energy bands, exactly to the boundary. In this case, the Fermi level falls within a band gap. Since there are no available states near the Fermi level, and the electrons are not freely movable, the electronic conductivity is very low. A metal consists of a lattice of atoms, each with an outer shell of electrons that freely dissociate from their parent atoms and travel through the lattice. This is also known as a positive ionic lattice. This 'sea' of dissociable electrons allows the metal to conduct electric current. When an electrical potential difference (a voltage) is applied across the metal, the resulting electric

field causes electrons to drift towards the positive terminal. The actual drift velocity of electrons is very small, in the order of magnitude of a meter per hour. However, as the electrons are densely packed in the material, the electromagnetic field is propagated through the metal at nearly the speed of light. Near room temperatures, metals have resistance. The primary cause of this resistance is the collision of electrons with the atoms that make up the crystal lattice. This acts to scatter electrons and lose their energy on collisions rather than on linear movement through the lattice. Also contributing to resistance in metals with impurities are the resulting imperfections in the lattice. The larger the cross-sectional area of the conductor, the more electrons per unit length are available to carry the current. As a result, the resistance is lower in larger cross-section conductors. The number of scattering events encountered by an electron passing through a material is proportional to the length of the conductor. The longer the conductor, therefore, the higher the resistance. Different materials also affect the resistance. In electrolytes, electrical conduction happens not by band electrons or holes, but by full atomic species (ions) traveling, each carrying an electrical charge. The resistivity of ionic solutions (electrolytes) varies tremendously with concentration – while distilled water is almost an insulator, salt water is a reasonable electrical conductor. Conduction in ionic liquids is also controlled by the movement of ions, but here we are talking about molten salts rather than solvated ions. In biological membranes, currents are carried by ionic salts. Small holes in cell membranes, called ion channels, are selective to specific ions and determine the membrane resistance (Giancoli, 1995).

Kanthal

Kanthal is the trademark for a family of iron-chromium- aluminium (FeCrAl) alloys used in a wide range of resistance and high-temperature applications. Kanthal FeCrAl alloys consist of

mainly iron, chromium (20–30%) and aluminium (4–7.5 %). The alloys are known for their ability to withstand high temperatures and having intermediate electric resistance. As such, it is frequently used in heating elements. Kanthal alloy has a melting point of 1,500 °C (2,730 °F). Special grades can be used as high as 1,425 °C (2,597 °F). Kanthal is moderately oxidative resistant and can handle temperature extremes (Kanthal.com, 2016).

Nichrome

Nichrome (NiCr, nickel-chrome, chrome-nickel,) generally refers to any alloy of nickel, chromium, and often iron and/or other elements or substances. Nichrome alloys are typically used in resistance wire. They are also used in some dental restorations (fillings) and in other applications. Nichrome wires are oxidative resistant and can handle most temperature extremes. It is silvery-grey in colour, is corrosion-resistant, and has a high melting point of about 1,400 °C (2,550 °F). Due to its resistance to oxidation and stability at high temperatures, it is widely used in electric heating elements, such as in appliances and tools. Typically, nichrome is wound in coils to a certain electrical resistance, and current is passed through it to produce heat. Nichrome is also used in fireworks and the pyrotechnics industry (Science Education Foundation, 2002).

Experimentation

The intent of phase one is to test cultured epidermal cells for their mitochondrial and cell reaction to stressors by monitoring the cell activity microscopically and testing conductivity before introduction, immediately after introduction and after five minutes of exposure to determine the level of mitochondrial function and reaction. This will lead to a database of cell reaction and conductivity that can be used in non-invasive indication of mitochondrial dysfunction that is an early sign of disease. The independent variables are the stressors and

stimuli being introduced to the cells. Dependent variables are the cell reaction and conductivity reaction which will be used as indicators of mitochondrial activity (function/dysfunction). The control group were the cells as they were plated without any additions, the constants were the sterility of the cell harvesting and plating, as well as the sustainability of the cells' viability through dextrose feeding and incubation. The amount of solution added and the assays used were also constant. The overall testing environment was also constant.

| Controls | Variables |
|--|-----------------------|
| Source of cells | Stressors and stimuli |
| Cell culturing procedures | |
| Cell feeding procedures | |
| Dye methodology/viability testing | |
| Amount of stressors and stimuli | |
| Method of observing mitochondrial changes | |
| Overall environment/temperature | |
| Method of observation | |
| Janus Green/Rhodamine B | |

The intent of phase two was to develop a sensor material that is sensitive enough to track cellular change at the mitochondrial/ATP production level. By testing existing materials such as silver as an electrode with a galvanic cell response sensor, then by testing copper, and wire hybrids Nichrome and Kanthal to determine their ability to detect cellular change in conductivity and resistivity.

| Controls | Variables |
|--|----------------------|
| Cell cultures and techniques | Type of wire |
| Incubation Time | Toxicity agents |
| Testing time and testing methods | Differing test areas |
| Sterility | |
| Amount of solutions added | |
| Feeding times and methods | |
| Epidermal cell source | |
| Dyes/Dye amount | |
| Software programs and measuring abilities | |

The following were the procedures that we used:

Cellular Monitoring through conductivity

- a. Make nutrient agar plates according to protocol. Pour slightly thicker than normal with wires included within the nutrient agar. Add 1mL Gentamycin to agar solution.
- b. Sterilize surface area of skin with 70% isopropyl alcohol (self and adult volunteer).
- c. Gently scrape surface area of skin to remove dead skin cells.
- d. Place nutrient agar disk onto area, flip agar back into plate.
- e. Place 5 drops of pre-prepared dextrose/nutrient broth solution on top of plate. Take conductivity/resistance of each plate before and after dextrose feed.
- f. Cover nutrient agar plate, place in incubator for 48 hours at 37 deg. C, adding more dextrose/nutrient broth solution as needed to prevent drying.
- g. Remove plates from incubator, set on microscope. View growth of cells
- h. Test conductivity/resistance of each cell plate.
- i. Add one microdrop per cell cluster of 50% dilution of Rhodamine B and 50% Janus Green B, allow to set for 9 minutes, while testing conductivity/resistance of each plate.
- j. Determine cell viability and mitochondrial viability through microscope observation, note colorimetric change.
- k. Add mitochondrial stressors to the top of the cells in 1 md doses and determine the conductivity/resistivity change in the cells. Place and compare all information into database for multiple trials.

Body Voltage Testing

- a. Using own self (diabetic) and an adult with no diabetes, test self-microvolt resistance and conductivity by zeroing out a voltmeter that has micro volt testing capabilities, then having the individual test their voltage by holding the ends of the probes every two hours for 12 hours. Repeat over 5 days (non-invasive).
- b. Test various areas around the body for best surface conduction/resistance production by placing the probes on surface areas easily accessed by personal accessory access (i.e. behind the ears, wrists, fingers, between fingers, etc.) (non-invasive)
- c.

Determining the most accurate sensor for cellular monitoring.

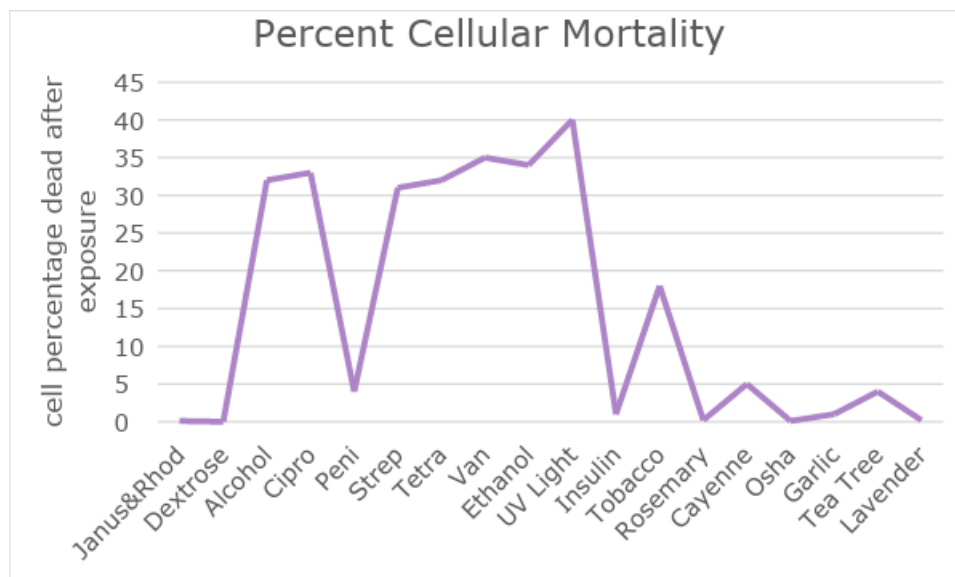
- a. Make nutrient agar plates according to protocol. Pour slightly thicker than normal. Add 1mL Gentamycin to agar solution.
- b. Sterilize surface area of skin with 70% isopropyl alcohol (self and adult volunteer).
- c. Gently scrape surface area of skin to remove dead skin cells.
- d. Place nutrient agar disk onto area of no skin cells, flip agar back into plate.
- e. Slip in sterilized (by alcohol) 10 cm electrode wires through agar, and gently bend outwards on either side.

- f. Place 5 drops of pre-prepared dextrose/nutrient broth solution on top of plate. Take conductivity/resistance of each plate.
- g. Cover nutrient agar plate, place in incubator for 48 hours at 37 deg. C, adding more dextrose/nutrient broth solution as needed to prevent drying.
- h. Remove plates from incubator, set on inverted microscope. View growth of cells away or towards the different wires.
- i. Test conductivity/resistance of each cell plate, through exposed wires.
- j. Add one microdrop per cell cluster of 50% dilution of Rhodamine B and 50% Janus Green B, allow to set for 9 minutes, while testing conductivity/resistance of each plate.
- k. Determine cell viability and mitochondrial viability through microscope observation, note colormetric change.
- l. Clamp wires to positive and negative terminals of volt meter, and test the conductivity/resistance/mV passing through each cell cluster by gently touching the end of a 3 cm wire to the voltage cluster.
- m. Add mitochondrial stressors to the top of the cells in 1 md doses and determine the conductivity/resistivity change in the cells.

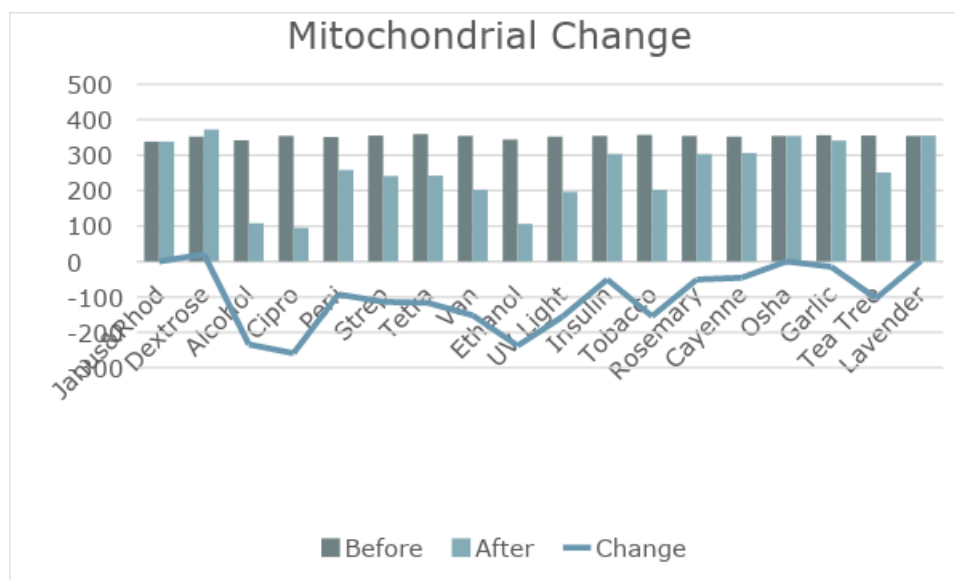
Building the sensor

- a. After determining the best wires for conductivity/resistance from step 2, a wire grid can be set up to produce a sensor network interlaced within a design that was determined to be the best for body conductivity/resistance from part 1.
- b. A controller chip and power production center will also be produced for the best design so that the sensor chip will simply lie against the skin and then transmit to a smart phone app a constant low level transmission of the amount of conductivity/resistivity currently being produced.

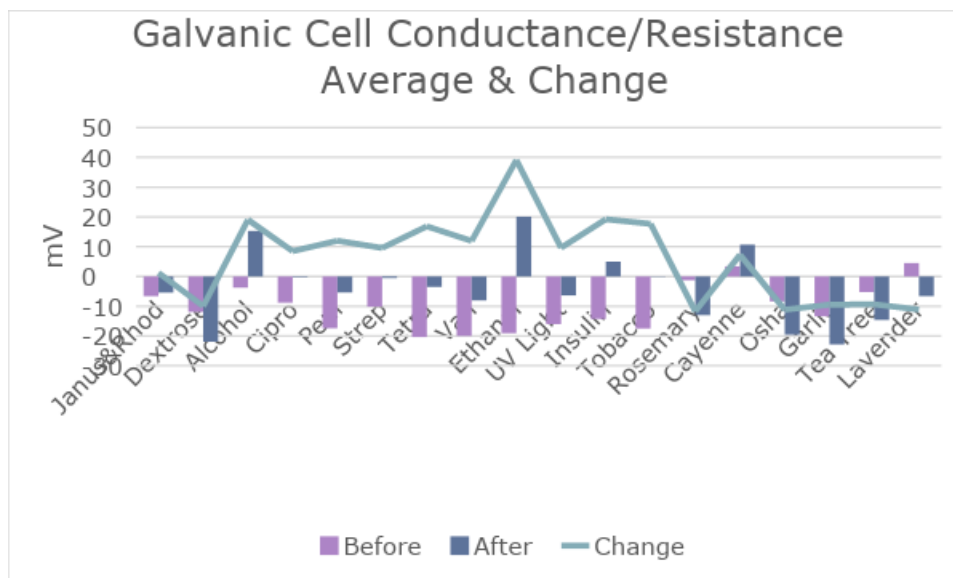
The following is data we collected over the years:



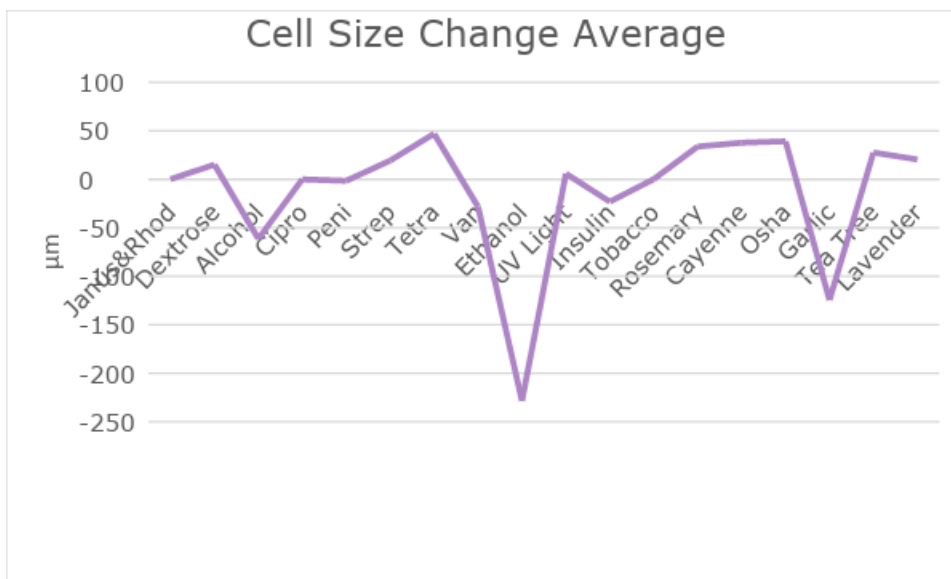
Cellular mortality indicates if the cell survived or not, and in the presence of some of the synthetically produced antibiotics and the tea tree oil solution, the cells had a high level of mortality.



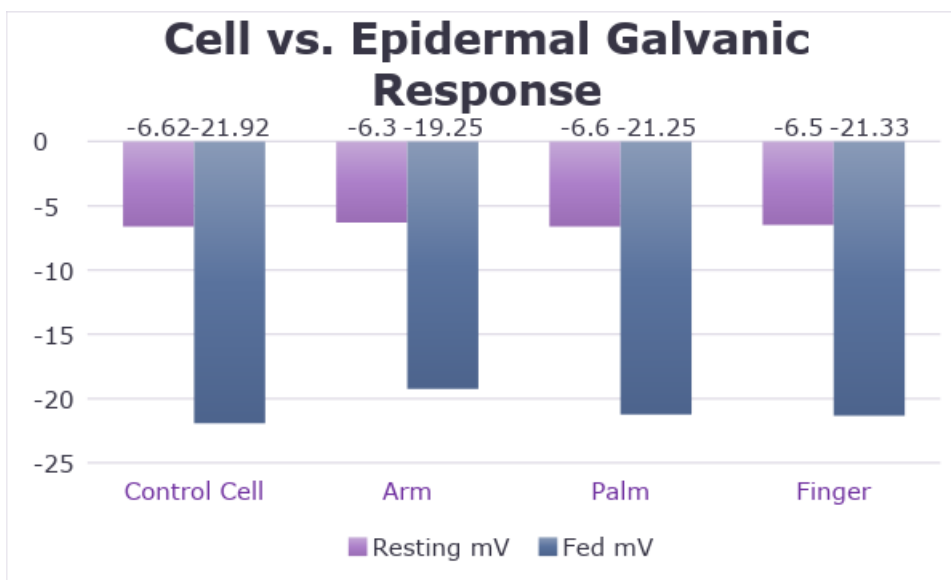
Mitochondrial change was observed through the Janus Green B illumination process, when the mitochondria winked out or turned pink, it was considered no longer viable and dysfunctional. There was a correlation to the dysfunction of the mitochondria and the conductivity level of the cells, once again most noticeable in the presence of the antibiotics and the tea tree oil.



The galvanic cell conductance and resistance was indicated by tapping each cell with the multimeter electrodes in the most sensitive mV reading. The voltage was taken before and after exposure to the solutions. Conductivity is a negative result, while resistivity is a positive result indicating a closer status to homeostasis.



Cell sizes were recorded before and after exposure to the different solutions and dyes, many of the solutions caused blebbing, which is the swelling or shrinking of a cell in tune with stimuli. If a cell shrinks or expands too far, it can lyse and destroy itself. Homeostasis is near zero.



The micro-voltmeter was tested on the cells, and then against the adult volunteer's arm, palm and finger tips by gently touching it to the arm. This was done before eating and 30 minutes after the subject consumed food, the results were comparable indicating a correlation between not the sweat produced, but the mitochondrial ATP function for galvanic response.

Phase 1 Conclusion:

Upon conducting the experiment and analyzing the data, it can be concluded that galvanic cell response correlates to galvanic skin response, as well as cellular response and mitochondrial response as a whole. Upon introduction of the varying solutions, the epidermal cell cultures had a notable reaction in mitochondrial function and mortality, and in some cases, cellular mortality. The cells reacted by blebbing within a matter of minutes of the exposure to the varying solutions, and the mitochondria began to change function and tone. As natural solutions were introduced to the cells, their resistivity increased, and as toxins and synthetically produced substances were introduced, their conductivity increased which correlates to galvanic skin response, with the only true exception being tea tree oil, which is a natural toxin. Tests were conducted with the epidermal cell subject by having them touch the multimeter on the arm, palm and fingertips, which indicated a very close correlation with cellular galvanic response and epidermal galvanic response. These experiments indicate a true correlation towards galvanic response at the cellular level in the face of potential medicines and toxins.

The hypothesis that stated that as the cells were stressed by toxins the conductivity will increase due to an immune response and the viable cells' attempt to work harder against the toxin, while the positive stimuli such as dextrose feeding will increase resistivity as the cell does not have to work as hard to maintain homeostasis is supported by the data. The second hypothesis which stated that copper and silver will be more sensitive to detecting these changes than other conductive materials is not supported as the copper and silver electrodes for the multimeter were not as sensitive as the gold and broke. The final hypothesis of as exposure to toxins increase, the more the conductivity will increase until the cell enters apoptosis or necrosis when the ATP function ceases. Cells that are not viable will not produce conductivity or resistance is fully supported by the data obtained. Future experimentation will include expanding

the test database with more medications and then to develop a full library for a testing strip to determine what toxins may be potentially causing cellular dysfunction at the mitochondrial level, which could lead to a paper testing strip or a simple galvanic test to determine the source of a disease.

Phase 2 Conclusion:

Upon conducting the experiment and analyzing the data, it can be concluded that the copper, then the Kanthal, and finally the Nichrome increased the conductivity/resistivity ability of the micromultimeter in detecting subtle changes in cell viability and function, and proved to be a more viable indicator of electrodermal and electrocellular activity than the silver electrodes used for galvanic skin response. A combination of the copper and Kanthal produces a galvanic cell sensor that is accurate to the cell reaction to stimuli within a matter of milliseconds. The sensor wires were capable of detecting resistance in healthy and functional cells, particularly those that were being fed with dextrose, and conductance in cells that were under stress from substances that are deemed toxins such as tobacco and alcohol. Measurements at the fingertips, between the fingers, at the wrist, and behind the ears were the most accurate in mV resistance and conductance in comparison to the cellular response resistance and conductance.

It was initially hypothesized that copper and silver would be sensitive enough to detect the changes in cellular conductivity and resistivity in accordance to stressors and normal stimuli, this was partially supported, as Kanthal does include silver oxide, and the copper was very accurate, moreso than the existing silver sensor included with a galvanic skin meter. Areas between the fingers and behind the ear should be most sensitive to conductivity/resistivity measurements. This hypothesis was partially supported, as the measurements behind the ear were

less accurate than the measuring points on the hands. The cells did show resistance the more healthy and homeostatic they were, because the cells were not having to work as hard to maintain homeostasis, while in the presence of stressors the conductance did increase because the cells were working harder to maintain homeostasis. Finally, the most conductive wires, copper and Kanthal can be integrated with small relays to translate the conductivity/resistivity readings to an application that will indicate cellular stress or dysfunction, this was also supported because the two wire types increased the sensitivity of a galvanic skin response sensor, therefore producing a galvanic cell response sensor. This data allows for the acceptance of the hypotheses as stated. Future experimentation will continue to test individuals who have known diagnosed ailments with the sensor apparatus, to program a separate app to measure in mV, and to downsize the sensor apparatus so that it is more portable and convenient, and can use a cell phone as a remote power source and processor.

The cost of experimentation is as follows:

| Item/Consumables | Purpose | Cost |
|-------------------------------------|--|-------------|
| Distilled Water | Solutions | \$97 |
| Petri Dishes w/Nutrient Agar | Culturing of cells | \$50.00 |
| Dextrose | Feeding of cell cultures | \$8.07 |
| Beral Pipettes | Feeding of cell cultures | \$7.25 |
| Nutrient Broth | Feeding of cell cultures | \$4.06 |
| 70% Isopropyl Alcohol | Chemical sterility of all equipment and area | \$4.56 |
| Hypodermic Needles | Application of dyes and solutions | \$5.56 |
| Janus Green | Mitochondrial dye | \$40.00 |
| Rhodamine B | Cell viability dye | \$10.45 |
| Toxin Solutions | Test items | \$20.00 |

| | | |
|---|------------|----------|
| Copper Electrodes | Test items | \$10.22 |
| Multimeter w/Gold Electrodes | Tester | \$21.97 |
| Kanthal Wire | Test items | \$20.00 |
| Nichrome Wire | Test items | \$6.95 |
| Galvanic Skin Sensor w/Silver Electrodes | Tester | \$97.00 |
| | Total | \$307.06 |

Problem Definition

Mitochondria are the powerhouses of the cell, but when something goes wrong, the mitochondria is responsible for cellular dysfunction. When the dysfunction is caused by outside sources, it is called mitochondrial toxicity, when it is caused by genetic sources, it is called mitochondrial dysfunction. The glycolytic and ATP progression in mitochondria produces electrical impulses, known as conductivity and resistivity, which in turn can be used as a signals for a wearable testing device to determine changes in body health. This has been experimented by our team for several years. The final product we are hoping to produce would be a lightweight, wearable sensor that is sensitive enough to detect cell signaling, galvanic cell response, and mitochondrial toxicity/dysfunction and relay this in real-time to a phone app to indicate immediate cellular impact, which would enable the user to take action to reverse the toxicity at the cellular level to avoid disease progression. The overall goal is to produce a fully functioning 3D model that shows how this sensor would work at the cellular level, as well as other important biological processes.

Problem Solution

From this experimentation and data, our team attempted to see if we could produce a database of cellular/mitochondrial reactions as indicators of disease/toxicity via mitochondria dysfunction from our scientific findings and enter it into a 3D model. Initially, we had hoped that

the model would provide a quicker way of gaining data and seeing if the sensor works. In doing this, our model needed to have a full library of ATP/electro activity that occurs within the mitochondria and cells during homeostasis and different diseases. This model would also help with reducing the confusion between electrodermal and electrocellular feedback.

However, after political battles and decisions made in our school district, our science fair program was shut down. This ended our team's experimentation and data gathering that we needed to complete the model. Because of this, we had to alter our problem solution. Instead, we just decided to model what happens during mitochondrial toxicity/dysfunction and how that can be measured through a sensor. We also wanted to highlight major biological processes that supported our findings from the updated model.

Data and End Results

Most of our data from these models are observational. However, in Version 8 we can get statistical data based off of the changes in permeability of the radius, amount of toxins introduced, the size of the radius of the membrane, and the number of good turtles/variables. From this, we can observe the amount of electrical signals that result from these changes.

Verification and Validation of Code

Our code is validated by the prior years of experimentation and data collection from our science fair project. We also validated the code through trial-and-error to ensure that it was working correctly and modeling what actually happens at the cellular level. This trial-and-error can be seen through the 8 different versions of code we have provided. For further validation please see the section of the report titled “Experimentation.” This section outlines what we did in the laboratory prior to modeling this and the conclusions we reached from it.

Conclusion

Through this model, we have established that a sensor used for the reasons we have put forth can theoretically work. Our model shows exactly what happens during mitochondrial toxicity and puts a visual aspect to the theory behind this project. We hope to continue work on this model to make it more realistic and add more variables. We would like to add variables such as other organelles, other cells, and organic molecules (CO₂, glucose, Oxygen, etc.). We would like to model the entire cell in relation to toxins, homeostasis (and disruption of homeostasis) within the cell and mitochondria, and disease of the cell. If we accomplish this, we feel that there will be even more supporting evidence for our scientific research/experimentation and we can

continue until a prototype of the sensor is made and we can distribute it. Ultimately, we want to finish the project and reach the goal we originally set at the beginning of the year.

Achievements

This year, the biggest achievement we have accomplished is being able to say that we have completed the challenge, as it was definitely challenging. As it was our junior year this year, we struggled managing our time between honors and AP classes, college classes, jobs, scholarships and activities like the Challenge. On top of this, our school shut down our beloved science fair program. This made our supercomputing project much harder, as we could no longer get data that we needed for our model.

Despite the struggles, we were able to learn and comprehend a lot about NetLogo. Since it was Sierra's first year in the Challenge, she learned about how the challenge works and what it's like to computer model real life experimentation. In doing so, we've learned more about my project and the components of it. Finally, we can say that we've achieved pride in what we have produced this year with the model, PowerPoint presentation, and presentation board. We hope to do very well at the Supercomputing Expo in April and bring some satisfaction and congratulations to a rather challenging year.

Acknowledgements

- Tracy Galligan
- Taos Municipal Schools
- Los Alamos National Labs
- Sandia National Labs
- Rowan Kinney
- Cyrus O'Hern
- Laura Tenorio
- NetLogo model library and community
-

Bibliography

- Adams, André A., et al. "Highly efficient circulating tumor cell isolation from whole blood and label-free enumeration using polymer-based microfluidics with an integrated conductivity sensor." *Journal of the American Chemical Society* 130.27 (2008): 8633-8641.
- Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 4th edition. New York: Garland Science; 2002. <https://www.ncbi.nlm.nih.gov/books/NBK21054/>
- Alters, Sandra & Alters, Brian (2008) *Introduction to Biology* "Chapter 12: Cell Reproduction" Wiley and Sons, NJ pgs. 179-193
- Boucsein, Wolfram (2012). *Electrodermal Activity*. Springer Science & Business Media. p. 2. ISBN 978-1-461-41126-0.
- Chen, Kang, et al. "Determination of urea in urine using a conductivity cell with surface acoustic wave resonator-based measurement circuit." *Talanta* 41.12 (1994): 2195-2200.
- Douglas C. Giancoli (1995). *Physics: Principles with Applications* (4th ed.). London: Prentice Hall.
- Gibco (2011) "Cell Culture Basics"
<http://www.invitrogen.com/site/us/en/home/References/gibco-cell-culture-basics.html>
- Haynes, William M., ed. (2011). *CRC Handbook of Chemistry and Physics* (92nd ed.). Boca Raton, FL: CRC Press. p. 3.246.
- Jackson, SM, Williams, M.L., Feingold, K.R., Elias, P.M. (1993) "Pathobiology of the stratum corneum" *West J. Med* 1993 March 158(3): 279-285. Retrieved from:
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1311754/#reference-sec>
- Kanthal Products. 2016. "Kanthal Applications and Products"
<http://www.kanthal.com/en/products/>
- Logsdon, John E.; Loke, Richard A (1999). "Propyl Alcohols". In Jacqueline I., Kroschwitz. *Kirk-Othmer Concise Encyclopedia of Chemical Technology*, (4th ed.). New York: John Wiley & Sons, Inc. pp. 1654–1656.
- Marshner's Mineral Nutrition of Higher Plants* (Third Edition) (2012). Petra Marshner, ed. "Plant Nutritions of Osha, Rosemary etc."

- Mulhall, H. J., et al. "Cancer, pre-cancer and normal oral cells distinguished by dielectrophoresis." *Analytical and bioanalytical chemistry* 401.8 (2011): 2455-2463.
- National Cancer Institute [NCI]. "Antioxidants and Cancer." 2014.
<http://www.cancer.gov/cancertopics/factsheet/prevention/antioxidants>
- National Center for Biotechnology Information (2015). PubChem Compound Database; CID=76123, <https://pubchem.ncbi.nlm.nih.gov/compound/76123>
- Oyagbemi AA, Saba AB, Azeez OI. (2010) Capsaicin: a novel chemopreventive molecule and its underlying molecular mechanisms of action.
<http://www.indianjancer.com/article.asp?issn=0019509X;year=2010;volume=47;issue=1;spage=53;epage=58;aulast=Oyagbemi> Indian J Cancer. 2010 Jan-Mar;47(1):53-8.
- Rheinwald, James G., and Howard Green. "Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cell is." *Cell* 6.3 (1975): 331-343.
- Rheinwald, James G., and Howard Green. "Epidermal growth factor and the multiplication of cultured human epidermal keratinocytes." *Nature* 265.5593 (1977): 421-424.
- Science Education Foundation. 2002 "Oxidative and Non-Oxidative Materials" http://www.sci-ed-ga.org/websites/gasef/images/pdf/light/section_13.pdf
- Schwiebert, Erik M. "ABC transporter-facilitated ATP conductive transport." *American Journal of Physiology-Cell Physiology* 276.1 (1999): C1-C8.
- Skulachev, V. P. "Membrane electricity as a convertible energy currency for the cell." *Canadian journal of biochemistry* 58.3 (1980): 161-175
- Sonksen P, Sonksen J (July 2000). "Insulin: understanding its action in health and disease". *British Journal of Anaesthesia*. 85 (1): 69–79.
- Webelements.com. (2016) "Silver, Copper, Gold" www.webelements.com
- Westerterp-Plantenga, MS; Reinbach, HC; Smeets, A; Martinussen, T; Møller, P (June 2009). "Effects of capsaicin, green tea and CH-19 sweet pepper on appetite and energy intake in humans in negative and positive energy balance.". *Clinical Nutrition*. 28 (3).

World Health Organization [WHO] “WHO REPORT on the global TOBACCO epidemic”
2008. http://www.who.int/tobacco/mpower/mpower_report_full_2008.pdf

