The UMAP Journal

Vol. 36, No. 3

Vol. 36, No. 3 2015

Table of Contents

Publisher's Editorial

MCM Modeling Forum

On Jargon

Publisher's Editorial
The Bedford Falls Effect

Solomon A. Garfunkel
Executive Director
COMAP, Inc.
175 Middlesex Turnpike, Suite 3B
Bedford , MA 01730–1459
s.garfunkel@comap.com

I have recently been reviewing COMAP activity over the past 35 years in order to update and clean up our Website. Aside from the pleasures and terrors of a walk down memory lane, it made me think about the place of COMAP in the mathematics education scene.

These days, I find myself more and more asking about whether or not there is a continued need for COMAP's existence, and if so, how that need may have changed since our inception. From the beginning, COMAP has always had one raison d'etre, namely to promote the teaching and learning of mathematics through mathematical modeling and contemporary applications. We have focused on the production of curriculum materials—Modules, textbooks, video series—in a variety of media and formats. And of course we have greatly expanded our modeling contests.

But modeling and applications are accepted now. In fact, we are somewhat "hot." No one seriously argues against the inclusion of applications in college math courses. Believe me, when we started, they did!—quite vociferously. Modeling is fashionable, and this *Journal* competes with many other mathematics journals for good applications-based articles. Once upon a time, we were it. This was the only journal where one could find serious applications of undergraduate mathematics. I would like to think that we played a role in shaping the current state of affairs.

The title of this editorial refers to the Frank Capra Christmas movie favorite, "It's a Wonderful Life," in which character George Bailey (played by Jimmy Stewart) gets to see what life in his home town of Bedford Falls would have been like if he had never been born. Well, we have no angel to show us an alternative history, but I believe that without this *Journal* and the outlet that UMAP and subsequent COMAP projects afforded like-minded instructors, we would not have come this far.

The UMAP Journal 36 (3) (2015) 195–196.

We have been an outlet and a clearinghouse—for your ideas, well before those ideas were part of the mainstream. And perhaps that is our greatest accomplishment and the strongest reason for our future work. There will be new directions to explore, in content, pedagogy, technology. There will always be a reluctance on the part of professional societies and other established organizations to embrace those new directions. And so there will always be a need for an organization willing to go "where angels fear to tread." And there will continue to be a need for an outlet for the talent and energy of our teachers and instructors who want and need to be heard.

The film closes with the tinkling of a Christmas tree bell and Bailey's daughter saying, "Look, Daddy; teacher says, 'Every time a bell rings, an angel gets his wings.'"

Think of us the next time you hear a school bell ring.

About the Author

Solomon Garfunkel is the founder and Executive Director of COMAP and Executive Publisher of this *Journal*.

He served on the mathematics faculties of Cornell University and the University of Connecticut at Storrs, but he has dedicated the last 35 years to research and development efforts in mathematics education. He was project director for the Undergraduate Mathematics and Its Applications (UMAP) and the High School Mathematics and Its Applications (HiMAP) Projects funded by NSF, and directed three telecourse projects, including *Against All Odds: Inside Statistics* and *In Simplest Terms: College Algebra*, for the Annenberg/CPB Project. He has been the Executive Director of COMAP, Inc. since its inception in 1980.

Dr. Garfunkel was the project director and host for the video series *For All Practical Purposes: Introduction to Contemporary Mathematics*. He was the Co-Principal Investigator on the ARISE Project, and Co-Principal Investigator of the CourseMap, ResourceMap, and WorkMap projects. In 2003, Dr. Garfunkel was Chair of the National Academy of Sciences and Mathematical Sciences Education Board Committee on the Preparation of High School Teachers.

Editor's Note

The views and opinions expressed in this issue by authors employed by the U.S. Department of Defense are theirs alone and not necessarily those of the Department of Defense or of any agency of the U.S. government.

MCM Modeling Forum

Results of the 2015 Mathematical Contest in Modeling

William P. Fox, MCM Director
Dept. of Defense Analysis
Naval Postgraduate School
1 University Circle
Monterey, CA 93943–5000
wpfox@nps.edu

Introduction

A total of 7,636 teams of undergraduates from hundreds of institutions and departments in 17 countries spent a weekend in February working on applied mathematics problems in the 31st Mathematical Contest in Modeling (MCM)®.

The 2015 MCM began at 8:00 P.M. EST on Thursday, February 5, and ended at 8:00 P.M. EST on Monday, February 9. During that time, teams of up to three undergraduates researched, modeled, and submitted a solution to one of two open-ended modeling problems. Students registered, obtained contest materials, downloaded the problems and data, and entered completion data through COMAP's MCM Website. After a weekend of hard work, solution papers were sent to COMAP on Monday. Two of the top papers appear in this issue of *The UMAP Journal*, together with commentaries.

Important News about the 2016 Contest

In additon to the customary two modeling problems calling upon continuous mathematics and discrete mathematics respectively, the 2016 MCM will feature a third problem, on data insights and statistical analysis. The On Jargon column in this issue offers some elaboration about this kind of problem.

The UMAP Journal 36 (3) (2015) 197–209.

Resources

In addition to this special issue of *The UMAP Journal*, COMAP offers a supplementary *2015 MCM-ICM CD-ROM* containing the press releases for the two contests, the results, the problems, unabridged versions of all the Outstanding papers, and judges' commentaries. Information about ordering is available at (800) 772–6627 or at

`http://www.comap.com/product/?idx=1466`

Results and winning papers from the first 30 contests were published in special issues of *Mathematical Modeling* (1985–1987) and *The UMAP Journal* (1985–2014). The 1994 volume of *Tools for Teaching*, commemorating the tenth anniversary of the contest, contains the 20 problems used in the first 10 years of the contest and an Outstanding paper for each year. That volume and the special MCM issues of the *Journal* for the last few years are available from COMAP. The 1994 volume is also available on COMAP's special *Modeling Resource* CD-ROM. Also available is *The MCM at 21* CD-ROM, which contains the 20 problems from the second 10 years of the contest, an Outstanding paper from each year, and advice from advisors of Outstanding teams. These CD-ROMs can be ordered from COMAP at

`http://www.comap.com/product/cdrom/index.html.`

Finally, the new volume *Mathematical Modeling for the MCM/ICM Contests Volume 1* is an exposition of the ideas, background knowledge, and modeling methodologies for solving the problems in the 2014 MCM/ICM contests: how to design rules to increase traffic throughput, how to rank the top coaches of a popular sport, and how to use networks to measure influence and impact.

That volume also presents a brief history of the MCM/ICM contests, offers ideas to help students prepare for the MCM/ICM contests, presents general modeling framework and methodologies, describes the judging procedure of the MCM/ICM papers, explains how to write successful MCM/ICM papers, and presents a sample scheduling of tasks during the contest. A number of exercise problems are included to help students understand the materials presented in the book.

Details and ordering are at

`http://216.250.163.249//product/?idx=1465.`

COMAP also sponsors:

- The MCM/ICM Media Contest (see p. 210).
- The Interdisciplinary Contest in Modeling (ICM)®, which runs concurrently with the MCM. Next year the ICM will again offer a modeling problem involving network science and a second problem involving human-environment interactons, plus a third problem on policy modeling. Results of this year's ICM are on the COMAP Website at

  `http://www.comap.com/undergraduate/contests.`

The contest report, an Outstanding paper, and commentaries appear in Vol. 36, No. 2.

- The High School Mathematical Contest in Modeling (HiMCM)®, which offers high school students a modeling opportunity similar to the MCM. Further details are at

  `http://www.comap.com/highschool/contests.`

2015 MCM Statistics

- 7,636 teams participated (with 2,137 more in the ICM)
- 12 high school teams (0.2%)
- 418 U.S. teams (5%)
- 7,218 foreign teams (93%), from Australia, Canada, China, Finland, Hong Kong SAR, India, Indonesia, Mexico, Netherlands, Scotland, Singapore, South Korea, Spain, Sweden, Taiwan and the United Kingdom
- 10 Outstanding Winners (1%)
- 12 Finalist Winners (1%)
- 641 Meritorious Winners (9%)
- 2,266 Honorable Mentions (31%)
- 4,685 Successful Participants (57%)

Problem A: Eradicating Ebola

Suppose that the world medical association has announced that their new medication could stop Ebola and cure patients whose disease is not advanced. Build a realistic, sensible, and useful model—that considers not only the spread of the disease, the quantity of the medicine needed, possible feasible delivery systems (sending the medicine to where it is needed), (geographical) locations of delivery, speed of manufacturing of the vaccine or drug, but also any other critical factors your team considers necessary as part of the model—to optimize the eradication of Ebola, or at least its current strain. In addition to your modeling approach for the contest, prepare a 1–2-page nontechnical letter for the world medical association to use in their announcement.

Problem B: Searching for a Lost Plane

Recall the lost Malaysian flight MH370. Build a generic mathematical model that could assist "searchers" in planning a useful search for a lost plane feared to have crashed in open water such as the Atlantic, Pacific, Indian, Southern, or Arctic Ocean while flying from Point A to Point B. Assume that there are no signals from the downed plane. Your model should recognize that there are many different types of planes for which we might be searching and that there are many different types of search planes, often using different electronics or sensors. Additionally, prepare a 1–2-page nontechnical paper for the airlines to use in their press conferences concerning their plan for future searches.

The Results

The solution papers were coded at COMAP headquarters so that names and affiliations of the authors would be unknown to the judges. Each paper was then read preliminarily by two "triage" judges at either Appalachian State University (Eradicating Ebola), Carroll College (Lost Plane Problem), or by a panel in China. Due to the very large number of entries for the Eradicating Ebola Problem, an additional triage site was created for that problem at the Naval Postgraduate School in Monterey, CA.

At the triage site, the summary, the organizations, and the paper's modeling content are the basis for judging the paper. The author and contest director put together a sheet for the triage judges that provides information for scoring the papers as well as maximum and minimum total points based upon content. When the two judges scores are divergent (differ by more than 3 points), a third judge scores the paper.

Final judging took place in Carmel, CA. The judges classified the papers as follows:

	Outstanding	Finalist	Meritorious	Honorable Mention	Successful Participation	Total
Eradicating Ebola Problem	5	5	430	1,618	3,281	5,356
Lost Plane Problem	5	7	211	648	1,404	2,280
	10	12	641	2,266	4,685	7,636

We list here the 10 teams that the judges designated as Outstanding; the list of all participating schools, advisors, and results is at the COMAP Website.

Outstanding Teams

Institution and Advisor	Team Members
Eradicating Ebola Problem	
Northwestern Polytechnical University Xi'an, Shaanxi, China Ligong Wang	Yizhou Wang Xiaojian Yang Yunqi Zhu
State University of New York at Buffalo Buffalo, NY John Ringland	Andrew Harris Dante Iozzo Nigel Michki
Chongqing University Chongqing, China Qu Gong	Fangliang Dong Daliang Xu Zhe Wang
Central South University Changsha, Hunan, China Xuanyun Qin	Yingzhe Zhou Sijia Xiong Peimin Li
University of Adelaide Adelaide, South Australia, Australia Sanjeeva Balasuriya	Parsa Kavkani Alexander Tam
Lost Plane Problem	
University of Colorado Boulder Boulder, CO Bengt Fornberg	Matthew Hurst Nathan Yeo Jordan Deitsch
University of Colorado Boulder Boulder, CO Bengt Fornberg	Derek Gorthy Christine Reilly Marc Thomson
Bethel University Arden Hills MN Nathan Gossett	Ben Visness Tyler Miller Philip Gibbens
Tsinghua University Beijing, China Zhenbo Wang	Xiang Chen Yuhe Tian Jiashen Tian
Colorado College Colorado Springs CO Andrea Bruder	Melissa Jay Nathan Mankovich Eleanore Campbell

Awards and Contributions

Each participating MCM advisor and team member received a certificate signed by the Contest Director and the appropriate Head Judge.

INFORMS, the Institute for Operations Research and the Management Sciences, recognized as INFORMS Outstanding teams two teams: the teams from the University of Adelaide (Eradicating Ebola Problem) and from Colorado College (Lost Plane Problem) and provided the following recognition:

- a letter of congratulations from the current president of INFORMS to each team member and to the faculty advisor;
- a check in the amount of $300 to each team member;
- a bronze plaque for display at the team's institution, commemorating team members' achievement;
- individual certificates for team members and faculty advisor as a personal commemoration of this achievement; and
- a one-year student membership in INFORMS for each team member, which includes their choice of a professional journal plus the *OR/MS Today* periodical and the INFORMS newsletter.

The Society for Industrial and Applied Mathematics (SIAM) designated one Outstanding team from each problem as a SIAM Winner. The SIAM Award teams were from Chongqing University (Eradicating Ebola Problem) and University of Colorado Boulder (team of Hurst, Yeo, and Deitsch) (Lost Plane Problem). Each team member was awarded a $300 cash prize. Their schools were given framed hand-lettered certificates in gold leaf.

The Mathematical Association of America (MAA) designated one North American team from each problem as an MAA Winner. The MAA Winners were from State University of New York at Buffalo (Eradicating Ebola Problem) and the Bethel University (Lost Plane Problem). Each team member was presented a certificate by an official of the MAA Committee on Undergraduate Student Activities and Chapters.

Ben Fusaro Award

One Meritorious, Finalist, or Outstanding paper is selected for the Ben Fusaro Award, named for the Founding Director of the MCM and awarded for the 11th time this year. It recognizes an especially creative approach; details concerning the award, its judging, and Ben Fusaro are in Vol. 25 (3) (2004): 195–196. The Ben Fusaro Award Winner was the Outstanding team from Central South University (Eradicating Ebola Problem) with team members Yingzhe Zhou, Sijia Xiong, and Peimin Li and team advisor Xuanyun Qin. A commentary about it appears in this issue.

Frank Giordano Award

For the third time, the MCM is designating a paper with the Frank Giordano Award. This award goes to a paper that demonstrates a very good example of the modeling process in a problem featuring discrete mathematics—this year, the Lost Plane Problem. Having worked on the contest since its inception, Frank Giordano served as Contest Director for 20 years. The Frank Giordano Award for 2015 went to the Outstanding team from Bethel University (Lost Plane Problem) with members Ben Visness, Tyler Miller, and Philip Gibbens and team advisor Nathan Gossett. A commentary about this paper appears in this issue.

Two Sigma Scholarship Award

The Two Sigma Scholarship Award went to the team from University of Colorado Boulder (Lost Plane Problem), with members Matthew Hurst, Nathan Yeo, and Jordan Deitsch and team advisor Bengt Forberg. They received a total of $10,000. This was the first year of this award, for which 463 U.S. teams competing in the MCM and ICM were eligible for one of two such scholarship prizes. We thank Two Sigma Investments for making this award possible.

Judging

Director
William P. Fox, Dept. of Defense Analysis, Naval Postgraduate School, Monterey, CA
Associate Director
Patrick J. Driscoll, Dept. of Systems Engineering, U.S. Military Academy, West Point, NY

Eradicating Ebola Problem

Head Judge
Marvin Keener, Dept. of Mathematics, Oklahoma State University, Stillwater, OK
Associate Judges
William C. Bauldry, Chair-Emeritus, Dept. of Mathematical Sciences, Appalachian State University, Boone, NC (Head Triage Judge)
Kelly Black, Dept. of Mathematics, Clarkson University, Potsdam, NY (SIAM Award Judge)
Karen Bolinger, Dept. of Mathematics, Clarion University, Clarion, PA (MAA Award Judge)

Zhijie Cai, School of Mathematical Sciences, Fudan University, Shanghai, China
Patrick J. Driscoll, Dept. of Systems Engineering, U.S. Military Academy, West Point, NY (Associate Contest Director)
Tim Elkins, Dept. of Systems Engineering, U.S. Military Academy, West Point, NY
Thomas Fitzkee, Dept. of Mathematics, Francis Marion University, Florence, SC (Fusaro Award Judge)
Ben Fusaro, Dept. of Mathematics, Florida State University, Tallahassee, FL (SIAM Judge)
Jerry Griggs, Dept. of Mathematics, University of South Carolina, Columbia, SC
Michael Jaye, Dept. of Defense Analysis, Naval Postgraduate School, Monterey, CA
Michael Tortorella, Dept. of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ

Triage Session at the U.S. Military Academy
Head Judge
Patrick J. Driscoll, Dept. of Systems Engineering
Associate Judges
James Enos, Christopher Green, Jacqueline Harris, Daniel McCarthy, Elizabeth Schott, and Russell Schott
—all from the Dept. of Systems Engineering, U.S. Military Academy, West Point, NY
Paul Heiney, Dept of Mathematics, U.S. Military Academy Preparatory School, West Point, NY
Ed Pohl, Dept. of Industrial Engineering
Letitia Pohl, Dept. of Civil Engineering
—both from University of Arkansas, Fayetteville, AR
Steve Henderson,U.S. Army Cyber Research Center, U.S. Military Academy, West Point, NY
Marie Samples, Dept. of Forensic Biology, Office of Chief Medical Examiner, New York, NY

Triage Session at Appalachian State University
Head Triage Judge
William C. Bauldry, Chair, Dept. of Mathematical Sciences
Associate Judges
William J. Cook, Ross Gosky, Jeffry L. Hirst, Lisa Maggiore, René Salinas, and Joel Sanqui
—all from the Dept. of Mathematical Sciences, Appalachian State University, Boone, NC
Kodwo Annan, Priya Shilpa Boindala, Amy H. Erickson, Keith Erickson, Junkoo Park, Katherine Pinzon, and James Price

—all from the Dept. of Mathematics, Georgia Gwinnett College, Lawrenceville, GA
Steven Kaczkowski and Douglas Meade,
—both from the Dept. of Mathematics, University of South Carolina, Columbia, SC
Samuel Kaplan, University of North Carolina at Asheville, Asheville, NC
Harrison Schramm, Office of the Chief of Naval Operations, Washington, DC
Rich West, Francis Marion University, Florence, SC

Triage Session at the Naval Postgraduate School
Head Judge
William P. Fox, Dept. of Defense Analysis, Naval Postgraduate School, Monterey, CA
Associate Judges
Robert Burks, Dept. of Defense Analysis
Michael Jaye, Dept. of Defense Analysis
Greg Mislick, Dept. of Operations Research
—all from the Naval Postgraduate School, Monterey, CA
Jeremiah Bartz and Thomas Fitzkee
—both from the Dept. of Mathematics, Francis Marion University, Florence, SC
Jay Belanger, Truman State University, Kirksville, MO
Karen Richey, Senior Cost Analyst, U.S. Government Accountability Office, Washington, DC
Thomas Smoltzer, Dept. of Mathematics and Statistics, Youngstown State University, Youngstown, OH

Triage Session in China
Head Judge
Zhijie Cai, School of Mathematical Sciences, Fudan University, Shanghai
Associate Judges
Yuan Cao, Fudan University, Shanghai
Zhongwen Chen, Soochow University, Suzhou
Hengjian Cui, Capital Normal University, Beijing
Qu Gong, Chongqing University, Chongqing
Mingfeng He, Dalian University of Technology, Dalian
Zhiqing He, East China University of Science and Technology, Shanghai
Haiyang Huang, Beijing Normal University, Beijing
Luming Jiang, East China Normal University, Shanghai
Qiyuan Jiang, Tsinghua University, Beijing
Laifu Liu, Beijing Normal University, Beijing
Liqiang Lu, Fudan University, Shanghai
Xiwen Lu, East China University of Science and Technology, Shanghai
Qiuhui Pan, Dalian University of Technology, Dalian

Chen Qiao, Xi'an Jiaotong University, Xi'an
Yongji Tan, Fudan University, Shanghai
Wenjuan Wang, Fudan University, Shanghai
Jianwen Xu, Chongqing University, Chongqing
Jinhai Yan, Fudan University, Shanghai
Wenyong Yan, Chengdu Technological University, Chengdu
Hu Yang, Chongqing University, Chongqing
Qifan Yang, Zhejiang University, Hangzhou
Jun Ye, Tsinghua University, Beijing
Jian Yuan, Southwest Jiaotong University, Chengdu
Hongyan Zhang, Central South University, Changsha
Yunjie Zhang, Dalian Maritime University, Dalian
Yicang Zhou, Xi'an Jiaotong University, Xi'an

Lost Plane Problem

Head Judge
Maynard Thompson, Mathematics Dept., University of Indiana, Bloomington, IN
Associate Judges
Fuping Bian, Tianjin University, Tianjin, China
Robert Burks, Operations Research Dept., Naval Postgraduate School, Monterey, CA (SIAM Award Judge)
William P. Fox, Dept. of Defense Analysis, Naval Postgraduate School, Monterey, CA (Contest Director)
Richard Marchand, Mathematics Dept., Slippery Rock University, Slippery Rock, PA
Veena Mendiratta, Lucent Technologies, Naperville, IL
David Olwell, Dept. of Systems Engineering, Naval Postgraduate School, Monterey, CA
Jack Picciuto, Director of Operations Analysis and Planning at IT Cadre, Ashburn, VA
Kathleen M. Shannon, Dept. of Mathematics and Computer Science, Salisbury University, Salisbury, MD (MAA Award Judge)
Dan Solow, Case Western Reserve University, Cleveland, OH (INFORMS Award Judge)
Rich West, Emeritus Professor of Mathematics, Francis Marion University, Florence, SC (Giordano Award Judge)

Triage Session at the Naval Postgraduate School
Head Judge
William P. Fox, Dept. of Defense Analysis, Naval Postgraduate School, Monterey, CA
Associate Judges
Michael Jaye, Dept. of Defense Analysis

David Olwell, Dept. of Systems Engineering
—both from the Naval Postgraduate School, Monterey, CA
Kelly Cline, Terry Mullen, Eric Sullivan, Marie Vanisko, and Theodore Wendt
—all from Dept. of Mathematics, Engineering, and Computer Science, Carroll College, Helena, MT
Richard West, Emeritus Professor, Dept. of Mathematics, Francis Marion University, Florence, SC
Katherine Fox, M.A., Licensed Counselor-I, Charleston, SC
Richard Marchand, Dept. of Mathematics, Slippery Rock University, Slippery Rock, PA
Veena Mendiratta, Lucent Technologies, Naperville, IL
Jack Picciuto, Director of Operations Analysis and Planning at IT Cadre, Ashburn, VA
Kathleen M. Shannon, Dept. of Mathematics and Computer Science, Salisbury University, Salisbury, MD
Bill Wilhelm, Lockheed Martin, Huntsville, AL

Sources of the Problems

Both the Eradicate Ebola Problem and the Lost Plane Problem were contributed by Contest Director William P. Fox (Dept. of Defense Analysis, Naval Postgraduate School, Monterey, CA).

Acknowledgments

Major funding for the MCM is provided by COMAP. Additional support is provided by the Institute for Operations Research and the Management Sciences (INFORMS), the Society for Industrial and Applied Mathematics (SIAM), and the Mathematical Association of America (MAA). We are indebted to these organizations for providing judges and prizes.

We also thank for their involvement and unflagging support the MCM judges and MCM Board members, as well as

- **Two Sigma Investments.** "This group of experienced, analytical, and technical financial professionals based in New York builds and operates sophisticated quantitative trading strategies for domestic and international markets. The firm is successfully managing several billion dollars using highly-automated trading technologies. For more information about Two Sigma, please visit `http://www.twosigma.com`."

Cautions

To the reader of research journals:

Usually a published paper has been presented to an audience, shown to colleagues, rewritten, checked by referees, revised, and edited by a journal editor. Each paper here is the result of undergraduates working on a problem over a weekend. Editing (and usually substantial cutting) has taken place; minor errors have been corrected, wording has been altered for clarity or economy, and style has been adjusted to that of *The UMAP Journal*. The student authors have proofed the results. Please peruse these students' efforts in that context.

To the potential MCM advisor:

It might be overpowering to encounter such output from a weekend of work by a small team of undergraduates, but these solution papers are highly atypical. A team that prepares and participates will have an enriching learning experience, independent of what any other team does.

COMAP's Mathematical Contest in Modeling and Interdisciplinary Contest in Modeling are the only international modeling contests in which students work in teams. Centering its educational philosophy on mathematical modeling, COMAP serves the educational community as well as the world of work by preparing students to become better-informed and better-prepared citizens.

About the Author

Dr. William P. Fox is a professor in the Department of Defense Analysis at the Naval Postgraduate School and teaches a three-course sequence in mathematical modeling for decision-making. He received his B.S. degree from the United States Military Academy at West Point, New York, his M.S. at the Naval Postgraduate School, and his Ph.D. at Clemson University. Previously he has taught at the United States Military Academy and at Francis Marion University, where he was the Chair of Mathematics for eight years. He has many publications and scholarly activities including books, chapters of books, journal articles, conference presentations, and workshops. He directs several mathematical modeling contests through COMAP: HiMCM and MCM. His interests include applied mathematics, optimization (linear and nonlinear), mathematical modeling, statistical models for medical research, and computer simulations. He is President-Emeritus of the NPS Faculty Council and President of the Military Application Society of INFORMS.

Editor's Note

The complete roster of participating teams and results is too long to reproduce in the *Journal*. It can be found at the COMAP Website, in separate files for each problem:

`http://www.comap.com/undergraduate/contests/mcm/contests/2015/results/2015_MCM_Problem_A_Results.pdf`
`http://www.comap.com/undergraduate/contests/mcm/contests/2015/results/2015_MCM_Problem_B_Results.pdf`

Media Contest

This year, COMAP again organized an MCM/ICM Media Contest.

Over the years, contest teams have increasingly taken to various forms of documentation of their activities over the grueling 96 hours—frequently in video, slide, or presentation form. This material has been produced to provide comic relief and let off steam, as well as to provide some memories days, weeks, and years after the contest. We *love* it, and we want to encourage teams (outside help is allowed) to create media pieces and share them with us and the MCM/ICM community.

The media contest is *completely separate* from MCM and ICM. No matter how creative and inventive the media presentation, it has *no* effect on the judging of the team's paper for MCM or ICM. We do not want work on the media project to detract or distract from work on the contest problems in any way. This is a separate competition, one that we hope is fun for all.

Further information about the contest is at

`http://www.comap.com/undergraduate/contests/mcm/media.html.`

There were 16 entries—12 of them from Dalian Maritime University! (Come on, you other schools!)

Outstanding Winner of the 5th MCM/ICM Media Contest:

- Dalian Maritime University, Dalian, China
  (Chenhui Xu, Mingzhu Tang, Fang Zhang)

Finalists:

- Dalian Maritime University
  (Mingqiushuo Wei, Minmin Xiong, Ren Hu)

- Dalian Maritime University
  (Lei Zhang, Chenfan Lian, Kangming Wu)

The remaining entries were judged Meritorious Winners. Complete results, including links to the Outstanding videos, are at

`http://www.comap.com/undergraduate/contests/mcm/contests/2015/solutions/index.html.`

How to Eradicate Ebola?

Yizhou Wang
Xiaojian Yang
Yunqi Zhu

Northwestern Polytechnical University
Xi'an, China

Advisor: Ligong Wang

Abstract

The breakout of Ebola in 2014 triggered global panic. How to control and eradicate Ebola has become a universal concern ever since.

First, we build an epidemic model SEIHCR (CT) that takes the special features of Ebola into consideration. These are treatment in *hospital*, *infectious corpses*, and intensified *contact tracing*. This model is developed from the traditional SEIR model. The model's results, whose parameters are determined using computer simulation, match perfectly with the data reported by WHO, suggesting the validity of our improved model.

Second, we study *pharmaceutical intervention* thoroughly. The amount of medicine needed is based on the cumulative number of individuals. Results calculated from the WHO statistics and from our SEIHCR (CT) model show only minor discrepancy, further indicating the feasibility of our model. In designing the delivery system, we apply the weighted Fuzzy Means Clustering Algorithm and select 6 locations that should serve as medicine delivery centers for other cities. We optimize the delivery locations and determine each location's share. The average speed of manufacturing should be no less than 106.2 unit doses per day, and an increase in the manufacturing speed and the efficacy of medicine will reinforce the intervention effect.

Third, other critical factors, *safer treatment of corpses*, and *earlier identification/isolation*, also prove to be relevant. Results show that these interventions will help reduce the number of infections more rapidly.

We then analyze the factors for control—and the time of eradication—of Ebola. For example, when the rate of the infectious being isolated is 33%–40%, the disease can be successfully controlled. When the introduction time for treatment decreases from 210 to 145 days, the eradication of Ebola arrives over 200 days earlier.

Finally, we select three parameters: the transmission rate, the incubation period and the fatality rate for *sensitivity analysis*.

The UMAP Journal 36 (3) (2015) 211–236.

Introduction

Problem Background

Ebola virus disease (EVD), formerly known as Ebola hemorrhagic fever, is a severe, often fatal illness in humans. The 2014 outbreak in West Africa is the largest and most complex outbreak since the Ebola virus was first discovered in 1976, with more cases and deaths than all other outbreaks combined. It started in Guinea and spread to Sierra Leone and Liberia. The situation as of 4 February 2015 in those countries can be seen in **Figure 1**.

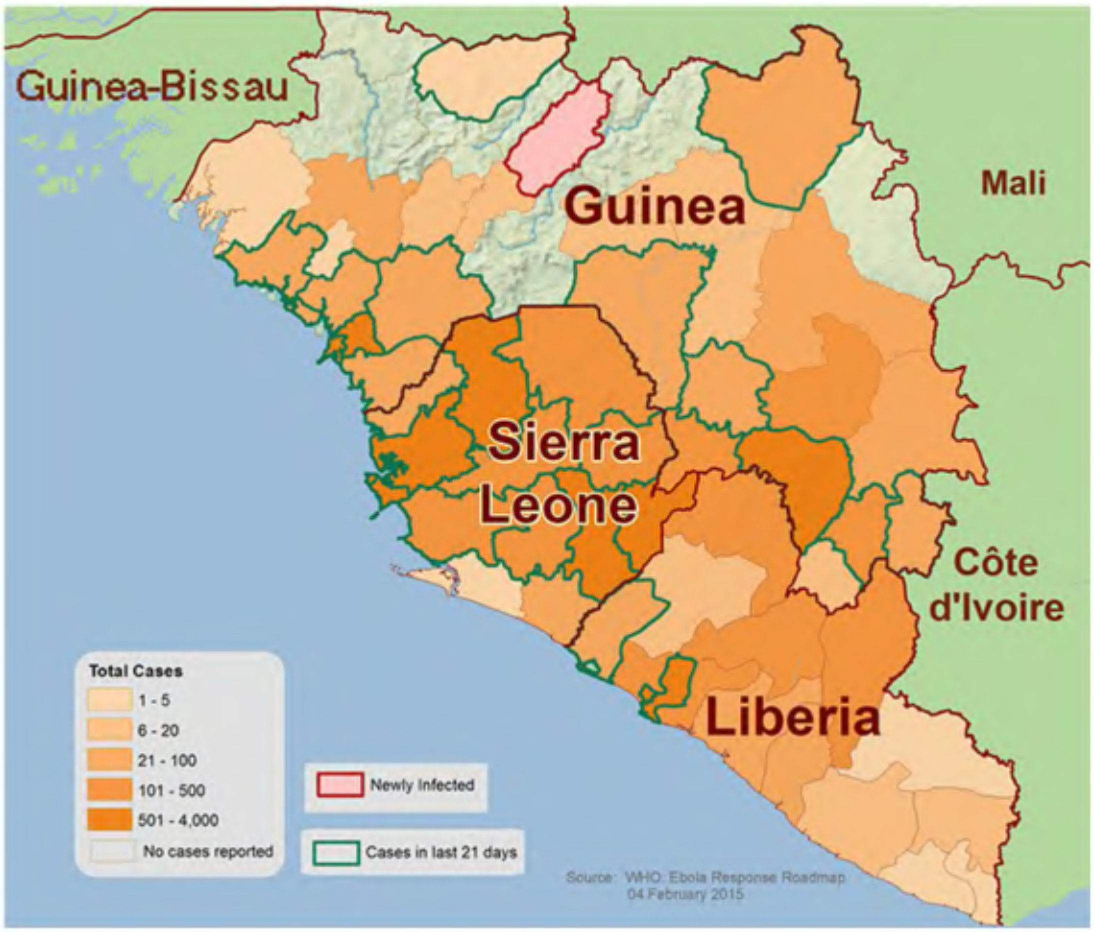

Figure 1. Ebola outbreak distribution in West Africa on 4 February 2015. Source: CDC [2015b].

Ebola was first transmitted from fruit bats to humans. It can now spread from human to human via direct contact with blood, secretions, organs or other bodily fluids of infected people, and from surfaces and materials contaminated with these fluids. Burial ceremonies can also play a role in the transmission of Ebola because the virus can also be transmitted through contact with the body of the deceased person.

Control of outbreaks requires coordinated medical services, alongside a certain level of community engagement. The medical services include rapid detection of cases, tracingof those who have come into contact with infected individuals, quick access to laboratory services, proper healthcare for the infected, and proper disposal of the dead through cremation or burial.

There are vaccines and drug treatments under development, but they have not yet been fully tested for safety or effectiveness. In the summer of 2014, the World Health Organization claimed that fast-tracking of testing was ethical in light of the epidemic [Tetro 2015], and the first batch of an experimental vaccine was sent to Liberia in January 2015 [Mundasad 2015].

Previous Research

The analysis of the spread of an epidemic has been a universal concern, and there has been much research into the 2014 Ebola epidemic in particular [Chowell and Nishiura 2014].

For example, Fisman et al. [2014] used a two-parameter mathematical model to describe the epidemic growth and control. Gomes et al. [2014] researched the risk of Ebola's spread; the risk of spread to outside Africa is relatively small compared to its expansion among West African countries.

The Centers for Disease Control and Prevention used the traditional SIR model to extrapolate the Ebola epidemic and project that Liberia and Sierra Leone could have 1.4 million Ebola cases by 20 January 2015 [Simon and Korman 2014]. Chowell et al. [2004] had used the SEIR model and studied the effect of Ebola outbreaks in 1995 in Congo and in 2000 in Uganda; the SEIR model takes the state of exposure into consideration. Based on that SEIR model, Althaus [2014] developed a model where the reproduction number is dependent on the time.

Valdez et al. [2015] found from modeling that reducing population mobility had little effect on geographical containment of the disease, but that rapid and early intervention that increases hospitalization and reducing transmission in hospitals and at funerals are the most important responses to any re-emerging Ebola epidemic.

Our Work

We are asked to build a realistic, sensible, and useful model to optimize the eradication of Ebola (or at least its current strain). Our model should consider not only the spread of the disease, the quantity of medicine needed, possible feasible delivery systems, locations of delivery, speed of manufacturing of the vaccine or drug, but also other critical factors that we consider necessary.

We apply the SEIR model to predict the spread of Ebola in the early stage.

Next, we analyze pharmaceutical intervention. We use the statistics provided by WHO [2015b] to determine the quantity of needed medicine.

We use our model of spread to predict the number of the patients and needed medicine, based on the current cumulative number of patients and the increasing rate of spread of the disease.

We sort the affected cities into several groups to determine delivery locations. We apply the Fuzzy Means Clustering Algorithm and select six delivery points to store the medicine and transport it to the nearby cities. Later, we take the speed of manufacturing and the efficacy of the medicine into consideration and test how those parameters could affect intervention.

We then consider other important factors that would help the eradication of Ebola: safer treatment of corpses, and intensified contact-tracing and earlier isolation.

Finally, we predict a time for Ebola's eradication.

General Assumptions

- Newborns are not counted in the total population.
- An individual who is exposed enters the incubation period and is not yet infectious.
- Recovered patients will not be infected again.
- Medicine is provided only in hospitals.
- Any location inside Guinea, Sierra Leone, and Liberia can be selected as the delivery center, and delivery routes between cities are straight lines.
- Medicine and vaccines can be delivered across borders.

Terminology and Symbol Description

Terminology

- **Susceptible individual** [Valdez et al. 2014]: "A person with a clinical illness compatible with EVD and within 21 days before onset of illness, either:
  - a history of travel to the affected areas OR
  - direct contact with a probable or confirmed case OR
  - exposure to EVD-infected blood or other body fluids or tissues OR
  - direct handling of bats, rodents, or primates from Ebola-affected countries OR
  - preparation or consumption of bush meat from Ebola-affected countries."

- **Exposed individual**: An infected person who is not yet infectious or symptomatic.
- **Basic reproduction number**: "The average number of secondary cases caused by a typical infected individual through its entire course of infection in a completely susceptible population and in the absence of control interventions" [Chowell and Nishiura 2014].
- **Contact tracing**: "Contact tracing is finding everyone who comes in direct contact with a sick Ebola patient. Contacts are watched for illness for 21 days from the last day they came in contact with the Ebola patient. If the contact develops a fever or other Ebola symptoms, they are immediately isolated, tested, provided care, and the cycle starts again—all of the new patient's contacts are found and watched for 21 days" [CDC 2014b].

Table 1.

Symbol table.

Symbol	Description
R_0	Basic reproduction number
N	Total population
$S(t)$	Number of suspected individuals at time t
$E(t)$	Number of exposed individuals at time t
$I(t)$	Number of infectious individuals outside hospital at time t
$H(t)$	Number of hospitalized individuals at time t
$C(t)$	Number of contaminated deceased at time t
$R(t)$	Number of removed individuals at time t
CUM(t)	Cumulative number of individuals at time t
β_0	Average transmission rate per person per day without control measures
$\beta(t)$	Average transmission rate per person per day after control measures
$(\beta_I, \beta_H, \beta_C)$	Transmission rate per person per day (outside hospital, inside hospital, by corpses)
λ	Total transmission rate = $\frac{1}{N}\left((\beta_I I(t) + \beta_H H(t) + \beta_C C(t)\right)$
$1/\sigma$	Average duration of incubation
α	Rate of identification/isolation of infectious individuals
$1/\gamma$	Average duration of infectiousness (basic SEIR model)
(γ_I, γ_H)	Average time from symptoms onset to recovery (outside hospital, inside)
(ν_I, ν_H)	Average time from symptoms onset to death (outside hospital, inside)
(δ_I, δ_H)	Fatality rate (outside hospital, inside)
$1/\psi$	Average time until the deceased is properly handled
κ	Average number of contacts traced per identified/isolated infectious individual
(π_E, π_I)	Probability a contact traced individual (exposed, infectious) is isolated without causing a new case
(ω_E, ω_I)	Ratio of probability that contact traced individual is (exposed, infectious) at time of originating case identification to the probability that a random individual in the population is (exposed, infectious)

Spread of Ebola

Traditional Epidemic Model

The SEIR Model

The transmission of Ebola follows the dynamics of the SEIR (susceptible-exposed-infectious-recovered) model, which is described by the following ordinary differential equations from [Chowell et al. 2004]:

$$\begin{aligned}
\dot{S}(t) &= -\beta S(t) I(t)/N, \\
\dot{E}(t) &= \beta S(t) I(t)/N - \sigma E(t), \\
\dot{I}(t) &= \sigma E(t) - \gamma I(t), \\
\dot{R}(t) &= \gamma I(t), \\
\dot{\text{CUM}}(t) &= \sigma E(t),
\end{aligned}$$

where the symbols are as defined in **Table 1** on p. 215. In the absence of control interventions, $\beta = \beta_0$ remains constant. However, after control measures are introduced at time τ, β is assumed to decay exponentially at rate k [Lekone and Finkenstädt 2006]. That is,

$$\beta(t) = \begin{cases} \beta_0, & t < \tau; \\ \beta_0 e^{-k(t-\tau)}, & t \geq \tau. \end{cases}$$

Outbreak Data

We use the SEIR model to estimate the cumulative cases in Guinea, Sierra Leone, and Liberia. We use the parameter estimates in **Table 2** and compare the model's predictions with the WHO data from March 1, 2014 to February 4, 2015 [WHO 2015a; CDC 2015a], for numbers of reported total cases (confirmed, probable, and suspected) and deaths.

Table 2.
Parameter estimates for the 2014 Ebola outbreak.

Parameter	Guinea	Sierra Leone	Liberia
Basic reproduction number, R_0	1.51	2.53	1.59
Transmission rate β_0 without control intervention	0.27	0.45	0.28
Transmission decay rate k	0.048	0.048	0.048
Population (millions) N	12.0	6.1	4.1

The *basic reproduction number is defined as*

$$R_0 = \beta \cdot \frac{1}{\gamma},$$

where $1/\gamma = 5.61$ days is the infectious duration from Chowell et al. [2004], who also give $1/\sigma = 5.3$ days as the average duration of the incubation period in a previous outbreak of Ebola in Congo in 1995. Values of the other parameter estimates in **Table 2** are from Althaus [2014].

Results of the SEIR Model

In **Figure 2** on p. 218, Day 1 is set to March 1, 2014. The graphs demonstrate that the model fits the data very well. Therefore, the SEIR model can serve as a good tool for analyzing the spread of the Ebola outbreak.

Improved Model

The SEIHCR (CT) Model

SEIR is a basic model for predicting the spread of a disease. However, for Ebola, we have to take other important factors into consideration:

- the potential threat posed by infectious corpses,
- the provision of hospital care, and
- the involvement of contact tracing (CT).

Figure 3 on p. 219 shows a schematic presentation of the improved SEIHCR (CT) model, indicating the compartmental states and the transition rates among the states.

The improved SEIHCR (CT) model [Rivers et al. 2014; Browne et al. 2014] can be written as

$$
\begin{aligned}
\dot{S}(t) &= \frac{1}{N}\big(-\beta_I S(t)I(t) - \beta_H S(t)H(t) - \beta_C S(t)C(t)\big),\\
\dot{E}(t) &= \frac{1}{N}\big(\beta_I S(t)I(t) + \beta_H S(t)H(t) + \beta_C S(t)C(t)\big) - \sigma E(t)\\
&\qquad - \kappa\big(\alpha I(t) + \psi C(t)\big)\pi_E\omega_E\big(E(t)/N\big),\\
\dot{I}(t) &= \sigma E(t) - I(t)\big(\alpha + (1-\delta_I)\gamma_I + \delta_I\nu_I\big) - \kappa\big(\alpha I(t)\\
&\qquad + \psi C(t)\big)\pi_I\omega_I\big(I(t)/N\big),\\
\dot{H}(t) &= \alpha I(t) - (1-\delta_H)\gamma_H H(t) - \psi C(t),\\
\dot{C}(t) &= \delta_I\nu_I I(t) + \delta_H\nu_H H(t) - \psi C(t),\\
\dot{R}(t) &= (1-\delta_I)\gamma_I I(t) + (1-\delta_H)\gamma_H H(t) + \psi C(t),\\
\dot{\mathrm{CUM}}(t) &= \sigma E(t).
\end{aligned}
$$

The cumulative number of infected individuals can be obtained from

$$\mathrm{CUM}(t+\Delta t) = \mathrm{CUM}(t) + \int_t^{t+\Delta t} \sigma E(s)\,ds.$$

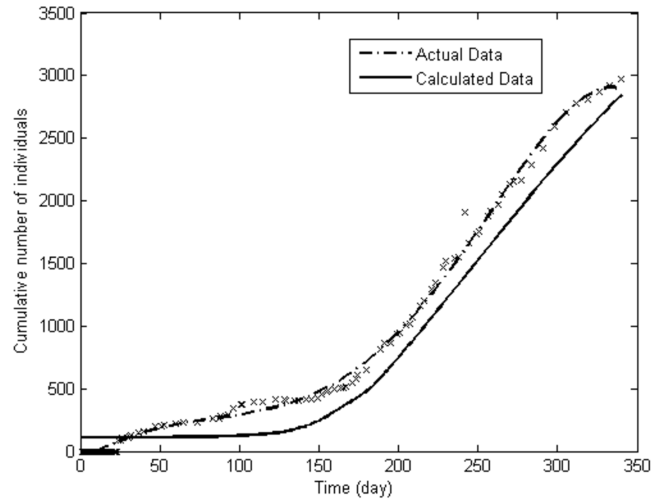

Figure 2a. Guinea.

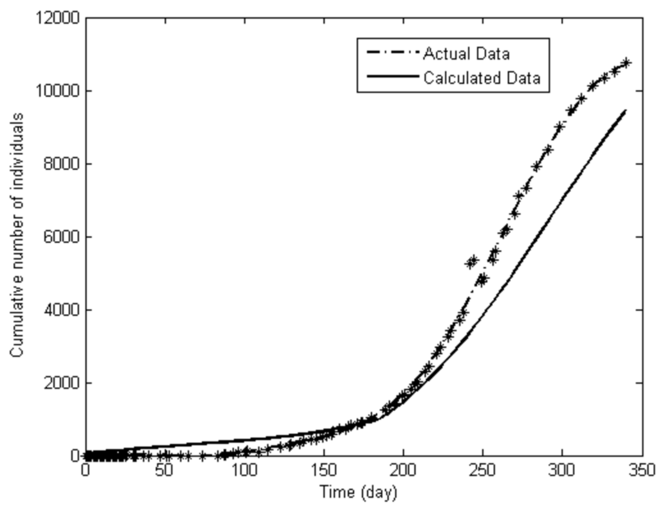

Figure 2b. Sierra Leone.

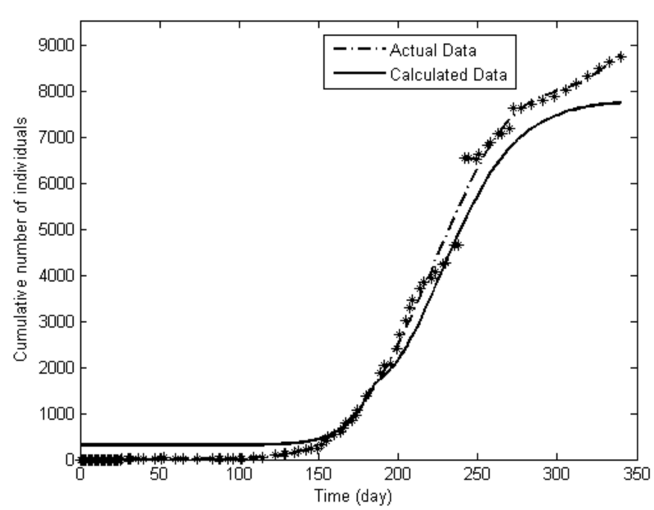

Figure 2c. Liberia.

Figure 2. Cumulative number of infections for Guinea, Sierra Leone and Liberia: Model predictions as dotted curve, actual data as dotted curve fitted to data points marked by Xs.

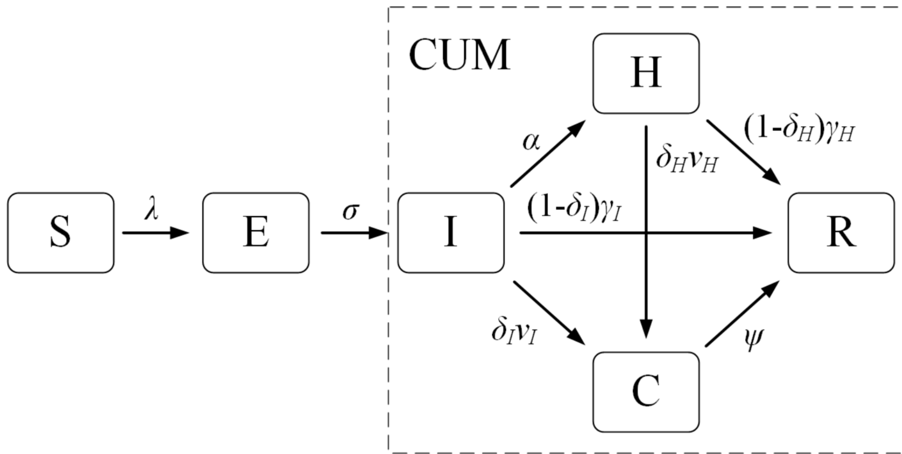

Figure 3. Compartmental flow of the SEIHCR (CT) model.

Reasons for considering corpses, hospitals, and contact tracing

- Providing hospital care to affected populations could be used as a basis for treating patients and prevent the disease from spreading to a larger scale. Regulating funerals and burials can reduce the risk of infection from corpses [Briand et al. 2014].
- Contact tracing is sometimes regarded as the key factor to stanch the Ebola outbreak in West Africa because if patients can be found earlier in their course of illness, the chance of exposure of family members and health workers will be reduced significantly.

We name the modified model the **SEIHCR (CT) model** to illustrate its improvement and differences from the traditional SEIR model.

Choosing Parameters

- **Estimation of R_0:** To obtain the reproduction number, we follow van den Driessche and Watmough [2002]:

  $$R_0 = \rho(GV^{-1}),$$

  where:

  - $\rho(A)$ is the spectral radius of matrix A,
  - F is the rate of appearance of new infections, and
  - V is the rate of transfer of individuals by all other means.

The F and V for the SEIHR (CT) model are

$$F = \frac{1}{N}\begin{pmatrix} -(\beta_I SI + \beta_H SH + \beta_C SC) \\ (\beta_I SI + \beta_H SH + \beta_C SC) \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$V = \begin{pmatrix} 0 \\ -\sigma E \\ \sigma E - \alpha I - \gamma_I I - \nu_I I \\ \alpha I - \gamma_H H - \nu_H H \\ \nu_I I + \nu_H H - \psi C \\ \gamma_I I + \gamma_H H + \psi C \end{pmatrix}.$$

- **Determination of other parameters**: We estimate the values of the parameters β_I, β_H, β_C, α, and ψ for the three countries using a least-squares curve-fitting algorithm. We use values of the parameters σ, γ_I, γ_H, ν_I, ν_H, δ_I, and δ_H from Althaus [2014].

  The basic reproduction number of the SEIHCR (CT) model is given by the following formula, computed by the next-generation method:

$$R_0 = \frac{\beta_I}{\alpha + (1-\delta_I)\gamma_I + \delta_I \nu_I} + \frac{\beta_H}{(1-\delta_H)\gamma_H + \delta_H \nu_H} + \frac{\beta_C}{\psi}.$$

  Among all these parameters, β_I is constant in the absence of control interventions, just as β_0 was in the SEIR model. After control measures are introduced at time τ, β_I is assumed to decay exponentially at rate k (just as β did in the SEIR model) [Lekone and Finkelstädt 2006].

  Since k and γ_H are both related to the efficacy of medicine, we assume their relationship to be

$$\gamma_H = \theta k,$$

  where θ is assumed to be 3.2 based on experience [Lekone and Finkenstädt 2006].

Results of the SEIHCR (CT) Model

Figure 4 compares for the three countries the results of the SEIHR (CT) model with the actual data. The values of the parameters are shown in **Table 3**.

Comparing **Figure 4** with **Figure 3** (p. 219), we see better fits with the SEIHCR (CT) model.

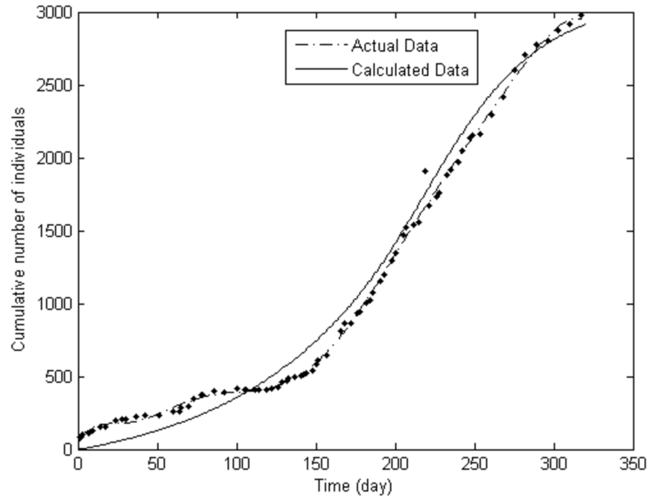

Figure 4a. Guinea.

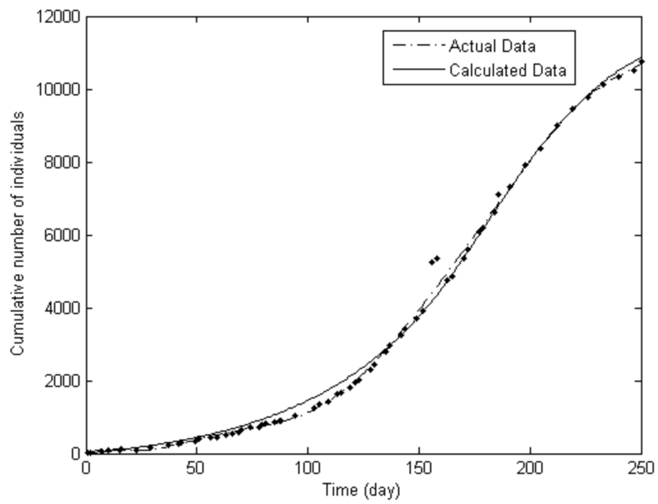

Figure 4b. Sierra Leone.

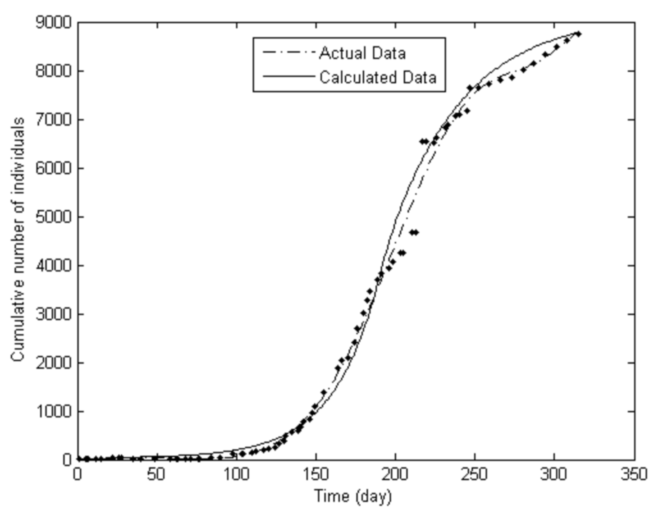

Figure 4c. Liberia.

Figure 4. Cumulative number of infections for Guinea, Sierra Leone and Liberia: Model predictions as dotted curve, actual data as dotted curve fitted to data points marked by Xs.

Table 3.
Parameter values for the SEIHCR (CT) model.

Parameter	Symbol	Guinea	Sierra Leone	Liberia	Combined
Population (millions)	N	12.0	6.0	4.1	22.1
Initial transmission rate outside hospital	β_{I0}	0.205	0.31	0.30	0.30
Transmission rate in hospital	β_H	0.0001	0.0001	0.0001	0.009
Transmission rate by corpses	β_C	0.224	0.125	0.23	0.125
Rate of identifying/isolating infectious	α	0.2	0.2	0.2	0.2
Days from symptoms to recovery, outside	$1/\gamma_I$	32	30	32	30
Days from symptoms to recovery, in hospital	$1/\gamma_H$	10	21	10	12
Days from symptoms to death, outside	$1/\nu_I$	8	8	8	8
Days from symptoms to death, in hospital	$1/\nu_H$	40	31	40	30
Days from death to proper handling of corpse	ψ	5.5	5.5	6.2	5.5
Traced contacts per identified infectious	κ	0	0	0	0
Transmission decay rate	k	0.03	0.015	0.03	0.021
Days to start of control measures	τ	200	160	190	160

Pharmaceutical Intervention

If medicine and vaccines can be available to patients and healthy people, the survival rate of the patients will increase and the situation of affected countries will improve to a great extent. Therefore, the manufacturing and delivery of medicine should remain an important factor in combating Ebola. Our tasks are to calculate the total amount of medicine, design a favorable delivery system, and compare the intervention extent of different manufacturing speed and different level of medicine efficacy.

Total Quantity of Medicine

Results from WHO Statistics

In deciding the quantity of medicine needed, we consider two important factors:

- the current number of patients, and
- the rate of increase of the disease.

WHO has been releasing a weekly report of newly confirmed cases by district, which gives the cumulative number of new patients [WHO 2015b]. Dividing by the duration (in this case, one week) gives the rate of increase of the disease. Meanwhile, by adding the newly infected cases, we can know the current cumulative patients as well.

By adding the current number of patients and the increase of the disease (we place equal weights on these two factors), we can obtain the number of unit doses of the medicine needed daily.

Due to the large number of cities (56) and the volume of the reported data (57 weeks till 1 February 2015), we present only aggregated results for each country (**Figure 5**).

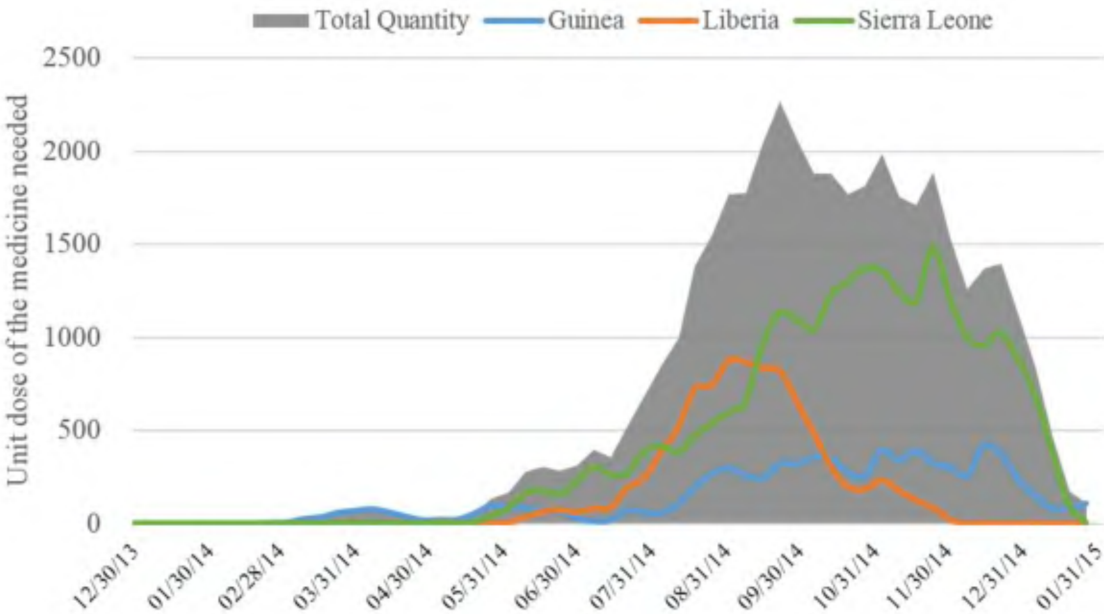

Figure 5. Number of doses of the medicine needed per day for Guinea (blue curve, with lowest peak), Liberia (orange curve, with second lowest peak), and Sierra Leone (green curve, with highest peak). The greyed area is the total doses needed by the three countries together.

Results from the SEIHCR (CT) Model

We take one city as an example to test the feasibility of our model, choosing Freetown, the capital of Sierra Leone, because this city has the largest number of cumulative patients.

We calculate from the SEIHR (CT) model the number of patients in Freetown and the rate of increase of the disease, week by week, and derive the quantity of needed medicine per day.

Figure 6a presents the cumulative numbers of patients in Freetown. In **Figure 6b**, the shaded area represents the total quantity of medicine as predicted from the model, while the solid curve is fitted WHO statistics. The two results match, with discrepancies within an acceptable range.

To conclude: SEIHCR (HC) can serve as a feasible model for both predicting the cumulative patients and determining the quantity of needed medicine.

Delivery System

To design the delivery system for medicine and vaccines, we

- Decide the locations for receiving drugs, then
- calculate the amount of medicine to deliver to each location.

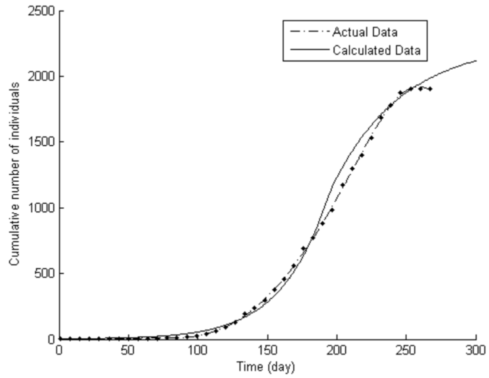

Figure 6a.

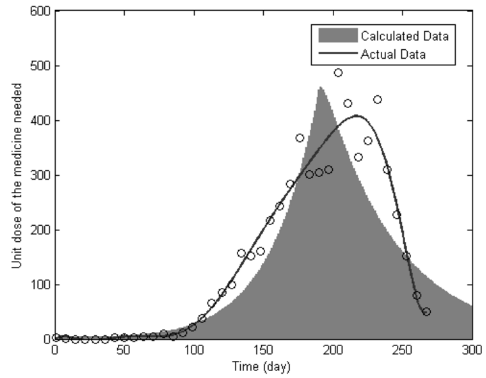

Figure 6b.

Figure 6. Cumulative number of patients in Freetown and quantity of needed medicine per week.

Locations for Delivery

We divide each country into areas and use the Fuzzy c-Means Clustering Algorithm to determine a depot for each. We then determine from which depot other cities will receive the drugs.

Weighted Fuzzy c-Means Clustering Algorithm

The key idea of the Fuzzy c-Means Clustering Algorithm is to represent the similarity a point shares with each of c clusters by means of a function whose values (memberships) are between 0 and 1 [Bezdek et al. 1984]. Each point has a degree of membership in every cluster. Membership close to 1 signifies a high degree of similarity between the point and the cluster, while membership close to zero implies little similarity.

The Weighted Fuzzy c-Means Clustering Algorithm differs from the traditional Fuzzy c-Means Clustering Algorithm in considering the quantity of medicine needed by each city.

Let $Y = \{y_1, \ldots, y_N\}$ be observations in R^n, with $y_k = [y_{k1}, \ldots, y_{kn}]$ an n-dimensional vector. Let $U = [u_{ik}]$ be a real $c \times N$ matrix, the matrix representation of the partition$\{y_i\}$ whose elements satisfy, for all k,

$$\sum_{i=1}^{n} u_{ik} = 1.$$

We denote the sets of all fuzzy c-partitions of Y by

$$M_{fc} = \{U_{c\times N} \mid u_{ik} \in [0,1]\}.$$

The most popular and well-studied clustering criterion for identifying optimal fuzzy c-partitions of Y is associated with the generalized least-squared errors functional. After weights, the quantity of medicine needed by each city, have been applied, the generalized least-squared errors functional can be written as

$$\min\ J_m(U,\nu) = \sum_{k=1}^{N}\sum_{i=1}^{c} D_k(u_{ik})^m ||y_k - \nu_i||^2,$$

$$\text{such that} \begin{cases} c < c_{\max}, \\ y_{ki\,\min} < y_{ki} < y_{ki\,\max}, \end{cases}$$

where

- m is a weighting exponent,
- D_k is the quantity of medicine needed by city k,
- $\nu = \{\nu_1, \ldots, \nu_c\}$ is the vector of centers of the clusters, and
- $\nu_i = \{\nu_{i1}, \ldots, \nu_{ic}\}$ is the location in n-space of the center of cluster i.

Optimal fuzzy clusterings of Y are defined as pairs $(\hat{U}, \hat{\nu})$ that locally minimize J_m. The necessary conditions for $m = 1$ are well-known. For $m > 1$, if $y_k \neq \hat{\nu}_j$, for all j and k, $(\hat{U}, \hat{\nu})$ may be locally optimal for J_m only if two particular equations hold [Bezdel et al. 1984]. [EDITOR'S NOTE: We omit the details.]

We can optimize via iteration by looping back and forth from these two equations until the iterate sequence shows only small changes in successive entries of $\hat{U}$ or $\hat{\nu}$. We formalize the general procedure for the Fuzzy Means Clustering Algorithm as follows:

We set $m = 2$. The value of c, the number of depots, should nether be too large nor too small. The objective function J_m decreases with c but more slowly as c is increased. We choose $c = 6$. Using WHO data, we calculate the means and membership matrix, carry out the algorithm, and obtain the six delivery locations shown in **Figure 7**.

Amount of Delivery

We apply the Fuzzy c-Means Clustering Algorithm to determine the number of unit doses needed for each depot, using 4 weeks as a cycle to reduce calculation. The average quantity needed is 106 unit doses per day. [EDITOR'S NOTE: We omit the details of the calculations.]

Speed of Manufacturing

Since the in-hospital fatality rate δ_H depends greatly on the amount of available medicine, if the daily rate u of manufacture is greater than the quantity D of medicine needed daily, we assume that the fatality rate will

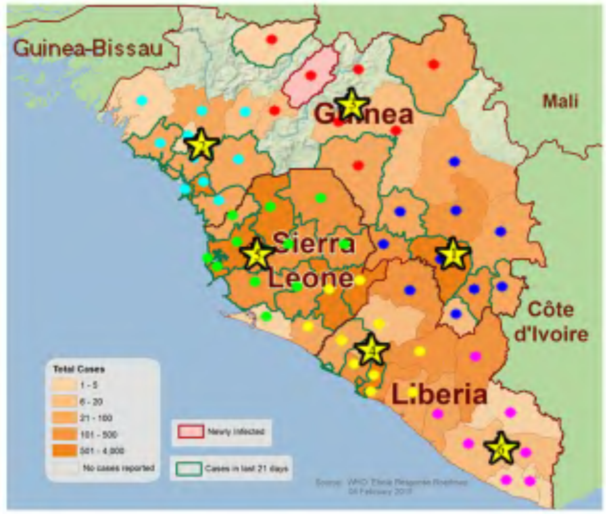

Figure 7. Depots (delivery locations) for medicine, indicated by numbered stars. Medicine is to be delivered from each star to the sites at colored dots surrounding it.

reach its lowest possible value δ_{H0}. However, if $u < D$, demand will not be wholly met and the fatality rate will be higher, with

$$\delta_H = \begin{cases} \delta_{H0} + \rho_u \left(1 - \dfrac{u}{D}\right), & u < D; \\ \delta_{H0}, & u \geq D, \end{cases}$$

where $\rho_u = 0.7$ and $\delta_{H0} = 20\%$.

If the rate of manufacture is at least 106, the needed number of unit doses per day, then the fatality rate is 20%. However, if the rate of manufacture is only half as much, the fatality rate is 57%, a huge increase..

We conclude: The rate of manufacture (and delivery) can greatly impact the efficiency of pharmaceutical intervention.

Other Important Interventions

Safer Treatment of Corpses

Swifter treatment of infectious corpses could help control the spread of Ebola. We vary the average time $1/\psi$ from 2 to 10 days. **Figure 8** shows the result for Sierra Leone, from which we observe that swift handling of corpses can make a big difference.

Intensified Contact Tracing and Earlier Isolation

To see the effect of intensified contact tracing and identification/isolation, we vary the two contact-tracing parameters

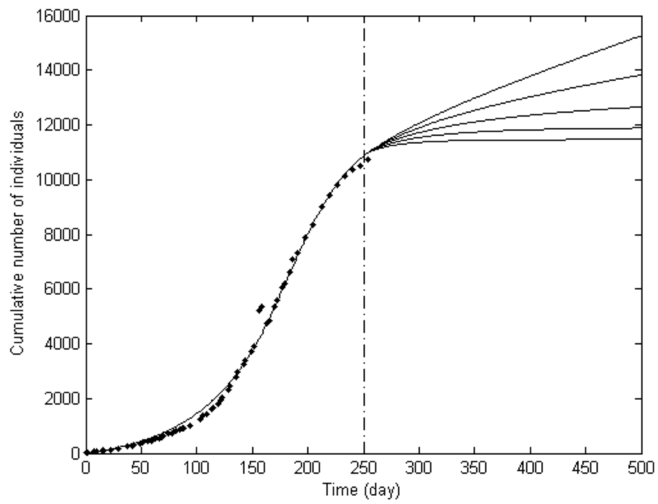

Figure 8. Simulation of predicted cumulative cases in Sierra Leone for $\psi = 2,\ 4,\ 6\,8\,10$.

- κ, the average number of contacts traced, and
- α, the rate of identifying and isolating infectious individuals

while keeping other parameters constant.

Figure 9a demonstrates that the number of contacts traced per case is important if the rate of identification and isolation of infectious individuals is low.

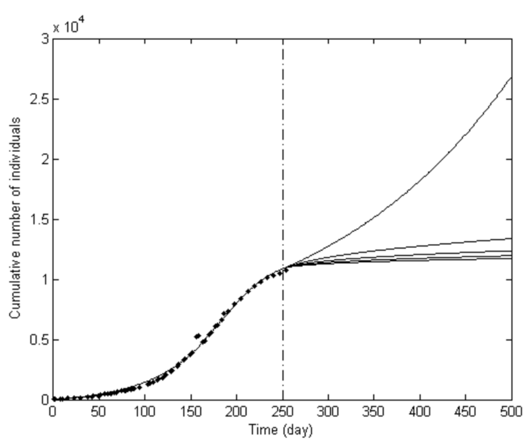

Figure 9a. $\alpha = 0.05$.

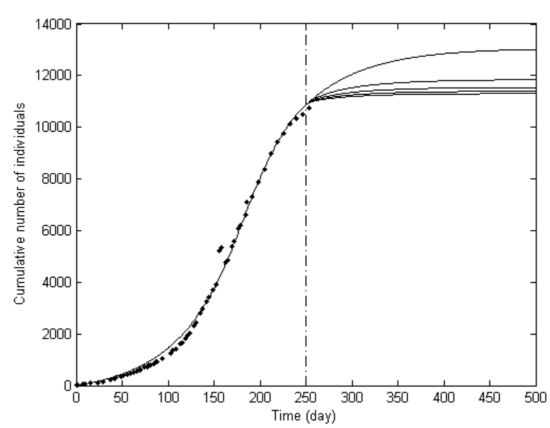

Figure 9b. $\alpha = 0.2$.

Figure 9. Simulation of predicted cumulative cases in Sierra Leone, for two different values of α, the rate of isolation, and each of the values $\kappa = 0$ (top curve), $10,\ 20,\ 30,\ 40$(bottom curve), the number of contacts traced. Note the differences in the vertical scales.

We conclude: Intensified contact tracing and earlier identification and isolation can be effective in eradicating Ebola.

Control and Eradication of Ebola

To predict the control and eradication of Ebola, we must consider all three countries together. For this situation, we take as values of the parameters the ones in the last column of **Table 2** (p. 216).

How Ebola Can Be Controlled

We define Ebola to be under control if the disease may still be growing but is not increasing its rate of growth:

- the first derivative of the cumulative number of cases is not 0: $\text{CUM}' \neq 0$; and
- the second derivative of the cumulative number of cases is less than or equal to 0: $\text{CUM}'' \leq 0$.

In **Figure 10** we trace the effects of four factors on controlling Ebola:

- α, the rate of identification/isolation of infectious individuals;
- κ, which reflects the intensity of contact tracing;
- $1/\gamma_H$, which reflects the cure rate; and
- δ_H, the in-hospital fatality rate.

The black straight line, and the region under it, represents that the situation is under control. The caption for each part of the figure gives the critical value of the parameter.

Figure 10a indicates that an increase in rate of identification/isolation of infectious individuals can lead to better control of the disease. The same is true for an increase in the the intensity of contact tracing (**Figure 10b**). Although decreasing the time to cure (**Figure 10c**) or decreasing the in-hospital fatality rate (**Figure 10d**) can lead to the control of Ebola, they are not effective approaches because the indicated critical values are highly unrealistic.

When Ebola Will Be Eradicated

Eradication of Ebola would mean no new cases, or

$$\text{CUM}' = 0.$$

Three parameters are especially essential in trying to eradicate Ebola:

- the introduction time τ of control measures,
- the efficacy η of the medicine, and
- the introduction time τ_{CT} of contact tracing.

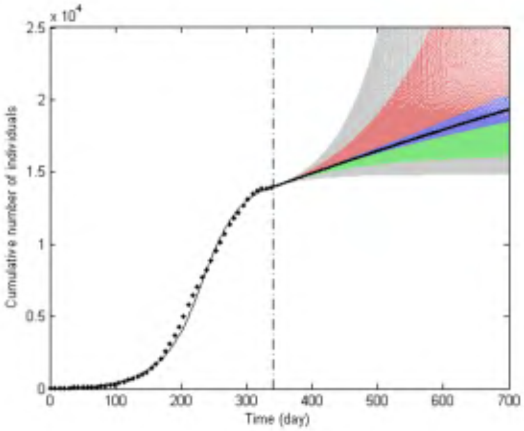

Figure 10a. $\alpha = 0.32$.

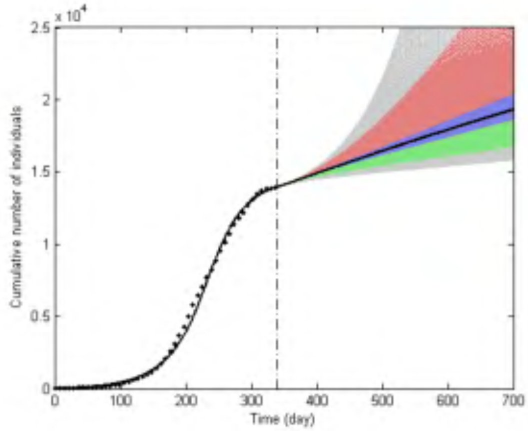

Figure 10b. $\kappa = 20$.

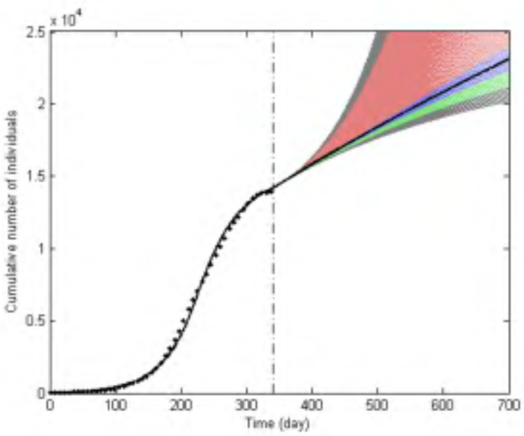

Figure 10c. $\gamma_H = 3.2$.

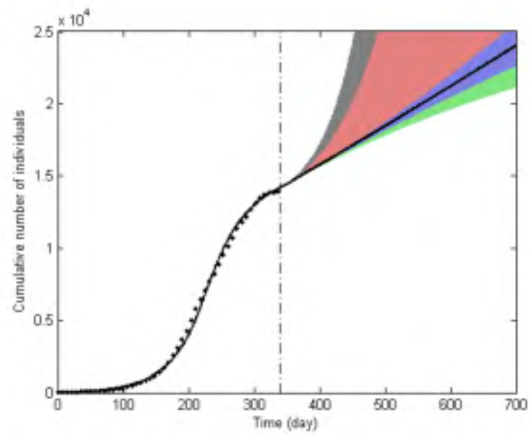

Figure 10d. $\delta_H = 0.07$.

Figure 10. The effect of each of four parameters on controlling the spread of Ebola, with the critical value for control identified.

By varying their values, we can see how these three parameters affect the time to eradication of Ebola.

The black line in **Figure 11a** represents the trend without change in the actual realized value of $\tau = 210$ days. We see that a decrease from 210 days to 145 days would reduce time to eradication by 200 days.

The black line in each of **Figures 11b** and **11c** also represent no change, while the dotted line represents time of eradication. Also worth noting from **Figure 11c** is that is that even a long delay until contact tracing won't cause the epidemic to expand without control.

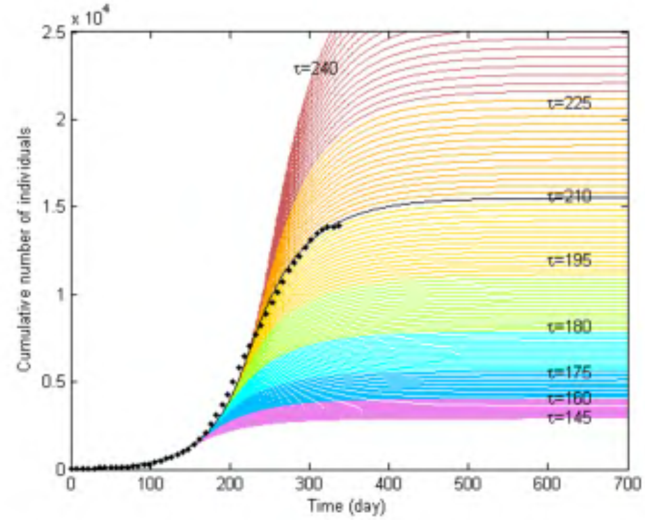

Figure 11a. Introduction time τ of control measures.

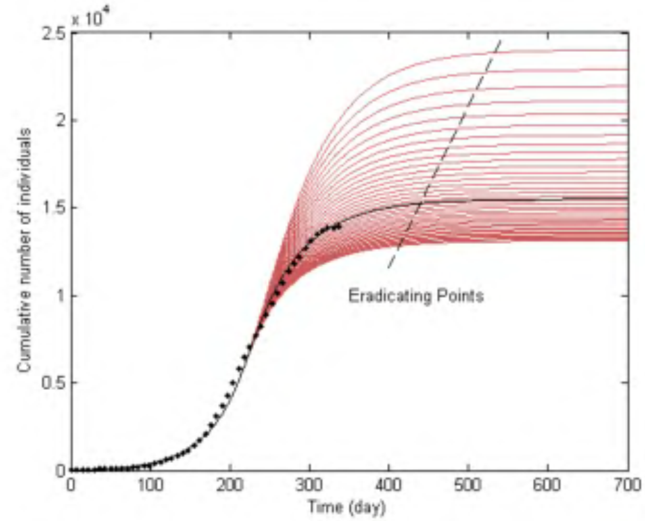

Figure 11b. Efficacy η of the medicine.

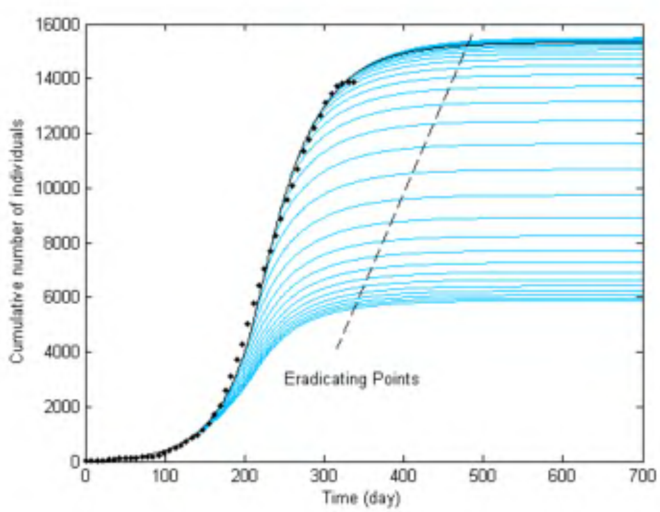

Figure 11c. Introduction time τ_{CT} of contact tracing.

Figure 11. Simulations of the eradication date for varying values of three parameters.

Sensitivity Analysis

Some parameters have a fixed value throughout our work. By varying their values, we can see their impact on the model.

Impact of Transmission Rates

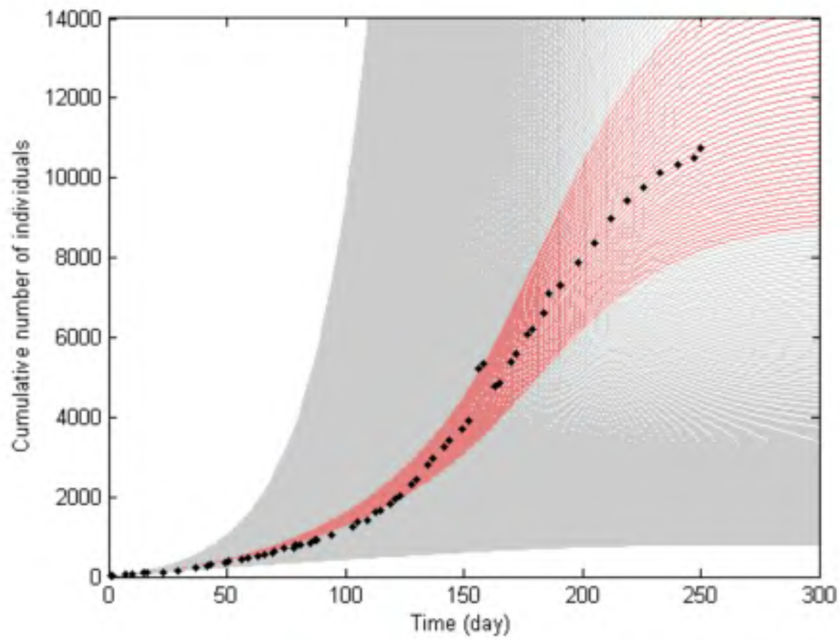

Figure 12. Impact of transmission rate on the cumulative number of cases.

The black dotted curve in **Figure 12** represents the actual statistics with transmission rate $\beta_I = 0.31$. The area (red) immediately surrounding this curve represents the range $(0.30, 0.32)$. We see that the transmission rate is a rather sensitive parameter. The same is true for the other transmission rates β_H and β_C.

Impact of the Incubation Period

The incubation period σ is 9 days for the reported data. **Figure 13** shows big variation for the values 5 days and 15 days, but both of those are beyond reasonable values.

Impact of the Fatality Rate

To test its sensitivity, we add noise to the fatality rate δ_H inside hospital. Values of 50% and 100% Gaussian noise make no appreciable difference in the cumulative number of cases. Since the fatality rate mainly reflects

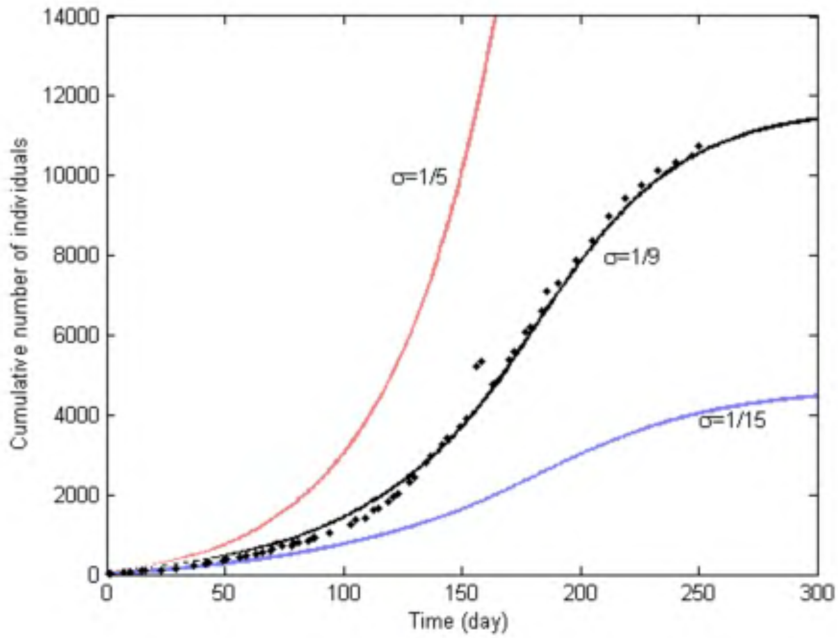

Figure 13. Impact of incubation period on the cumulative number of cases.

variability in the rate of manufacture of medicine and the efficacy of the medicine, we conclude that such variability would not have much impact on the cumulative number of cases.

Strengths and Weaknesses

Strengths

- We improve the traditional epidemic SEIR model, transforming it into the SEIHCR (CT) model which considers more factors and is thus better suited to the special features of West Africa.
- We divide the factors impacting Ebola into three parts:
  - factors that should be taken into consideration when improving the spread model;
  - factors regarding pharmaceutical intervention, such as medication efficacy;
  - other important interventions, such as treatment of corpses and contact tracing.
- We apply the Weighted Fuzzy c-Means Clustering Algorithm to determine delivery locations. We calculate daily needed medicine quantities for different cities from the SEIHCR (CT) model and use those as the weights in the algorithm, thereby making the algorithm an integral part of the whole epidemic model.

- We analyze the influence of various factors on the control and eradication of Ebola and predict time to final eradication.

Weaknesses

- The model has a large number of parameters, some of them poorly identifiable.
- Due to the large quantity of statistics for the cities in Guinea, Sierra Leone and Liberia, we selected only Freetown as an example for detailed analysis.

Future Work

- The values of some of the parameters are related to one another, and we would like to determine how. Eisenberg et al. [2015] offer insight into this problem.
- We could use our SEIHCR (CT) model to analyze more cities than just Freetown.
- We should test our model by applying it to other epidemics that share similar features with Ebola. Furthermore, we could test our model by applying it to earlier outbreaks of Ebola.
- All outbreaks of Ebola in history have been short. Why?

References

Althaus, Christian L. 2014. Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4169395/`.

Bezdek, James C., Robert Ehrlich, and William Full. 1984. FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences* 10 (2–3): 191–203. `http://www.researchgate.net/profile/James_Bezdek/publication/220580020_Fuzzy_C-means_cluster_analysis/links/0046351a132da435cd000000.pdf`.

Briand, Sylvie, Eric Bertherat, Paul Cox, et al. 2014. The international Ebola emergency. *New England Journal of Medicine* 371: 1180–1183. `http://www.nejm.org/doi/full/10.1056/NEJMp1409858`.

Browne, C., X. Huo, P. Magal, et al. 2014. A model of the 2014 Ebola epidemic in West Africa with contact tracing. `http://arxiv.org/pdf/1410.3817.pdf`.

Centers for Disease Control and Prevention (CDC). 2014a. Ebola virus disease treatment. http://www.cdc.gov/vhf/ebola/treatment/index.html.

________. 2014b. Contact tracing can stop Ebola in its tracks. http://www.cdc.gov/vhf/ebola/outbreaks/what-is-contact-tracing.html.

________. 2015a. Ebola outbreak in West Africa — Reported cases graphs. http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/cumulative-cases-graphs.html.

________. 2015b. 2014 Ebola outbreak in West Africa — Outbreak distribution map. http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/distribution-map.html.

Chowell, G., N.W. Hengartner, C. Castillo-Chavez, et al. 2004. The basic reproductive number of Ebola and the effects of public health measures: The cases of Congo and Uganda. *Journal of Theoretical Biology* 229: 119–126. http://chowell.lab.asu.edu/publication_pdfs/The%20basic%20reproductive%20number%20of%20Ebola%20and%20the%20effects%20of%20public%20health%20measures%20the%20cases%20of%20Congo%20and%20Uganda.pdf.

Chowell, Gerardo, and Hiroshi Nishiura. 2014. Transmission dynamics and control of Ebola virus disease (EVD): A review. *BMC Medicine* 12: 196. http://www.biomedcentral.com/1741-7015/12/196.

Driessche, P.V.D., and J. Watmough, J. 2002. Reproduction numbers and sub-threshold endemic equilibrium for compartmental models of disease transmission. *Mathematical Biosciences* 180 (November 2002): 29–48. http://www.researchgate.net/profile/James_Watmough/publication/11073829_Reproduction_numbers_and_sub-threshold_endemic_equilibria_for_compartmental_models_of_disease_transmission/links/02bfe51114449be4c9000000.pdf.

Eisenberg, Marisa C., Joseph N.S. Eisenberg, Jeremy P. D'Silva, et al. 2015. Modeling surveillance and interventions in the 2014 Ebola epidemic. http://arxiv.org/pdf/1501.05555.pdf.

Fisman, David, Edwin Khoo, and Ashleigh Tuite. 2014. Early epidemic dynamics of the West African 2014 Ebola outbreak: Estimates derived with a simple two-parameter model. doi:10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4169344/.

Gomes, Marcel F.C., Ana Pastore y Piontti, Luca Rossi, Dennis Chao, Ira Longini, M. Elizabeth Halloran, and Alessandro Vespignani. 2014. Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. doi:10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4169359/.

Legrand, J., R.F. Grais, P.Y. Boelle, A.J. Valleron, and A. Flahault. 2007. Understanding the dynamics of Ebola epidemics. *Epidemiology and Infection* 135 (4): 610–621. http://www.medbox.org/preview/53f50e73-e1ac-48da-bf2b-70f41fcc7b89/doc.pdf.

Lekone, Phenyo E., and Bärbel F. Finkenstädt. 2006. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics* 62 (4) (December 2006): 1170–1177. http://seor.gmu.edu/~klaskey/OR680/MSSEORProjectsSpring07/BAMWeb/My_Homepage_Files/Download/Statistical%20Inference%20in%20a%20Stochastic%20Epidemic%20SEIR%20Model.pdf.

Mundasad, Smitha. 2015. Ebola crisis: Experimental vaccine 'shipped to Liberia.' http://www.bbc.com/news/health-30943377.

New Zealand Ministry of Care. 2015. Case definitions for Ebola virus disease. http://www.health.govt.nz/our-work/diseases-and-conditions/ebola-updates/case-definitions-ebola-virus-disease.

Partners in Health. 2014. Need to know: How contact tracing can stop Ebola.
http://www.pih.org/media/need-to-know-contact-tracing.

Rivers, C.M., E.T. Lofgren, M. Marathe, et al. 2014. Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia. *PLOS Currents* (2014 November 6) http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4325479/.

Simon, Cory, and Bernard Korman. 2014. Mathemathinking: The mathematics behind the Ebola epidemic. http://mathemathinking.com/uncategorized/the-mathematics-behind-the-ebola-epidemic/.

Tetro, Jason. 2015. The Ebola vaccine situation: A cause for hope. http://www.popsci.com/ebola-vaccine-situation-cause-hope.

Valdez, L.D.,Rêgo, H.H. Aragão, H.E. Stanley, and L.A. Braunstein. 2015. Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies. http://arxiv.org/pdf/1502.01326v4.pdf.

World Health Organization (WHO). 2014. Ebola virus disease fact sheets, update Sep 2014. http://www.who.int/mediacentre/factsheets/fs103/en/.

________. 2015a. Ebola situation report. http://www.who.int/csr/disease/ebola/situation-reports/en/.

________. 2015b. Ebola data and statistics. http://apps.who.int/gho/data/node.ebola-sitrep.ebola-country?lang=en.

Yizhou Wang.

Xiaojian Yang.

Yunqi Zhu.

Advisor Ligong Wang.

Judges' Commentary: Eradicating Ebola

Kelly Black
Dept. of Mathematics
Clarkson University
P.O. Box 5815
Potsdam, NY 13699–5815
kjblack@gmail.com

Introduction

The Eradicating Ebola Problem required teams to estimate the optimal way to distribute a new medication to areas stricken by the Ebola virus, so as to eradicate the virus. The tasks included modeling the spread of the disease, determining how to ship and store the medications, and how to deliver the medications to people.

The majority of teams focused their efforts on modeling the spread of the disease. Many teams provided additional models to examine where to ship and store stockpiles of the medication as well as how to distribute the medication. A smaller number of teams were able to construct models that included the interactions between these two parts of their models.

We start the discussion of the teams' efforts with the modeling associated with the spread of Ebola. Next, we briefly indicate the distribution issues, followed by a brief note on how to examine the interactions between these two tasks. Once we have discussed the modeling, we examine the issue of parameter fitting, and then explore the practice of examining the sensitivity of the models. Finally, we offer a few notes about writing in general.

We do not provide here an overview of the judging process. It is an important consideration, described in previous commentaries [Black2009; Black 2011; Black 2013]. Faculty advisers and team members should be aware of the process, since it provides valuable insight into why particular parts of their document are important and how they are viewed at different stages in the judging process.

The UMAP Journal 36 (3) (2015) 237–244.

Disease Modeling

The vast majority of teams first discussed modeling the spread of Ebola. The most common model was a compartmental model, but a large minority of teams used cellular automata models. The compartmental models provide an excellent starting point for a whole population, but the cellular automata models make it easier to examine the spatial dimension of the spread of the disease.

Compartmental Models

Many teams with a compartmental model tended to start with an SIR (susceptible–infected–recovered) model [Murray 2008]. It was common then to follow with refinements such as an SEIR (susceptible–exposed–infected–recovered) model, or an SIQR (susceptible–infected–quarantined–recovered) model, or an SEIR model with a delay. The addition of an exposed class is motivated by the time required by the virus until the infected person is able to pass the disease to other people. The addition of the delay is motivated by the practices associated with local customs and how bodies are treated prior to burial.

One thing missing from many of the models that the teams examined is the impact of the spatial dynamics associated with the spread of the disease. The teams that did try to address the spatial component usually examined separate populations in different regions, and treated each population using their model. In some cases, the models were adapted to accommodate movement between the different regions.

Cellular Automata Models

In the cellular automata approach, a simple set of rules was determined, and the results of numerical experiments were provided. Just as in the compartmental models, it was common to divide the area in question into separate regions and then include terms to accommodate movement between the different regions.

The teams with a cellular automata approach had the added burden of formulating nontrivial rules to define the behavior of their computational model. Additionally, the numerical trials result in a random variable, so their results should be stated in the context of a statistical framework.

Distribution

Another part of the task included modeling the transport, storage, and the distribution of the medicines used to combat Ebola. This was one of the more difficult aspects of this year's event. The methods used to address this part of the question varied widely. The most common approach was simply to choose cities in areas with a high density of Ebola cases that also have the infrastructure to handle large shipments. The model was then constructed by treating the separate cities as nodes on a connected graph.

Most teams simply stated the locations with little or no motivation. Some teams provided a more careful analysis that motivated their choices, and some teams assembled models that calculated the impacts for different combinations of cities. The teams then devised schemes to search for the best choice of cities.

Once the locations were chosen, most teams built on the idea of a connected graph, with the cities the nodes of the graph. Once the model was constructed, an objective function was generally defined. Some teams created a weighted model that used the distances between nodes or the average distance from patients; others created a weighted model based on the need, which came from their model of the spread of Ebola; still others used a combination that tried to balance the distances with the need.

Once the locations were determined and the objective function defined, the teams then had to determine the best way to disseminate a fixed amount of medicine that would minimize/maximize their objective function. Because of the discrete nature of the models, the approaches generally included approaches such as a genetic algorithm, a clustering algorithm, or swarm optimization.

Connecting the Disease Model and the Distribution Model

The majority of teams were able to create, approximate, and describe their model for the spread of Ebola. A large number of teams were then able to create a model for the distribution of the medications that were available. Relatively few were able to bring the two models together and explore their interaction.

Most teams used their disease model to determine the level of needs in the various areas. Once the need was determined, they were then able to apply their distribution model to determine how to address the predicted need. The problem with this approach, though, is that the predicted need is closely tied to the ability to deliver medicines.

A small number of teams addressed this important aspect of the problem. Of those that did, many simply applied an iterative process in which they applied their disease model, determined the optimal way to deliver medicine, and then repeated the process to determine the need within the new context of the assumed availability of the medicines.

An even smaller number of teams were able to incorporate their disease model directly into the objective function that determined the level of need for a given region. This is an impressive accomplishment for a relatively short event. We certainly do not expect many teams to be able to complete such a difficult task, but it is important for a team at least to acknowledge this important aspect of the modeling effort.

One of the downsides of the MCM is that the idea of model refinement and the process of updating a model is skipped because of time constraints. It is important, though, for a team to at least recognize what should be changed and what should be retained when updating their model. This is why we expect the teams to include addressing the strengths and weaknesses of their model, and they should identify what they would consider changing or at least provide some idea of their priorities when refining their model.

Too many teams simply add an extra section titled "Strengths and Weaknesses" and then provide a bulleted list with brief sentences. This is certainly good; but to do better, the team should demonstrate that they have thought deeply about their model and can identify a deep understanding of the model that they created.

Parameter Fitting

Every year, teams develop models that include a large number of parameters. In this year's event, the teams were asked to examine a specific system and provide insight into the system. In this context, it is not appropriate to non-dimensionalize the system and provide an analysis of the overall behavior of the system.

A number of teams spent considerable time to determine the proper values of the parameters to use. Some of the choices came from examining a variety of papers whose focus is on the spread of the Ebola virus. In these cases, it could be difficult for the judges to determine how appropriate the choice of values are given that the original paper likely made use of a different model, and the values of the parameters are not necessarily the same. In these cases, though, the judges tended to give the teams the benefit of doubt considering the time constraints.

Other teams made use of various data sources and attempted to determine the best choice of the parameters, using a variety of approaches. Many teams used relatively straightforward least-squares estimation. This is a difficult task, and the judges made every effort to give the teams credit for their efforts. The primary issue is that the teams provide a strong written description of what they did and provide citations as well as references for other works they used.

Sensitivity

Every year, we say that sensitivity is important. Every year, this has been an area of the modeling process that tends to receive the least attention. This year was different. For the first time, we saw a large number of papers that provided an organized strategy to explore the sensitivity of the teams' models.

The majority of the teams' efforts focused on examining how the conclusions changed based on small changes in one or more parameters. This is an extremely important part of the modeling process, and the appearance of this simple exploration is an exciting development. We recognize that the time constraints make a complete sensitivity analysis difficult, but exploring the impact that different parameters have on the model is a relatively straightforward task.

In addition to the exploration of the parameters, for the first time we saw multiple teams that provided a brief analysis of the sensitivity of a model to basic assumptions. Several papers examined the impact of changing one or more assumptions, changing their model, and then examining how the conclusions change. This is an exciting development and represents a nontrivial leap in the modeling process.

The teams that were able to conduct this kind of analysis explicitly demonstrated a deep understanding of the modeling process. If this trend continues, then it will soon become vital that a team conduct some kind of basic sensitivity analysis in order to be recognized in the highest tier of the papers. This year's event required that teams make important predictions. It is important for the teams to provide some level of insight into how much trust can be placed in their predictions, and a sensitivity analysis is an important tool to determine how robust a model is.

Miscellaneous Comments

We survey a variety of unrelated but common issues that we observed. These issues include the use of figures and annotation of figures, citations and references, and the adaptation of existing models. The problems associated with these aspects of writing and modeling are important topics that represent stumbling blocks for a large number of teams. These are areas that advisers can discuss before the competition, and treatment of them can make a big difference in how a team's work will be received by a judge.

Figures

Every year we see large numbers of teams that provide a long list of figures but provide little to no discussion about the figures. The assumption is that a person reading the paper will understand what is important about a figure and will interpret the figures in the same way the teams interpret the figures.

Every figure should be described and discussed in the narrative of the report. The team should explain what is in the figure and why it is important. For example, this year it was common to see multiple approximations for the number of people infected by the Ebola virus from different regions, but it was not always clear what conclusion could be drawn from approximations that were not clearly labeled.

Additionally, every figure should have a title. Every figure should have a caption. The axes should be clearly labeled including units. A large number of teams provide plots without labels on the axes!

Use of References and Citations

Many teams include a list of references at the end of their report. A minority of teams provide citations within the narrative. To present ideas and not clearly provide the source of the motivation and where the idea originates from is plagiarism. A paper that provides clear citations and consistent references stands apart from many of the other reports.

How a Model is Adapted from Other Models

This is a source of concern every year, but this year it was a bigger issue than usual. Most of the models presented for the modeling of the spread of the Ebola virus made use of standard models and then adapted the model to accommodate the specific circumstances in this case.

A large number of teams simply assumed that the reader is aware of those standard models and that those models require little explanation. However, it can be difficult to understand a paper when a model is simply stated with little motivation or little discussion.

When a model is provided, the different parts of the system should be discussed, and the individual terms should be discussed. In doing so, the team has an opportunity to impress the judges and let us know that they understand the model. If a model is stated with no overview and no discussion, then we cannot determine if it is copied from some other source or has been adapted from other sources. We also cannot determine if the team understands the model or is simply restating something that they happened to find.

Discussing the model and providing motivation for how a model is adapted results in a paper that is much easier to read. When the paper is easier to read, and has proper citations, it immediately makes a strong impression on a judge. In the early stages of the judging, such a paper sends a message that the paper should be read in more detail. In the later stages of the judging, it sends a message that the team was respectful of the work of others, and they have a deeper understanding of the work.

Conclusion

The Eradicating Ebola Problem required teams to examine the connection between two models. Each topic on its own—disease modeling and distribution of resources—is a topic of intense interest in the research community. Many teams were able to build on the existing work and extend existing models. A smaller number of teams were able to bring the disparate models together into a cohesive whole.

In addition to the creation of the models, the teams had to determine values for the resulting parameters. This was done in a variety of ways, and teams that were able to discuss their approach and included particular details were more likely to make a stronger impression on the people reading their work.

Another important aspect of modeling is the analysis of the resulting model. In this year's event, we were pleasantly surprised to see that teams provided a much stronger analysis of their models than in the past. In particular, many teams provided a sensitivity analysis with respect to parameters, and a small number of teams went further with a sensitivity analysis with respect to their assumptions.

Finally, as is the case every year many teams had difficulties in using figures and graphs to share their insights about their results. Also, a large number of teams failed to provide citations within their narrative as well as consistent references. Additionally, many teams struggled with how to discuss their model and share how they adapted and extended other models.

References

Black, Kelly. 2009. Judge's Commentary: The Outstanding Traffic Circle papers. *The UMAP Journal* 30 (3) (2009): 305–311.

________. 2011. Judges' Commentary: The Outstanding Snowboard Course papers. *The UMAP Journal* 32 (2) (2011): 123–129.

________. 2013. Judges' Commentary: The Ultimate Brownie Pan papers. *The UMAP Journal* 34 (2) (2013): 141–149.

Centers for Disease Control and Prevention (U.S.). 2014. Generic Ebola response (ER): Modeling the spread of disease impact & intervention. Version 2.3. `http://dx.doi.org/10.15620/cdc.24900`. 11 September 2014. Last accessed 27 June 2015.

Meltzer, Martin I., Charisma Y. Atkins, Scott Santibanez, Barbara Knust, Brett W. Petersen, Elizabeth D. Ervin, Stuart T. Nichol, Inger K. Damon, and Michael L. Washington. 2014. *Estimating the future number of cases in the Ebola epidemic—Liberia and Sierra Leone, 2014–2015. Morbidity and Mortality Weekly Report (MMWR)* Supplements September 26, 2014 / 63(03); 1-14. `http://www.cdc.gov/mmwr/preview/mmwrhtml/su6303a1.htm`. Last accessed 27 June 2015.

Murray, J.D. 2008. *Mathematical Biology: I: An Introduction*. 3rd ed. New York: Springer-Verlag.

About the Author

Kelly Black is a faculty member in the Department of Mathematics and Computer Science at Clarkson University. He received his undergraduate degree in Mathematics and Computer Science from Rose-Hulman Institute of Technology and his daster's and Ph.D. degrees from the Applied Mathematics program at Brown University. He has wide-ranging research interests including laser simulations, ecology, and spectral methods for the approximation of partial differential equations.

Judge's Commentary: The Ben Fusaro Award for 2015

Thomas Fitzke
Dept. of Mathematics
Francis Marion University
Florence, SC
tfitzkee@Fmarion.edu

Introduction

The Ben Fusaro Award honors the Founding Director of the MCM (who continues to serve as a judge for the contest). First awarded in 2004, it recognizes an entry for an "especially creative approach" to the discrete problem in the contest.

The Fusaro Award for 2015 goes to the Outstanding team from Central South University, Changsha, China. Their paper, one of the top group of papers designated as Outstanding, stands out for a remarkably improved model and detailed analysis.

Problem Statement

The world medical association has announced that their new medication could stop Ebola and cure patients whose disease is not advanced. Build a realistic, sensible, and useful model that considers not only the spread of the disease, the quantity of the medicine needed, possible feasible delivery systems, locations of delivery, speed of manufacturing of the vaccine or drug, but also any other critical factors your team considers necessary as part of the model to optimize the eradication of Ebola, or at least its current strain. In addition to your modeling approach for the contest, prepare a 1–2 page non-technical letter for the world medical association to use in their announcement.

The UMAP Journal 36 (3) (2015) 245–247.

My Comments

Here are my comments on the paper, organized by the categories deemed appropriate for this problem by the judges, who gave the most credit for model development, analysis / validation, and conclusions.

Summary

As always, a strong summary is essential. The summary sets the scope and conclusions of the paper. This summary presents a good overall plan for the paper, listing in detail the models and considerations of the team. Some conclusions or calculated results would improve the paper even more.

Format, Clarity, Writing

An important component of papers is the writing style. Papers should be easy to read with correct English grammar. Although variation in the nouns, verbs, and adjectives used is understandable, correct use of noun verb relationships is necessary. Tables and graphs should be labeled and described in the text. This paper does an adequate job in its writing style. More notably, its tables and graphs are referenced and explained in great detail. They provide important analysis of simulations and listing of results.

Assumptions

The assumptions are well-thought-out and reasonable. Additional justification in some instances would show the reader why the assumption is needed.

Model Development

Model development is an outstanding achievement of this paper. The improvements to the SEIR model in the presentation of the SEIIR model with intervention and with medication are outstanding. The models are presented clearly and not rushed into. The variables and equations are not thrown at the reader without explanation as with most other papers, but rather defined and described before being stated.

The table and graphs are also explained well. They illustrate the analysis and results of models. Projected values from models can be easily compared with actual data. Graphs or pictures included from other sources should be denoted with the appropriate resource.

Sensitivity Analysis

The paper includes extensive analysis of several parameters in their model, which is more than almost all entries. This analysis shows that the team has full understanding of their models. This leads to reasonable listing of strengths and weaknesses.

Strengths/Weaknesses

The paper concludes with a collection of relevant observations. It is clear that with more time and effort the models could be expanded to encompass many more features such as treatment centers at locations other than capitals. But still this paper is a remarkable achievement for a few days' work.

About the Author

Dr. Fitzkee is the Professor and Chair of Mathematics at Francis Marion University in Florence, South Carolina. He has been a national judge for the MCM for two years and the HiMCM since 2009. He is a member of the MAA Minicourse Committee since 2011 and was coordinator of the Francis Marion Undergraduate Mathematics Conference in 2004–06.

Searching for a Lost Plane: A Neighborhood-Based Probabilistic Model

Melissa Jay
Nathan Mankovich
Eleanore Campbell

Colorado College
Colorado Springs, CO

Advisor: Andrea Bruder

Abstract

We propose a model that determines the most probable locations of a missing transoceanic flight. Since transoceanic flights are not detected by radar when they are more than 200 miles from a coast, radio communication and the occasional satellite imagery is essential to an effective search plan. Our search plan is built on the last distance in miles on the intended flight path (IFP) where air traffic control heard from the pilot.

We create a quadratic distance-to-casualty probability density function to predict the distance from the point of last contact to where the plane likely fell. Through trajectories and the possibility of veering off the IFP, we determine a two-dimensional search region and discretize it into smaller areas. The trajectories account for any model of aircraft, since parameters such as cruising altitude and glide ratio are assessed. Probabilities of containment (POC) (probability that the lost plane is in the given region) are then assigned to each search region.

We next consider the probabilities of detection (POD) during the search. We apply Koopman's equation for the probability of detecting a missing boat, which provides flexibility for the model of a search plane and its instrumentation. Upon determining POD, we create an algorithm to search the bounded area. Our algorithm applies the POD to the POCs and redistributes probabilities if the lost plane is not found in a region while still considering the chance that the plane was actually in the region searched. If the plane is not found in a given cell, the algorithm outputs a neighboring region to search next. This customized neighborhood search plan maximizes resources while remaining time-efficient. An example of a hypothetical missing Transatlantic Flight ABCD is included as a demonstration of our methodology.

The UMAP Journal 36 (3) (2015) 249–267.

Our search plan is built on assumptions, include that the IFP is linear and that the plane disassembled only upon impacting the water. Our model is sensitive to specific parameter changes but accommodates different plane types. Strengths of the model include non-uniform probability distributions modeling distance to casualty and impact and a neighborhood-based, continuously-updating search mechanism.

Introduction

Since 1948, 83 aircraft have vanished, some of which have still not been found [Merrill et al. 2014]. Lost transoceanic flights are among the hardest flights to locate due to the lack of radar coverage. Average radar coverage is approximately 200 miles in any direction [Hartmann and Diehn 2014]. Beyond radar coverage, air traffic control (ATC) relies on radio or messaging communication with a pilot on a pre-determined regular basis. If there are no location updates in the designated time frame, signals alert ATC that there has been a loss of communication.

In the past, Bayesian statistical methods have been used to track the whereabouts of missing flights. Air France Flight 447 went down in 2009 and was not found for over a year. Toward the end of the search, statisticians used Bayesian methods, narrowed the search plane, and picked up the blackbox's radio transmitter signal, locating the plane [Seidel 2014]. The available prior information, such as weather conditions on the flight path, floating bodies found soon after, and the signal of the beacon eliminated search regions. On the other hand, Malaysian Flight 370 contained little prior information, the search began later, no floating remnants were found, and individuals presumed that the blackbox signal had been turned off (Seidel). To this day, the location of flight MH370 remains unknown [Seidel 2014].

We create a mathematical model and search plan. Our model accounts for an indefinite time until casualty and impact with the water. We consider potential deviations from the intended flight path (relevant to the Malaysian flight) and prioritize time-efficiency.

Model

We consider the probabilities of containment of the plane's whereabouts, the probabilities of detecting the plane while efficiently and conservatively using resources, and the continuously updating probabilities of the most likely location of the missing airplane.

Constructing Probabilities of Containment

The containment model maps the probability of the airplane's containment to specific regions surrounding the flight's last known position.

Our first consideration is estimating how far the plane most likely traveled after the signal was lost. The variable y_0 is the distance (in miles) from the starting location when air traffic controllers last heard from the pilot. We assume that the plane stayed on its Intended Flight Path (IFP).

The random variable Y_c refers to the distance to casualty, or the distance along the IFP from the time at which the plane started to malfunction. We assume that when a plane malfunctions, its descent to the water begins. The distance to casualty is measured from $y_0 = 0$. We model the probability density function of Y_c as follows:

$$f_{Y_c}(y_c) = \begin{cases} \dfrac{3}{(y_B - y_R)^3}\left[y_C - (y_B - y_R)\right]^2, & y_0 \le y_C < y_B - y_R; \\ \\ 0, & \text{elsewhere,} \end{cases}$$

where

- y_B is the remaining distance to the planned destination from y_0, and
- y_R is the length of the final part of the journey where radar detection is available.

Therefore, $y_B - y_R$ represents the flying distance left before radar could have spotted the missing flight.

In framing the probability density function for Y_c, we assume that it is more likely for the plane to reach a casualty state immediately or soon after the last communication with ATC. This intuition is reasonable because a loss of communication often occurs due to abnormal plane operation. The probability continues to decrease quadratically and becomes 0 in the radar-detection region. The parameter y_R can be determined using the equation

$$y_R = \sqrt{(200)^2 - h^2} = \sqrt{40,000 - h^2},$$

where h is the cruising altitude in miles of the plane. Radar detectors can track planes within a 200 mile radius, on average [Hartmann and Diehn 2014]), so y_R can be determined via the Pythagorean theorem.

We determine the possible locations where the airplane could have fallen based on a distribution of trajectories. We let Y_I be the distance to impact, or the distance from y_0 at which the airplane hit the water. This variable takes into account the glide of the airplane:

$$Y_I = Y_C + Gh,$$

where G is the glide ratio of the plane [Benson n.d.]. Each airplane model has an average glide ratio, defined as

$$G = \frac{\text{units of distance forward}}{\text{unit of distance downward}}.$$

We multiply the glide ratio by the cruising altitude to determine the horizontal distance that the plane traveled during its downward fall. The probability density function of Y_I can therefore be modeled using the following transformation of variables technique:

$$\begin{aligned}
P(Y_I < y_I) &= P(Y_C + Gh < y_I) \\
&= P(Y_C < y_I - Gh) \\
&= \frac{3}{(y_B - y_R)^3} \int_0^{y_I - Gh} \big(y_C - (y_B - y_R)\big)^2 dy_C \\
&= \frac{3}{(y_B - y_R)^3} \left[\big(y_C - (y_B - y_R)\big)^2 dy_C\right]\Big|_0^{y_I - Gh} \\
&= \frac{3}{(y_B - y_R)^3} ([(y_I - Gh - y_B + y_R)^2][(0 - y_B + y_R)^2],
\end{aligned}$$

which implies

$$\begin{aligned}
f_{y_I}(y_I) &= \frac{d}{d_{y_I}} \frac{3}{(y_B - y_R)^3} ([(y_I - Gh - y_B + y_R)^2][(0 - y_B + y_R)^2] \\
&= \begin{cases} \dfrac{3\,[y_I - Gh - (y_B - y_R)]^2}{(y_B - y_R)^3}, & Gh \le y_I < y_B - y_R; \\ 0, & \text{elsewhere.} \end{cases}
\end{aligned}$$

The probable distances from the point of lost contact to the point of impact are measured along the IFP.

While this model has accounted for a single dimension, we can easily extend it to the left and to the right. The next portion of our model accounts for the possibility of veering off the intended linear path.

Veering away from the IFP is considered by assuming that possible locations of the plane are normally distributed to the sides of y_I. We determine the standard deviation σ_x of this normal probability distribution by

$$\sigma_x = Gh \sin \frac{\pi}{8}.$$

Since it is unlikely that a plane veers beyond $\frac{\pi}{8}$ from the IFP, $\frac{\pi}{8}$ will be used to find one standard deviation. Thus, the normal probability distribution

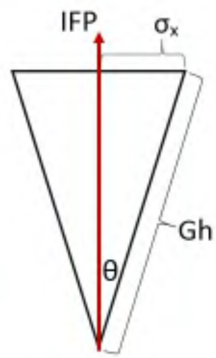

Figure 1. One standard deviation is calculated to be the distance from the IFP where a plane gliding $\frac{\pi}{8}$ off course would land.

perpendicular to the IFP is

$$f_X(x) = \frac{1}{Gh\left(\sin\frac{\pi}{8}\right)\sqrt{2\pi}} \exp\left[\frac{-x^2}{2\left(Gh\sin\frac{\pi}{8}\right)^2}\right], \qquad -\infty < x < \infty.$$

Because $f(x)$ and $f(y_I)$ are independent distributions, we can create a joint probability density function by multiplying $f(x)$ and $f(y_I)$:

$$f_{X,Y_I}(x,y_I) = f_X(x) f_{Y_I}(y_I)$$
$$= \begin{cases} \dfrac{3\left[y_I - Gh - (y_B - y_R)\right]^2}{Gh\sin\frac{\pi}{8}\sqrt{2\pi}(y_B - y_R)^3} \exp\left[\dfrac{-x^2}{2\left[Gh(\sin\frac{\pi}{8}\right]^2}\right], \\ \qquad\qquad -\infty < x < \infty,\ Gh \le y_I < y_B - y_R; \\ 0, \qquad \text{elsewhere.} \end{cases}$$

Partitioning the Search Region

We partition the search field and assign each region a probability of containment (POC) that the region contains the missing plane.

Since 99.7% of all observations are contained within 3 standard deviations of the mean of a normal distribution, we consider the search region bounded on both sides by three standard deviations. We segment y_I over its entire range. Therefore, we consider $N-1$ partitions and N vertical regions via the following equation:

$$N = \left\lceil \frac{y_B - y_R - Gh}{t_{0.5} v_s} \right\rceil,$$

where $t_{0.5}$ represents the quantity 0.5 hours and v_s is the velocity of the search plane. This segmentation guarantees that each search plane spends up to half an hour sweeping a length of each search region. Assuming that more than one plane searches a given region at a time, these partitions lead to a highly efficient search.

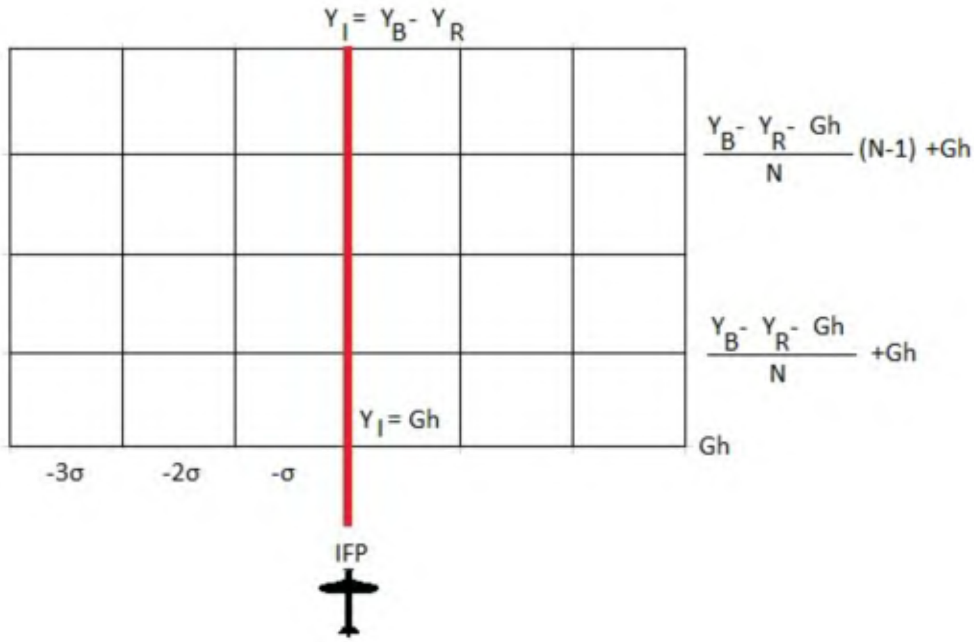

Figure 2. The discretization method for an $N \times 6$ search region. In this case, $N = 4$.

To find the probabilities of containment for each of the rectangular regions, we integrate the joint probability density function $f_{X,Y_I}(x, y_I)$ over the region.

Finding Latitude and Longitude for Search Regions

To direct the searchers, we must specify the boundaries of each region in latitude and longitude coordinates.

Initially, the bounds of a given uniform probability region $r_{u \times l}$ can be represented as $0 \leq u \leq N$ and $-3 \leq l \leq 3$ where l, u are integer multiples of $n =$ region number along the IFP and σ_x, respectively.

We define a coordinate plane with y-axis along the IFP, x-axis in the direction of σ_x, and origin at $(y_I, 0\sigma_x)$. The bounds of a region $r_{u \times l}$ in miles are:

$$-l\left(Gh \sin\frac{\pi}{8}\right) \leq x \leq l\left(Gh \sin\frac{\pi}{8}\right)$$

and

$$0 \leq y \leq \frac{u(||\mathbf{y_B} - \mathbf{y_0}|| - Gh)}{N}.$$

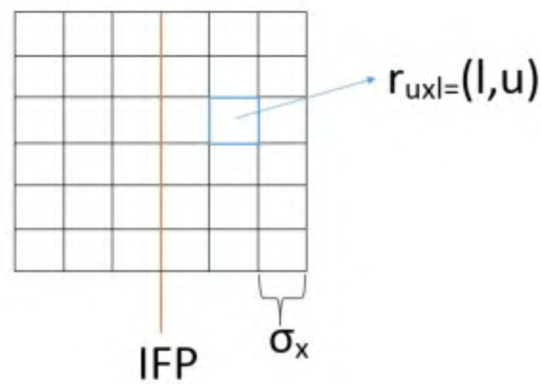

Figure 3. The location of a given $r_{u \times l} = (l, u)$ in reference to an IFP and σ_x.

To translate a number m of region-bounding coordinates in (x, y) to latitude and longitude, we build a $2 \times m$ matrix

$$C = [\mathbf{c_1} \ \ldots \ \mathbf{c_m}],$$

where each bounding coordinate $\mathbf{c_i}$, where $0 < i \leq m$ is of the form

$$\begin{bmatrix} c_x \\ c_y \end{bmatrix}.$$

Next, we use a rotation matrix, vector addition, and vector scaling to shift an "IFP-referenced" coordinate to latitude and longitude. $\mathbf{y_0}$ and $\mathbf{y_B}$ are known in the "IFP-referenced" system and in traditional coordinates. The components of $\mathbf{y_B} - \mathbf{y_0}$ in latitude and longitude are designated r_e, r_n. To build a rotation matrix from the "IFP-referenced" system to latitude and longitude, we first must find the counterclockwise angle ϕ between the two coordinate systems as follows:

$$\phi = \arctan\left(\frac{r_e}{r_n}\right).$$

Using ϕ, the rotation matrix R is either:

$$R = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}, \qquad \frac{\pi}{2} \leq \phi < 0$$

or

$$R = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \qquad 0 \leq \phi \leq \frac{\pi}{2}.$$

After finding a suitable R, we multiply C by R to express every $\mathbf{c}$ in C in miles of latitude and longitude with reference to $\mathbf{y_0}\rho$, where ρ is the average distance in miles of 1° North and 1° East from $\mathbf{y_0}$. Finally, we add the resulting vector to $\mathbf{y_0}$ to get coordinates in degrees North and East.

In summary, changing C in the "IFP-referenced" system to the corresponding coordinate matrix C_L in latitude and longitude amounts to:

$$C_L = \rho RC + \mathbf{y_0}.$$

Probability of Detection

A standard model for finding the probability of detection (POD), due to B.O. Koopman, is the exponential function

$$\text{POD} = 1 - e^{-\text{coverage}},$$

where coverage is the area effectively swept divided by the entire search area A_S [Frost 1999, Part II, 2 and 10].

We define the search area as

$$A_S = \left(Gh \sin\frac{\pi}{8}\right)\left(\frac{y_B - y_R - Gh}{N}\right),$$

where the A_S is the area of a single partitioned region of the overall search region. The area swept depends on the sweep width S and on the effort. The sweep width depends on the detection distance $\kappa\epsilon$ of the instrument (e.g., human eyes, radar, etc.) used for the search and the elevation h_s of the search plane.

We modify Koopman's model to account for variations in weather, conditions, search abilities, and more. We scale this variation as κ, which takes values between 0 and 1. The distance from the search plane's path projected on the ocean to the furthest possible detected object is $\sqrt{(\kappa\epsilon)^2 - h_s^2}$. Assuming that the search plane flies in parallel swaths, the sweep width would be

$$S = 2\sqrt{(\kappa\epsilon)^2 - h_s^2}$$

in order to cover all of A_S. The effort is said to be $d_s n$, where the distance searched by an individual aircraft per swath is d_s and the number of search planes included in the search is n. Assuming that the search plane flies parallel to the IFP, the number of search swaths is the number of subregions N, so d_s is

$$d_s = \frac{y_b - y_r - Gh}{N}.$$

We assume that search planes search simultaneously and at the same speed. Thus, d_s is the same for each plane. Incorporating each of these components into a single model yields the following POD equation:

$$\text{POD} = 1 - \exp\left[\frac{-2d_s n\sqrt{(\kappa\epsilon)^2 - h_s^2}}{\left(Gh\sin\frac{\pi}{8}\right)\frac{(y_B - y_R - Gh)}{N}}\right].$$

Unless the search effort is changed based on the search region, we can assume that the POD is uniform among all regions of probable containment.

The POC-Updating Algorithm

If there were no such thing as instrumentation error or human error, POD would equal 1 and thus we would not need to adjust the POCs of the discretized regions. Since search conditions are nowhere near perfect, we let $1 - \text{POD}_q$ to be the probability that the lost plane in region q went undetected. Given that we do not find the plane after searching the region q, we update the POC in q to be

$$\text{POC}_{q'} = (\text{POC}_q)(1 - \text{POD}_q).$$

This is the probability that the plane is actually in the searched region, even though it was not detected. Assuming that the sum of probabilities of containment over all discretized regions is constant, the probability lost when POC_q becomes $\text{POC}_{q'}$ must be accounted for in the other regions. To do this, we add

$$\frac{\text{POC}_q - \text{POC}_{q'}}{6N - 1}$$

to each of the POC values except $\text{POC}_{q'}$. This process can be iterated during a search to update all the POC values per searched discrete probability region.

Determining the Search Pattern

To determine the best order to search the $6 \times N$ different regions, we use the POC-updating algorithm described above, using gradients to choose the next search region.

By the "best" search order, we mean the most efficient in terms of resources and time. Thus, we want to minimize flight between nonadjacent regions, because that increases less-useful or redundant coverage and increases time needed to cover the entire area. The search plan algorithm involves updating the POC map after completing the search of one region to find the new probability of containment map, given that the downed aircraft was not detected in the recently-searched region. We represent the map of POC by a matrix in Mathematica and create a loop to evaluate the differences between the updated POC's of the just-searched region k and its adjacent neighbors. We take the minimum of the differences, and then locate the row n_{new} and column σ_{new} of that minimum element, to find the next region $r_{n_{\text{new}},\sigma_{\text{new}}}$ to search. Essentially, this method consists of updating the POC matrix, calculating a gradient/difference to identify the highest adjacent POC, and taking that region as the next to be searched.

Assumptions in the Methodology

- The lost plane's pre-established course (IFP) is a linear path from departure point to intended arrival point.
- The lost plane's glide affects its trajectory into the ocean. The plane does not explode in the sky.
- Any debris disassembles upon impacting the water.
- The plane will not veer more than $3Gh\sin(\frac{\pi}{8})$ from the IFP.
- No forces (e.g., currents) move the debris or plane from where they first landed.
- All planes in the model—the lost plane and the search plane(s)—fly at a constant speed.
- The lost plane is more likely to be found near the last point of communication.
- The lost plane remains detectable during the entire search.
- Every searching plane has the same instrumentation, searches at the same speed, and flies at the same altitude.

Sensitivity Analysis

We investigate the effects of parameters on our search plan.

We consider how the value of $f_{X,Y_I}(x, y_I)$ changes with G and with θ. In particular, we fix x and y_I and examine f as a function of G and then f as a function of θ, denoting those functions—with, for simplicity, a small abuse of notation—as $f(G)$ and $f(\theta)$.

From experimentation with values of x and y_I, we find that the shapes of the curves of $f(G)$ and $f(\theta)$ are strikingly similar (**Figure 4**). As G and θ increase, the respective $f(G)$ and $f(\theta)$ exhibit a rapid decrease. Since the bounded integral of $f_{X,Y_I}(x, y)$ produces a POC, G and θ have the same effect on POC as they do on $f_{X,Y_I}(x, y)$. In short, as $G + \theta$ decreases, the probability of finding the lost plane increases.

Increasing θ directly increases the search area, because the X distance is measured in mile-multiples of $\sigma_x = Gh\sin\theta$, which increases with θ.

We now examine the effects of G and of θ on POD.

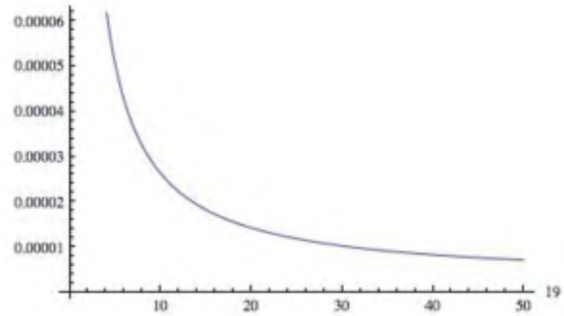

Figure 4. $f(G)$ with $(x, y_I) = (0.5, 0.5)$ exemplifies the curvature of $f(G)$ and of $f(\theta)$ that holds regardless of the x and y_I values.

As θ increases, POD rapidly decreases from 1 to 0 (**Figure 5**). The most intriguing aspect is that the graph suggests that POD$(0) \approx 1$. Also, POD is extremely small for any plane that has veered more than 90° off the IFP.

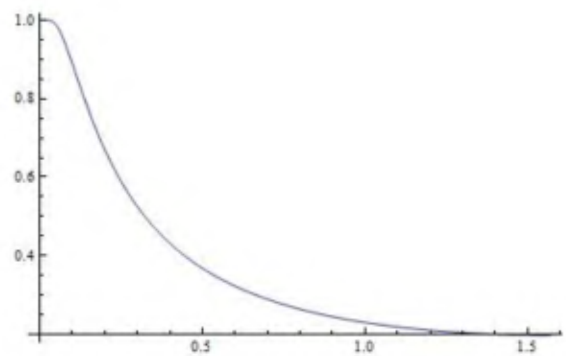

Figure 5. POD(θ), with θ in radians.

Increasing the glide ratio G affects the POD similarly to increasing θ, as shown in **Figure 6**. Increasing the exponent's denominator has an inverse effect on the POD. Thus, a search following the methodology explained in the **Working Example** will be more likely to find a plane with a small glide ratio.

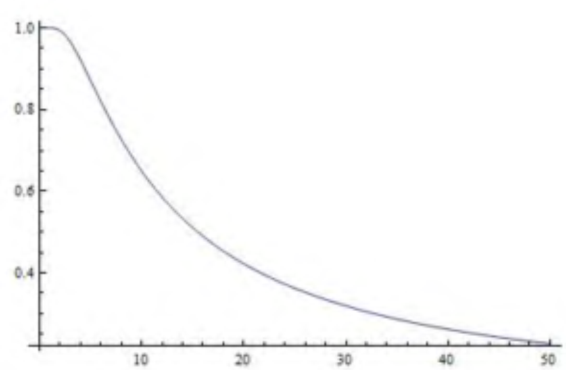

Figure 6. POD(G).

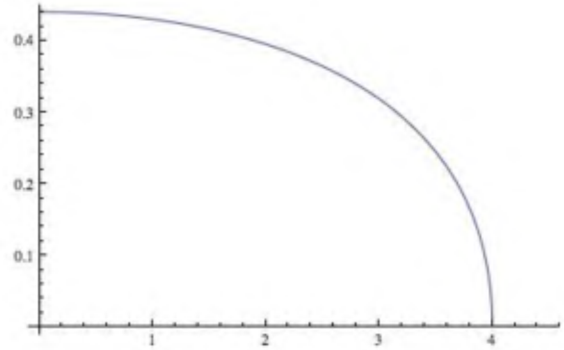

Figure 7. POD(h_s) for $0 \leq h_s < \epsilon$.

The graphs of **Figures 5** and **6** are generated using the constant parameter values found in the subsequent section **Working Example**.

Another variable that clearly affects the search plan is the cruising altitude h_S of the search plane, because the POD depends on $\sqrt{\kappa\epsilon^2 - h_s^2}$. Varying h_s within $0 \leq h_s < \epsilon$ produces an elliptic graph of POD centered around the origin: **Figure 7** illustrates that when h_s/ϵ increases, the POD increases. This is a logical relationship, because each swath covers more ground, which in turn decreases the search effort. Thus, it is best to run the search at the lowest possible altitude in order to improve the chance of finding the lost plane.

The number n of search planes also affects the POD, because n is in the numerator of the search effort fraction. **Figure 8** gives a graphic illustration. As expected, using more search planes increases the likelihood of detection.

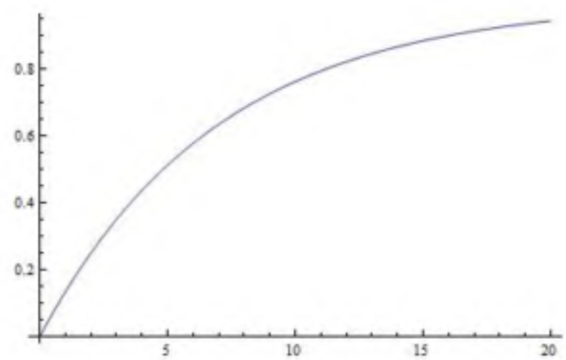

Figure 8. POD(n).

Of course, a search plane must fly low enough for the effective range of the instrumentation $(\kappa\epsilon)$ to reach the ground, or else the POD would be effectively zero.

Our model is sensitive to changes in several parameters and therefore may not be considered robust. However, the sensitivity is a necessary concomitant of a model that accommodates different types of planes and search instruments.

Strengths and Weaknesses of Our Model

Strengths

- We use a quadratic probability density function rather than the commonly-used uniform distribution. In many search-and-rescue methodologies, it is assumed that the time of casualty is uniformly distributed over all possible times in a given range [Stone 1977]. We instead construct a decreasing, concave-down distribution to account for the large likelihood that complications arise early after y_0.
- Our model also accounts for a flight trajectory to map the flight's motion at the time of casualty to a region of the ocean surface.
- We locate the search regions using latitude and longitude coordinates rather than tracking angles and distances from the point of lost contact.
- The POD incorporates several factors, including an added constant k that can incorporate search conditions. This constant helps to better explain the rate of coverage while still using Koopman's commonly-used Exponential Model.
- Our search plan is a constantly-updating algorithm that keeps time-efficiency in mind. If the lost plane is not detected in a highly-probable region, the POCs of the other search regions are updated. The cell with the next highest probability in the searched region's "neighborhood" is investigated. Implementing such a neighborhood restriction speeds up the search process, wasting no travel time while still searching regions with high POCs. Our search plan can easily be generalized to search planes of different speeds and larger or smaller regions of the ocean.

Weaknesses

- The POC of the regions closest to Gh but more than one standard deviation away in the X direction are slightly over-allocated. The standard deviation was calculated using trigonometry, indicating that our rectangular probability region is based on a triangular calculation (see **Figure 1**). Because of this over-allocation, there is a slight under-calculation of probabilities one or more standard deviations away from the IFP towards the latter partitions of Y_I.
- We do not consider ocean currents that may transport debris away from the original point of ocean contact.

Working Example

Imagine that on February 9, 2015, Flight ABCD of a Boeing 777 aircraft from Boston, MA to Lisbon, Portugal disappeared. The point of last communication occurred at coordinates 40.9275° N, 46.2575° W. Let $y_B = 1{,}912$ miles. The plane was cruising at 560 mph at an altitude of 40,000 feet (7.58 miles) over open ocean. The Lisbon airport is the nearest location that can detect an airplane, so $y_R = 199.86$ is measured from the Lisbon airport along the IFP.

We give below the probability distribution of Y_C, the distance in the direction of the IFP when the plane experiences a casualty that causes it to begin a downward trajectory.

$$f_{y_c}(y_c) = \begin{cases} \frac{3}{(1912-199.86)^3}\left[y_C - (1912 - 199.86)\right]^2, & 0 \le y_C < 1712.14; \\ 0, & \text{elsewhere.} \end{cases}$$

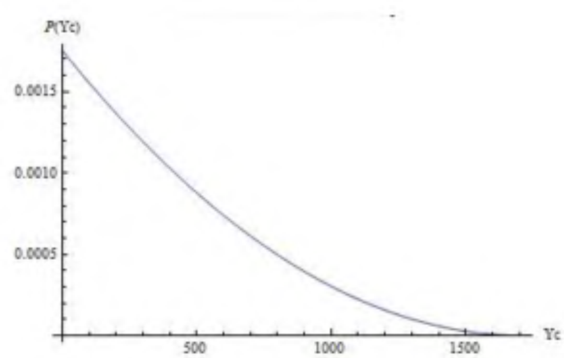

Figure 9. Probability density function for Y_C, the distance in the direction of the IFP when the plane begins a downward trajectory.

This airplane has a glide ratio of approximately 19:1, similar to other large aircraft. With this information, Y_I can be modeled as follows:

$$f_{Y_I}(y_I) = \begin{cases} \frac{3}{(1712.14)^3}(y_I - 1568.12)^2, & 144.02 \le y_I < 1712.14; \\ 0, & \text{elsewhere.} \end{cases}$$

After considering the distribution of distances of impact, we must now account for the possibility of veering off the IFP. The following X distribution, distributed in the direction perpendicular to the intended flight path, accounts for Flight ABCD's potential off-course location of impact:

$$f_X(x) = \frac{1}{\left(144.02 \sin\frac{\pi}{8}\right)\sqrt{2\pi}} \exp\left[\frac{-x^2}{\left(2(144.02 \sin\frac{\pi}{8}\right)^2}\right], \quad -\infty < x < \infty.$$

By multiplying the independent probability density functions $f_X(x)$ and $f_{Y_I}(y_I)$, we get the joint probability density function

$$f_{X,Y_I}(x, y_I) = \begin{cases} \dfrac{3(y_I - 1568.12)^2}{\left(144.02 \sin \frac{\pi}{8}\right) \sqrt{2\pi}(1712.14)^3} \exp\left[\dfrac{-x^2}{\left(2(144.02 \sin \frac{\pi}{8}\right)^2}\right], \\ \qquad\qquad -\infty < x < \infty, 144.02 \leq y_I < 1712.14; \\ 0, \qquad\qquad \text{elsewhere,} \end{cases}$$

which we exhibit in **Figure 10**.

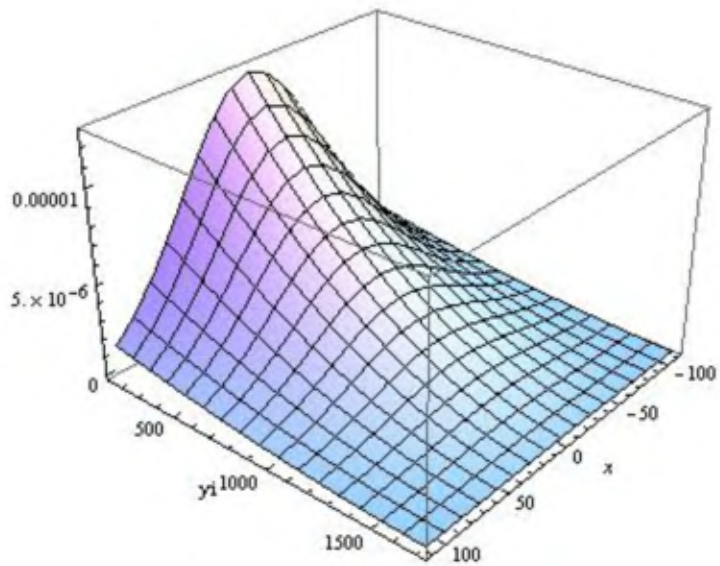

Figure 10. The joint probability density function for the POC.

The probability density function for the POC can be discretized into 36 subregions, as described earlier and illustrated in **Figure 11**.

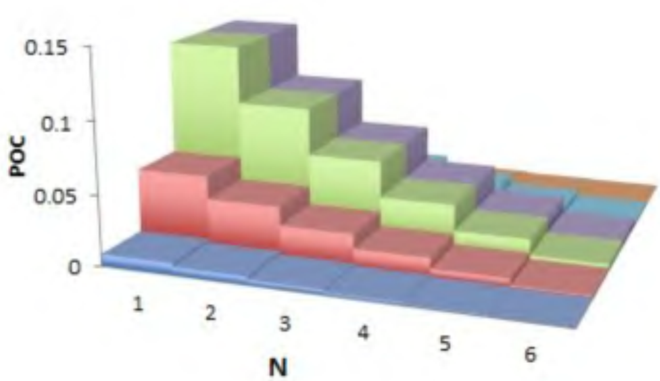

Figure 11. Discretized POC of Flight ABCD.

Next, we find the POD:

$$\text{POD} = 1 - \exp\left[-\frac{2d_s n\sqrt{(k\epsilon)^2 - h_s^2}}{\left(Gh\sin\frac{\pi}{8}\right)(y_B - y_R - Gh)/N}\right]$$

The list of parameters for this model, either estimated or researched values, is as follows:

- Remaining distance of flight from y_0 to y_R, $y_B = 1912$
- Glide constant for a large plane such as a Boeing 777-200, $G = 19$
- Average altitude at which a commercial airliner cruises, $h = 7.58$
- $y_R = \sqrt{200^2 - h^2}$
- Velocity of the search planes (all of which are the same model in order to facilitate easier coordination and timing in the search), $v_s = 550$
- $t_{.05} = 0.5$ hours
- The number of partitions in the y direction, $N = \left\lceil \frac{y_B - y_R - Gh}{t_{.05} v_s} \right\rceil$
- Distance traveled by search planes in searching a single region, $d_s = \dfrac{(y_B - y_R - Gh)}{np}$
- Detection distance of instrument, $\epsilon = 4$
- Scales detection distance based on search conditions such as weather, $\kappa = 0.98$, note $< 0\kappa < 1$. In this case, κ is near 1, indicating clear calm search conditions.
- Number of search planes n, chosen in this example to be 8.

Then, for our example, we have

$$\text{POD} \approx 0.68.$$

We input the values for the 36 regions in **Figure 11** into Mathematica to form the initial POC matrix. We then determine the order and path of the most efficient search pattern. We also utilize the model to update probabilities that certain regions contain the lost plane, given that the plane was not detected when a region was most recently searched.

Comparing **Figures 12** and **13**, we observe some key differences. When adjacency is not a requirement in our algorithm, frequently the subsequent search region is the reflection of the previous one due to the symmetry of the joint probability distribution across the IFP. This leads to much more back-and-forth movement, and as a result, more distance traveled, thus wasting time and fuel.

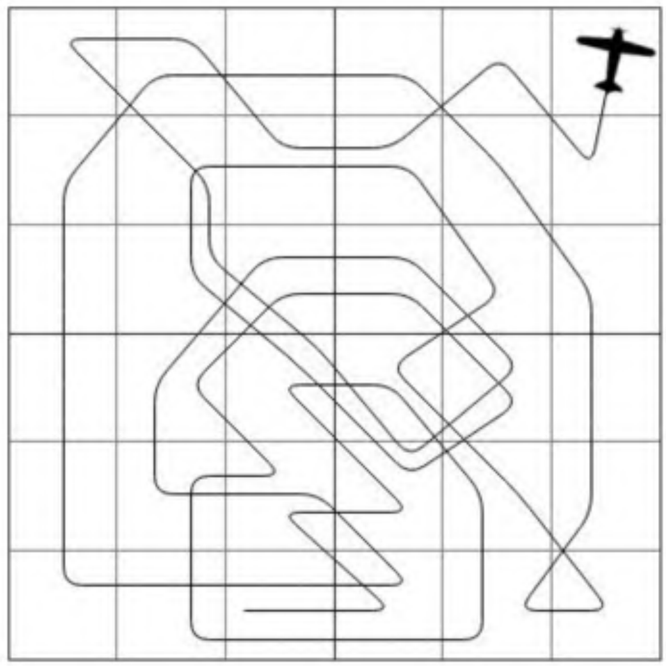

Figure 12. The search path produced by our adjacency algorithm.

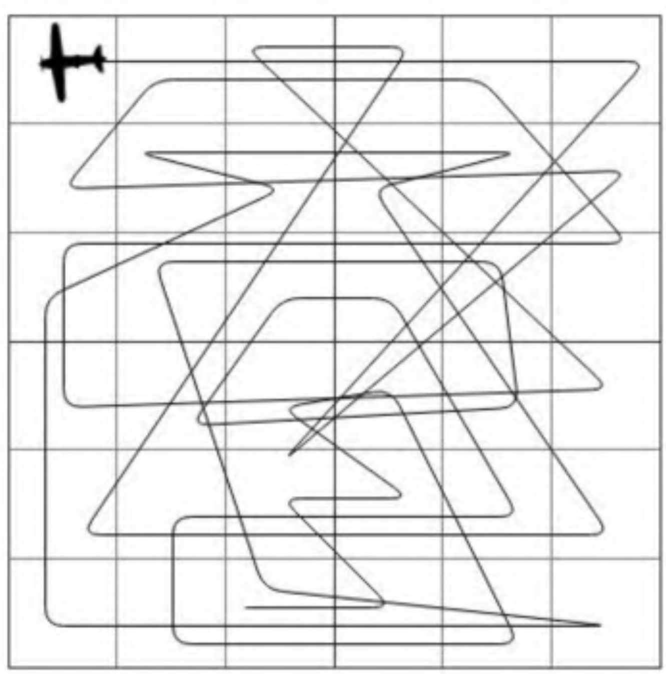

Figure 13. The search path produced by an algorithm similar to ours but without requiring subsequent search regions to be adjacent.

To search a certain region, that region must be described by latitude and longitude. Continuing with the example, say that the region in need of searching is the discrete probability region σ, N from 2 to 3. The algorithm in action is as follows: First, the rotation matrix

$$\begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} 0.9983 & -0.0579 \\ 0.0579 & 0.9983 \end{bmatrix}$$

is defined using the rotation angle $\phi = -0.057939$. The algorithm calculates the four latitude and longitude bounding coordinates of the rectangular search region σ, N from 2 to 3 as shown in Table 1.

Table 1.
Example of calculation of latitude and longitude bounding coordinates.

Shifted Coords:			
IFP	IFP	σ	σ
2	2	3	3
40.26564161	39.93471242	41.06305591	40.73212672
-34.84693598	-29.14165398	-34.80068276	-29.09540075

Conclusion

The final results of our model present a valid, time-efficient search plan for locating missing flight ABCD from Boston to Portugal. Our methodology accounts for all sizes and types of airplanes as well as all types of search planes. Our sensitivity analysis stresses the importance of careful consideration of resources and the number of search planes. A greater number of search planes drastically increases the probability of detecting the missing plane. While some parameters are more robust than others, carefully choosing the number of search planes per search region optimizes time-efficiency and overall detection. Our neighborhood-based search plan makes this model a strong choice when considering especially large search regions. While we did not have time, we hoped to incorporate the impacts of currents in the movement of airplane debris. We plan to further explore this topic to eventually extend our model to account for debris search plans too.

References

Benson, Tom. n.d. Glide angle. http://tinyurl.com/qfcftu2 = http://www.grc.nasa.gov/WWW/k-12/VirtualAero/BottleRocket/airplane/glidang.html .

Frost, J.R. 1999. Principles of search theory. http://www.sarinfo.bc.ca/Library/Planning/PrincSrchThry_S.pdf .

Hartmann, Christoph, and Sonya Angelica Diehn. 2014. The technologies behind the search for Malaysia Airlines Flight 370. http://dw.com/p/1BOg6 .

Stone, Lawrence D. 1977. Search theory: A mathematical theory for finding lost objects. *Mathematics Magazine* 50 (5) (1977): 248–256.

Merrill, Dave, Chloe Whitaker, and Alex Tribou. 2014. Dozens of planes have vanished in post-WWII era. http://tinyurl.com/nedeygb = http://www.bloomberg.com/infographics/2014-03-13/vanishing-planes-mapped-since-1948.html .

Seidel, Jamie. 2014. Missing Malaysia Airlines Flight MH370: Search goes back to basics turning to the power of maths to solve the mystery. http://tinyurl.com/lhzrufa = http://www.news.com.au/travel/travel-updates/missing-malaysia-airlines-flight-mh370-search-goes-back-to-basics-turning-to-the-power-of-maths-to-solve-the-mystery/story-fnizu68q-1226941119783 .

Melissa Jay, Andrea Bruder (team advisor), Eleanore Campbell, and Nathan Mankovich, with a celebratory cake in the foreground.

Judges' Commentary: Lost Plane Problem

Robert Burks
Defense Analysis Dept.
Naval Postgraduate School
Monterey, CA
reburks@nps.edu

The Problem

Recall the lost Malaysian flight MH370. Build a generic mathematical model that could assist "searchers" in planning a useful search for a lost plane feared to have crashed in open water such as the Atlantic, Pacific, Indian, Southern, or Arctic Ocean while flying from Point A to Point B. Assume that there are no signals from the downed plane. Your model should recognize that there are many different types of planes for which we might be searching and that there are many different types of search planes, often using different electronics or sensors. Additionally, prepare a 1–2 page nontechnical paper for the airlines to use in their press conferences concerning their plan for future searches.

Introduction and Overview

This year's Lost Plane Problem focused on identifying the factors and metrics for developing a coordinated search plan for a lost plane over open water. The problem required teams to develop a modeling approach based on these factors to achieve an undefined objective related to searching for a lost plane. In addition, there was the traditional required nontechnical paper, with this year's focus on providing the airlines information concerning the search plan for use in press conferences.

I start this commentary with a short review of the mechanics of this year's judging process. I follow the mechanics with a discussion and observations from the judging on various elements of the problem. I then discuss the importance of sensitivity analysis, assumptions and identifying

The UMAP Journal 36 (3) (2015) 269–276.

the strengths and weaknesses of a developed model. I finish by addressing some points concerning communication and conclude with a summary.

The Process

I believe that it is beneficial to once again review several elements of the judging process for this year's contest.

The criteria used to distinguish outstanding papers from good papers gradually changes as the judging progresses through the triage to the final rounds, with the final papers standing out as the best papers submitted under a wide variety of criteria.

Triage

The primary objective of the triage is to identify papers that should be given more detailed consideration from the judges. Every paper is read by at least two judges seeking to determine if the paper contains all of the required and necessary elements that make it a candidate for more detailed readings. If a paper addresses all of the issues and appears to have a reasonable model, then judges are likely to identify it as a paper that deserves more attention.

A paper must be clear and concise to do well in the triage, and the paper's summary is critical at this point in the judging. A good summary provides a brief overview of the problem, the paper's structure, and specific results stated in a clear and concise manner. Small things that make a paper stand out include having a table of contents and ensuring that all required questions are clearly addressed in the paper. Many papers do not do well in the triage because they fail to address all of the questions and the judge decides that a team's efforts will not compare well with the better papers. For example, one critical element overlooked by many papers this year was directly addressing the prompt that the model should recognize that there are many different types of search planes, often using different electronics or sensors.

Fully developing all of the required elements is a critical area often overlooked in papers. For example, sensitivity analysis remains one of the weakest elements and is often entirely missing in many papers—and these papers do not do well during the triage.

In addition, it is vital that the team express their general approach and results as clearly and concisely as possible in the nontechnical position paper. This means providing a broad overview of the problem, the approach, and specific results in clear, concise, nontechnical terms. In other words, can the nontechnical paper be read and understood by someone without an education in mathematics? In this year's problem, the aim of the nontechnical paper was to provide the airlines with a structure to organize a coordinated search process that the airlines could use in future press conferences. Many papers actually provided a press release versus a nontechnical paper to the airlines. This did not hurt papers during the triage process but became more of a factor as papers advanced through the judging process.

These small things make it much easier for a judge to identify the team's effort and for the paper to do well in the triage round. However, the best models and the best effort is not effective if the results are not adequately communicated. It is important to remember that this is a modeling competition and that effective communication is a critical part of the modeling process.

Final

The final consists of multiple rounds of judging over several days. As the rounds progress, the judging criteria shifts from identifying papers that warrant further consideration to a process of identify the very best papers.

The first round of the final begins with each judge reading a set of papers and then all judges meeting to discuss the key aspects of the question and what should be included in a "good" paper. This year these aspects included, in addition to all of the required elements,

- a clear discussion of the search process,
- the development of a truly integrated search plan that incorporated a variety of search assets, and
- emphasis on the sensitivity analysis portion of the paper.

As the final progresses, each paper is read multiple times with the final set of papers being read by all judges. In these last rounds, the modeling process and the mathematical integrity of a paper begin to identify the outstanding papers in the competition.

Components of the Question

This year's Lost Plane Problem consisted of one question, but successful approaches required modelers to address three major components:

- The first component required teams to discuss what happens from the moment of lost contact with the plane until the plane crashes in open water.
- The second component required teams to address potential spatial and temporal conditions that might impact the size and shape of the search area.
- The last component focused on developing an integrated search process to cover the identified search area.

As a general comment, many papers failed to develop a clear overall objective for their model: Was the objective to

- minimize the expected amount of time to find the plane?
- maximize the probability of finding the plane?
- minimize some function of the time and cost of the search?

As is the case with all models, there was a need for data. In many cases, teams forgot, or did not include, specific values for their data or for their model's parameters. For example, saying that "the data can be obtained from the literature" is inadequate.

As another example, many papers developed a model to provide the probability of the plane's location in the ocean, for which it was not unreasonable to use a normal distribution centered around the Intended Flight Path of the plane after lost contact. However, a number of papers including such a model and presenting a lovely picture of the normal distribution did not provide values for the mean and standard deviation of this distribution. These papers did not do as well in the final rounds of judging.

First Component: Loss of Contact

What happens after contact is lost with the plane is one major aspect a team must address in their paper. This question helps to determine the initial size, shape, and location of the search area. Did the plane immediately plummet to the ocean or did it continue to fly for a set amount of time? Did the plane make some aerial maneuvers before its descent or did some combination of scenarios occur? Teams did a good job of addressing this component of the problem, with it typically encompassing the majority of the pages in a paper. In general, modeling this component consisted of two basic approaches:

- The first set assumed that the plane began some form of an immediate descent to the ocean, as if it broke apart in flight or there was an immediate lost of power.
- The second approach assumed some sort of controlled trajectory to the ocean surface.

The judges were not looking for a specific approach but expected to see a modeling process that clearly identified and developed a set of assumptions and first principles that the team could build upon.

Many papers treated this component as a multi-stage problem. Popular first-stage modeling approaches included some type of approach to cover the descent of the plane to the ocean: Bayesian analysis, calculus-based parabolic, or physics-based falling-body. This was then followed by a Monte Carlo, simulation, or particle-distribution element to help shape the initial search area.

The most common first-stage modeling approach was the use of Bayesian analysis with some sort of probability distribution over the map of the initial wreckage area. The next-favored approach by teams was a physics approach that included a wide variety of characteristics including the weight, speed, NS air resistance of the plane, and gravity to determine the impact zone. These approaches produced an initial spherical-to-conical-shaped search area overlaid with a location probability grid. The judges considered fully-developed models with supporting data for their parameters as a critical criterion for "good" papers.

Second Component: Changing Size/Shape of the Search Area

The second component that teams considered consisted of the temporal and environmental impact on the changing shape and size of the initial search area. These considerations included developing various forms of "drifting" models that included multiple factors from prevailing ocean currents, ocean winds, and the Coriolis effect.

Unfortunately, many papers talked in generalities and presented a good deal of theoretical discussion but failed actually to integrate this discussion into a clear model with data for the change in size and shape of their initial search area. The judges considered as better papers those that actually adjusted their findings from the initial search area to account for these factors.

Once a "search region" was found, many teams broke that region into a collection of smaller $N \times N$ rectangular regions. While teams chose a specific value for N, no paper seems to have performed sensitivity analysis on the value of N, which is somewhat important because that value determines the actual size of a search area.

Third Component: Integrated Search Plan

The last component of the problem required the development of an integrated search plan to cover the developed search area. Many papers took some form of optimization approach to address this element. These approaches ranged from nonlinear programming to simulated annealing, with objective functions ranging from minimizing search time or cost to maximizing the probability of successfully finding the plane or maximizing the searched area.

The typical team approached this component using the characteristics of a single search platform. However, these teams failed to recognize that this component contains one of two required elements addressed in the problem prompt. The problem required teams to consider the impact of different types of planes with different search capabilities. The judges considered this prompt to imply that teams should consider an integrated approach to solving this problem that included coordinating the efforts of different search assets in the same plan. Teams that compared the results of different types of search platforms to determine the best platform did not suffer during the process, but the judges felt that comparing several platforms simultaneously in a single integrated search plan was a criterion for better papers.

Sensitivity, Assumptions, and Strengths and Weaknesses

The judges realize the limited time available to the teams to complete their models is a considerable constraint and they do not expect perfect models. However, the judges do expect teams to analyze their models in a structured way and to assess their models critically. A vital part of the mathematical modeling process is this critical analysis of the model.

This analysis ranges from examining the impact of the basic assumptions on the modeled conclusions to examining the shortcomings of the techniques employed in the model. As in previous years, the judging criteria placed a large emphasis on assumptions, sensitivity analysis, and testing. Many papers neglected to fully consider these issues and were scored lower by the judges.

Assumptions

The basic assumptions that a team makes are the starting point for their modeling efforts. The judges did not place restrictions on the basic assumptions other than that they need to make sense and be necessary.

However, simply listing assumptions is not enough; papers should include a discussion of why they are making an assumption and its potential impact/influence on the model.

It is also important to recognize that stating the assumption is not the end of the process; examining the impact on the modeled conclusions if the assumption changes is a vital part of the modeling process.

An example is the assumption of the type of plane that is lost. Many teams, when they developed their model for what happened after contact was lost, assumed that a particular type of plane was lost. However, few teams took advantage to address the second required element, to recognize that there are different types of planes and see how changing the type of plane impacted their analysis.

If changing an assumption results in a change of the results, then the team should indicate that as a potential weakness. The judges considered addressing the impact of different planes being lost as a criterion for "good" papers.

Sensitivity Analysis and Testing

Sensitivity analysis and testing were appropriate and necessary for all modeling approaches. Many papers included a sensitivity analysis section in their paper but only addressed the theoretical aspects of sensitivity analysis versus actually changing the value of an assumption or parameter to understand the impact. This year's problem was a perfect candidate for testing against a historical lost aircraft case study. Many teams provided a historical case study but they only used it to test the development of their initial search area and not against the actual search process. The inclusion of either a robust sensitivity analysis or a model-testing section was viewed as a critical criterion for "better papers."

Communication

Papers were judged on the quality of the writing, with special attention to the summary and to the nontechnical letter. The quality of writing, in general, is continuing to improve. The strongest summaries this year included a clear objective with a description of the search process, a general overview of the modeling process, and an explicit result of the model analysis. The judges continue to be surprised by the number of papers where the summary only describes what the team will attempt or the general theory without describing the results. Similarly, many of the nontechnical articles focused more on providing an actual press release instead of a structure for a search process.

A nontechnical article does not mean that numbers are not included. It means that the article can be read meaningfully by someone without an education in advanced mathematics.

In addition, for this year's problem, the nontechnical letter was supposed to be written as consultants to the airline executives as a guide that they could use in a press conference to describe their efforts in locating the downed aircraft. As such, the letter was supposed to summarize the team's overall approach to solving the problem. Instead, the vast majority of teams wrote this letter in the form of a press release for the airlines, which is not at all what was intended.

Conclusions

The outstanding teams modeled and presented all the aspects of the problem described in the problem statement, including the fully-developed standard elements (assumptions, sensitivity analysis, strengths and weaknesses, etc.), developed an effective model, explained the modeling choices made, and were clearly and concisely written. The judges continue to be impressed with the quality of the submissions, especially considering the time constraints. The growth in the quality and number of submissions is very encouraging to those who work to promote the practice of good mathematical modeling.

About the Author

Robert Burks is a senior lecturer in the Dept. of Defense Analysis at the Naval Postgraduate School. He received his undergraduate degree in Aerospace Engineering from the U.S. Military Academy, his master's in Operations Research from the Florida Institute of Technology, and his Ph.D. in Operations Research from the Air Force Institute of Technology. He has wide-ranging research interests, including diffusion of information, agent-based modeling, and the mathematics of 3D animation and gaming. Dr. Burks served as both a triage and final judge on this year's Lost Plane Problem.

Judges' Commentary: The Frank Giordano Award for 2015

Rich West
Francis Marion University
Florence, SC
rwest@fmarion.edu

Introduction

The Frank Giordano Award goes to a paper that demonstrates an excellent example of the modeling process.

For the fourth year in a row, the award goes to a paper for "Problem B"—the MCM Problem oriented toward methods of discrete mathematics This year, that problem was the Lost Plane Problem.

Having worked on the contest since its inception, Frank Giordano served as Contest Director for 20 years. As Frank says:

> It was my pleasure to work with talented and dedicated professionals to provide opportunities for students to realize their mathematical creativity and whet their appetites to learn additional mathematics. The enormous amount of positive feedback I have received from participants and faculty over the years indicates that the contest has made a huge impact on the lives of students and faculty, and also has had an impact on the mathematics curriculum and supporting laboratories worldwide. Thanks to all who have made this a rewarding and pleasant experience!

The Frank Giordano Award for 2015 goes to a team from the University of Colorado Boulder in Boulder, CO, for their solution to the Lost Plane Problem. This solution paper was in the top group, receiving the designation of Outstanding, and is characterized by a high-quality application of the complete modeling process, including

- assumptions with clear justifications, a well-defined, logical series of models, model testing, and sensitivity analysis of six parameters;

The UMAP Journal 36 (3) (2015) 277–280.

- originality and creativity in the modeling effort to solve the problem as given; and
- clear and concise writing, making it a pleasure to read.

The Problem

Recall the lost Malaysian flight MH370. Build a generic mathematical model that could assist "searchers" in planning a useful search for a lost plane feared to have crashed in open water such as the Atlantic, Pacific, Indian, Southern, or Arctic Ocean while flying from Point A to Point B. Assume that there are no signals from the downed plane. Your model should recognize that there are many different types of planes for which we might be searching and that there are many different types of search planes, often using different electronics or sensors. Additionally, prepare a 1–2 page nontechnical paper for the airlines to use in their press conferences concerning their plan for future searches.

The University of Colorado Paper

Executive Summary Sheet and Position Nontechnical Paper for the Airlines

The summary is well-written and gave the reader a good idea of what to expect. It contains the appropriate specifics with regard to issues and is both concise and thorough. The team's nontechnical paper, written in an appropriate nontechnical manner, summarizes the unique approach of this team for future searches in large bodies of water.

Assumptions and Parameters

After a brief introduction of their modeling process, the team states clearly and justifies specific assumptions and the parameters used. For the most part, the assumptions are reasonable and useful in addressing the model. Further, these assumptions are cross-referenced throughout the paper. The parameters are described clearly in words and individually analyzed for sensitivity later in the paper.

The Model

Considering the generic nature of a lost plane in an ocean, the team deftly approaches the problem by combining the same three aspects that most of the Finalist papers used:

- identify the crash area,
- adjust the search area for current and drift of debris, and
- allocate resources to conduct the search for debris, the plane, and the crash site.

Finally, using the parameters identified and the current and drift data from the Malaysian flight MH370 search, they developed a simulation and ran over 900 scenarios.

The approach to modeling the crash site was unique in describing in detail how, after losing power, the plane's turning would impact the worst case of a search area. The team's assumptions make it clear that they would be looking for only floating debris. But they demonstrate that they could backtrack the drift and current of the debris to find the original crash site, where they might find sunken items from the plane such as the "black boxes."

Of note regarding modeling the ocean current and drift in the search area is that the team had access only to specific data from Malaysian flight MH370. The team later declares this particularity as a weakness, with a recommendation of future work for research or development for all large bodies of water.

Finally, the allocation of planes for searching utilizes three specific patterns for a rectangular search area. Searching for debris is only visual from a plane flying one of these three patterns. The pattern of search is determined by the number of aircraft available. The simulations determine the efficiency of one pattern over another. Again, the team notes that this part is a weakness of their model, with a need for future work in investigating other patterns.

Model Testing

Model testing for the search area is done on the 2009 lost Air France Flight 447. In comparing the radii for the areas, the University of Colorado team shows how their "worst case" approach produces a much larger area for search. But of more significance for this judge are the many different simulations run with "three different search methods with varying numbers of search planes, weather conditions, visibility distances, and sizes of search areas." This versatility allows the team to make recommendations on allocations of planes and search patterns in many diverse circumstances.

Sensitivity Analysis

One particular item that makes this team's presentation stand out is the sensitivity analysis. They establish a realistic baseline with the six parameters identified initially and then vary each parameter to see its impact on the overall outcomes of the model. In each case, the results are analyzed to conclude the effect on stability of the model and recommendations for changes or future work.

Evaluation of the Model and Recommendations for Future Work

In their analysis of strengths and weaknesses and recommendations for future work, the team identifies the need for access to locally specific current data. This would be important to the airlines and / or agencies directing the search. Further, the team identifies the need for identifying take-off and landing bases for the search planes. These locations would also be specific to the search area. Finally, realizing that the size and shape of the search area will change from day to day, they recommend incorporating a parameter in the model for time elapsed.

References and Bibliography

The list of references is fairly thorough, and it is very good to see specific documentation of where those references were used in the paper.

Conclusion

The careful exposition in the development of the mathematical model, significant model testing, and extensive sensitivity analysis make this paper one that the judges felt was worthy of the Outstanding designation. The team is to be congratulated on their analysis, their clarity, and the use of the mathematics that they knew to create and justify their own model for searching for a lost plane in a large ocean.

About the Author

Rich West is a Mathematics Professor Emeritus from Francis Marion University in Florence, SC, where he taught for 12 years. Prior to his time at Francis Marion, he served in the U.S. Army for 30 years, 14 of which were spent teaching at the U.S. Military Academy. He is currently working on a National Science Foundation Grant on freshmen placement tests. He has judged for both the MCM and HiMCM for more than 12 years.

On Jargon

Big Data and Statistics Problems

William P. Fox, MCM Director
Dept. of Defense Analysis
Naval Postgraduate School
1 University Circle
Monterey, CA 93943–5000
wpfox@nps.edu

Introduction

In addition to the traditional choices of one problem each oriented toward continuous mathematics and discrete mathematics, the 2016 Mathematical Contest in Modeling (MCM)® will offer for the first time the choice of a problem featuring data insights (including "big data") and statistical analysis.

Below we give a sample problem of each kind.

Sample Big Data Problem: Let's Tweet

Background

According to IBM, we create 2.5 quintillion bytes of data every day. The creation of this data is so rapid that 90% of the data in the world today has been created in the last two years. Often these data can be messy and unpredictable. But what can these data tell us when analyzed? This is the concept of "big data."

Social media has played a huge role in the increasing data footprint, and Twitter alone accounts for 500 million posts per day. Large companies are chasing these data to find meaning, such as seeking insights into public opinion, providing social recommendations, predicting trends on

The UMAP Journal 36 (3) (2015) 281–284.

the stock market, and even triggering response to catastrophic events and emergences.

For this problem, you are been tasked with helping to improve a restaurant recommendation engine based on what people are tweeting. Think of the result as similar to the online reviews sites Yelp or a TripAdvisor, but where the reviews are tweets. It's your job to help create mathematical models that will automatically score and analyze tweets as they come in.

The Problem

An important element to the big data analysis of social media data is measuring a user's influence and credibility.

Measuring influence, especially in a specific context of what is being tweeted, can help companies weigh the importance of a tweet. Looking at credibility is equally as important, since it can determine whether a user is a trustworthy source of information. You are provided a dataset in Excel that includes 10,000 real tweets, with the following attributes:

- Number of followers of the user
- Number of retweets of the tweet
- Timestamp of tweet
- Gender of the user
- Text of the tweet
- Number of tweets that user has tweeted
- Numerical Sentiment Score of a tweet (where negative values mean negative sentiment, positive values mean positive sentiment, and 0 is neutral)

Using mathematical modeling, propose a model with rationale to score and rank the tweets, to measure the influence and credibility of each tweet in the dataset, in the following ways:

1. Number of followers is often used as an indicator of influence on Twitter. Create a scoring algorithm that gives each tweet a score on a set scale, based on number of followers.
2. Number of retweets is often used as an indicator of influence in Twitter. Now create a scoring algorithm, built on the one that you developed for number of followers, that gives each tweet a score on a set scale, based on number of followers *and* number of retweets.
3. Now take a look at the various other attributes that you have access to, such as gender and time of the tweet. How can you use these attributes to make an even better scoring algorithm? For example, does time of day affect the probability that a tweet will be retweeted? Could time of day be an indicator of predicted influence on its own?

4. One of the biggest criticisms of Twitter is the number of fake accounts and fake tweets. In fact, it's estimated that 10% of all Twitter accounts are fake. Using all the attributes you have access to, create a credibility score for each tweet. How confident are we that the tweet is from a "real" person?
5. Write a report on the rationale for your influence and credibility models you've created in 1-4, and why you think they are good indicators of a tweet's impact on Twitter.

In addition to the contest format, prepare a 1–2-page nontechnical paper to explain your methods and results to the CEO of a social media analysis company.

Data Sources

Tweet data from the Web.

Sample Statistical Analysis Problem: Causes for Violence

Background

Data are available for a country that, for each major region, include

- violence,
- poverty levels,
- literacy levels,
- government satisfaction rates from surveys,
- ethnicity, and
- employment rates.

The Problem

Can we determine the most likely cause or causes for violence in the regions and overall? Build one or more mathematical models to enable your team to answer this problem for this country.

As a result of your model:

- What steps should/could the central government take to reduce violence?

- How confident are you in the accuracy of this data, and how should that affect your decision?
- How well will your model work for any country's data on violence and causes?

In addition to the contest format, prepare a –2-page nontechnical paper to the president of this country to explain what they should do and—most importantly—why?

Data Sources to Accompany the Problem Statement

For the Philippines:

- Quantitative data on ethnicity, poverty, literacy, and good governance, mainly from the National Statistics Office and the National Statistics Coordination Board of the Philippines.
- Data on insurgency and terrorism-related activities, extracted from the archives of the Armed Forces of the Philippines and the Core Lab of the U.S. Naval Postgraduate School.

About the Author

Dr. William P. Fox is a professor in the Department of Defense Analysis at the Naval Postgraduate School and teaches a three-course sequence in mathematical modeling for decision-making. He received his B.S. degree from the United States Military Academy at West Point, New York, his M.S. at the Naval Postgraduate School, and his Ph.D. at Clemson University. Previously he has taught at the United States Military Academy and at Francis Marion University, where he was the Chair of Mathematics for eight years. He has many publications and scholarly activities including books, chapters of books, journal articles, conference presentations, and workshops. He directs several mathematical modeling contests through COMAP: HiMCM and MCM. His interests include applied mathematics, optimization (linear and nonlinear), mathematical modeling, statistical models for medical research, and computer simulations. He is President-Emeritus of the NPS Faculty Council and President of the Military Application Society of INFORMS.