

AJUR

American Journal of
Undergraduate Research

Volume 13 | Issue 4 | December 2016

www.ajuronline.org

AJUR

American Journal of
Undergraduate Research

Volume 13 | Issue 4 | December 2016

www.ajuronline.org

- 2 **AJUR History and Editorial Board**
- 3 **Special Thanks to AJUR's Sponsors**
- 5 **Green Chemistry: Comparison of Ionic Liquid vs Conventional Solvent for Indole Alkylation**
Jonathan G. Redrico & John L. Krstenansky
- 11 **Use of the Bruker AXS SMART BREEZE™ System for Macromolecular X-ray Data Collection**
Christiana Standler, G. Blake Overson, Cody A. Black, Guizella A. Rocabado, & Bruce R. Howard
- 21 **Ebola Impact and Quarantine in a Network Model**
Anca Radulescu & Joanna Herron
- 37 **Analysis of Vortex Pool-and-Chute Fishway**
Mathew Nyberg, Brian Draeger, Brian Weekly, Eileen Cashman, & Michael Love
- 59 **Dynamic Visualization of Time-Based Changes to Data Generated by Reddit.com: The Real Time Conversation Project**
Daniel Wang, Andy Luse, & Jim Burkman
- 69 **Effects similar to anxiolysis in an organic extract of *Stypopodium zonale* on an anxiety related behavior in *Drosophila melanogaster***
Keysh Mejías, Grisel Robles, Zulmari Martínez, Anamaris Torres, Lee Algarín, Genesis López, & Ricardo Chiesa
-

American Journal of Undergraduate Research (AJUR) is a national, peer-reviewed, open-source, quarterly, multidisciplinary student research journal. It is indexed internationally by EBSCO, and is listed via the Library of Congress under ISSNs of 1536-4585 (for print) and 2375-8732 (for web). The journal was established in 2002.

EDITORIAL TEAM Volume 13 | Issue 4 | December 2016

Dr. Kestutis G. Bendinskas, Editor-in-Chief, editor@ajuronline.org
 Dr. Anthony Contento, Technical Editor, technicaleditor@ajuronline.org
 Rose Throop, Print Shop Liason
 Daniel Laird, Web Master, webmaster@ajuronline.org
 Bonita Graham, LaTeX and Copy Editor

EDITORIAL BOARD by subject area

ACCOUNTING

Dr. Dean Crawford,
dean.crawford@oswego.edu

ART HISTORY

Dr. Lisa Seppi,
lisa.seppi@oswego.edu

ASTROPHYSICS

Dr. Shashi Kanbur,
shashi.kanbur@oswego.edu

BEHAVIORAL NEUROSCIENCE

Dr. Aileen M. Bailey,
ambailey@smcm.edu

BIOCHEMISTRY

Dr. Pamela K. Kerrigan,
pamela.kerrigan@mountsaintvincent.edu

BIOENGINEERING

Dr. Jorge I. Rodriguez,
jjorger@clemsun.edu

BIOINFORMATICS

Dr. Kevin Daimi,
daimikj@udmercy.edu
 Dr. John R. Jungck,
jungck@udel.edu

Dr. Isabelle Bichindaritz,
ibichind@oswego.edu

BIOLOGY, PHYSIOLOGY

Dr. David Dunn,
david.dunn@oswego.edu

BIOLOGY, DEVELOPMENTAL

Dr. Poongodi Geetha-Loganathan,
p.geethaloganathan@oswego.edu

BIOLOGY, MICROBIOLOGY

Dr. Peter Newell,
peter.newell@oswego.edu

BOTANY

Dr. William R. Bromer,
wbromer@stfrancis.edu
 Dr. Julien Bachelier,
julien.bachelier@oswego.edu

CHEMISTRY

Dr. Alfredo Castro,
castroa@felician.edu

Dr. Charles Kriley,
cekriley@gcc.edu

Dr. Douglas Mulford,
douglas.mulford@emory.edu

Dr. Mark A. Benvenuto,
benvenma@udmercy.edu

COMMUNICATION STUDIES

Dr. Jennifer Gerometta,
jgerometta@iona.edu

COMPUTER SCIENCES

Dr. Dele Oluwade,
deleoluwade@yahoo.com

Dr. Kevin Daimi,
daimikj@udmercy.edu

COMPUTATIONAL CHEMISTRY

Dr. Alexander Soudackov
asouda@illinois.edu

ECOLOGY

Dr. William R. Bromer,
wbromer@stfrancis.edu

ECONOMICS

Dr. Elizabeth Schmitt,
elizabeth.schmitt@oswego.edu

EDUCATION

Dr. Marcia Burrell,
marcia.burrell@oswego.edu

EDUCATION, PHYSICS

Dr. Andrew D. Gavrin,
agavrin@iupui.edu

ENGINEERING, ELECTRICAL

Dr. Michael Omidiora,
michael.omidiora@nyu.edu

ENGINEERING, ENVIRONMENTAL

Dr. Eileen M. Cashman,
eileen.cashman@humboldt.edu

ENGINEERING, MANUFACTURING AND CONSTRUCTION, ROBOTICS

Dr. Haoyu Wang,
wanghao@mail.ccsu.edu

ENGINEERING, SOFTWARE

Dr. Kevin Daimi,
daimikj@udmercy.edu

ENVIRONMENTAL SCIENCE

Dr. Eileen M. Cashman,
eileen.cashman@humboldt.edu

GEOLOGY

Dr. Larry Davis,
ldavis@csbsju.edu

HISTORY

Dr. Richard Weyhing,
richard.veyhing@oswego.edu

Dr. Murat Yasar,
murat.yasar@oswego.edu

HONORARY EDITORIAL BOARD MEMBER

Dr. Lorrie Clemo,
lorrie.clemo@oswego.edu

KINESIOLOGY / EXERCISE SCIENCE

Dr. David Senchina,
david.senchina@drake.edu

LITERARY STUDIES

Dr. Douglas Guerra,
douglas.guerra@oswego.edu

MATHEMATICS

Dr. John Emert,
emert@bsu.edu

Dr. Jeffrey J. Boats,
boatsjj@udmercy.edu

Dr. J.D. Phillips,
jophilli@nmu.edu

Dr. Dele Oluwade,
deleoluwade@yahoo.com

Dr. Christopher Baltus,
christopher.baltus@oswego.edu

Dr. Mark Baker,
mark.baker@oswego.edu

MEDICAL SCIENCES

Joan Newell, MD
joannewellmd@gmail.com

METEOROLOGY

Dr. Steven Skubis,
steven.skubis@oswego.edu

MUSIC

Dr. Juliet Forshaw,
juliet.forshaw@oswego.edu

NANOSCIENCE AND CHEMISTRY

Dr. Gary Baker,
bakergar@missouri.edu

NEUROSCIENCE

Dr. Pamela E. Scott-Johnson,
pamela.scottjohnson@morgan.edu

Dr. Amy Overman,
aoverman@elon.edu

PHYSICS

Dr. Carolina Ilie,
carolina.ilie@oswego.edu

POLITICAL AND SOCIAL SCIENCES

Dr. Dean Dohrman,
dean.dohrman@csuglobal.edu

PSYCHOLOGY

Dr. Amy Overman,
aoverman@elon.edu
 Dr. Pamela E. Scott-Johnson,
pamela.scottjohnson@morgan.edu

SPEECH PATHOLOGY

Dr. Kim Tillery,
kim.tillery@fredonia.edu

STATISTICS

Dr. Mark Ecker,
mark.ecker@uni.edu

TECHNOLOGY, ENGINEERING

Dr. Recayi Pecen,
regpecen@na.edu

SPECIAL THANKS

AJUR is made possible through the assistance of our sponsors.

Support for this issue has been provided by Wiley
as well as the Office of the Provost at
the State University of New York at Oswego. Thank you!

WILEY



Interested in supporting quality undergraduate research?
Request sponsorship information at editor@ajuronline.org

Green Chemistry: Comparison of Ionic Liquid vs Conventional Solvent for Indole Alkylation

Jonathan G. Redrico^a and John L. Krstenansky^{*b}

^aSchmid College of Science and Technology, Chapman University, Orange, CA

^bDepartment of Biopharmaceutical Sciences, Keck Graduate Institute School of Pharmacy, Claremont, CA

Student: jonred624@gmail.com

Mentor: John_Krstenansky@kgi.edu*

ABSTRACT

A comparison of the effectiveness of two synthetic methods was done for a green chemistry method using an ionic liquid-based solvent relative to a conventional method using organic solvents for the *N*-alkylation of indole. The green method used potassium hydroxide in the ionic liquid, 1-butyl-3-methylimidazolium tetrafluoroborate, followed by addition of the alkyl halide. The conventional method used sodium hydride in dimethylformamide. Possible advantages of the green method would be the use of a non-volatile solvent, the possibility of recycling the solvent, and the use of a less reactive base. However, its reaction yield was lower than for the conventional method, the reagent expense was higher, and recycling and reuse of the ionic solvent require considerable amounts of conventional solvent to clean it up sufficiently for reuse. On a laboratory scale, the ionic liquid method yielded neither a “green” nor an efficiency advantage. It is possible that with optimization on an industrial scale involving solvent reclamation that the green method could realize its promise.

KEYWORDS

Green Chemistry; Sustainable Chemistry; Ionic Liquids; *N*-alkylation; Indole; Solvent Recycling; Organic Chemistry; 1-Butyl-3-methylimidazolium Tetrafluoroborate; Synthesis

INTRODUCTION

The promise of green chemistry is to develop processes that offer improved effectiveness and environmental impact in terms of reducing hazardous waste, reducing workplace hazards, using sustainable materials, or reducing energy requirements. One development in the field of green chemistry involves utilizing ionic liquids in place of conventional solvents.¹ Ionic liquids are salts with low melting points, where one or both of the ions are large and the cation has a low degree of symmetry that results in a reduced lattice energy for the crystalline form of the salt. These liquids exhibit very low vapor pressures and thermal stabilities, and can often be recovered and reused as solvents. Ionic liquids allow for a triphasic system with aqueous and hydrocarbon phases, adding further to their utility as solvents.

These potential advantages of ionic liquids were explored as an alternative solvent in the *N*-alkylation of indole as part of a larger combinatorial chemistry effort. The low vapor pressure of ionic liquids allows for use in combinatorial setups that are open to the atmosphere. The conventional method treats indole with sodium hydride (NaH) in dimethylformamide (DMF).²

In the conventional reaction, NaH is used as a strong base to create the indole *N*-anion. NaH is commercially available as a 60% dispersion in mineral oil. The mineral oil is removed by trituration of the NaH with hexane. Besides the handling issues in combinatorial synthesis, NaH readily reacts with moisture in the air or in solvents, therefore requiring closed reaction systems and anhydrous solvents. Upon reaction with indole, flammable hydrogen gas is produced. Workup involves quenching in water, extracting with ether, and eluting on a short silica column. No solvents are recycled in the process. Prior experience in our lab with the conventional method² shows that the crude product obtained is usually about 90% by weight product with the remainder being primarily solvents as determined by gas chromatography-mass spectrometry (GC-MS) and nuclear magnetic resonance spectroscopy (NMR). This material is sufficiently pure to use for subsequent acylation reactions.

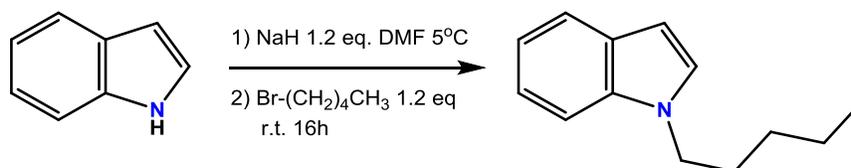
In contrast, the green method in the ionic liquid, 1-butyl-3-methylimidazolium tetrafluoroborate ([BMIM]⁺[BF₄]⁻), uses potassium hydroxide (KOH) as the base in place of NaH.³ Although KOH is hygroscopic, in a dry atmosphere it is easily handled without special apparatus. This feature, along with the low vapor pressure of the ionic liquid, allows for the use of apparatus that is not as rigorously dry and shut off from the atmosphere. Also, there is no generation of hydrogen and the solvent is not flammable, so fire risk is eliminated.

An additional advantage to the ionic liquid is the opportunity for re-purification⁴ and reuse, which offers the possibility of producing less hazardous solvent waste. This recoverability is also important from a cost standpoint since the ionic liquid is significantly more expensive than the organic solvent of the conventional method. Therefore, the purpose of this study is to examine the utility of this alternative green method for the production of *N*-alkylindole intermediates in a combinatorial synthetic scheme involving subsequent acylation of the intermediates.

METHODS AND PROCEDURES

All reagents and solvents were purchased from Sigma-Aldrich, the solvents were HPLC grade, and all were used without further purification.

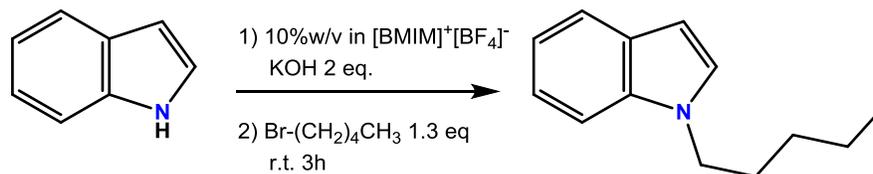
Conventional *N*-alkylation of Indole



Scheme 1.

Commercially available 1H-indole (11.8 g, 101 mmol) in 90 mL anhydrous *N,N*-dimethylformamide (DMF) was added to a solution of sodium hydride (NaH) (4.90 g, 122.4 mmol, 60% dispersion in mineral oil, which was removed by triple trituration with hexane) in 70 mL DMF, at 5°C. The mixture was allowed to warm to ambient temperature and was stirred for 30 min. The mixture was cooled to 5°C and 1-bromo-*n*-pentane (14.9 mL, 120 mmol) in 50 mL DMF was added drop-wise to the solution. After addition was complete, the mixture was allowed to warm to ambient temperature. After 16 h, the mixture was cooled to 5°C and quenched with water, then extracted with 4 × 200 mL diethyl ether (Et₂O), dried over anhydrous sodium sulfate (Na₂SO₄), and filtered. The combined organic fractions were concentrated *in vacuo*. Silica gel chromatography (30 mm high × 85 mm diam; 250 mL *n*-hexane followed by 2 × 250-mL portions of dichloromethane (DCM)) yielded 17.7 g of a yellow oil that by NMR consisted of 16.1 g 1-*n*-pentyl-1H-indole (85.2% yield) with 4.8% DMF, 1.9% water, 0.7% DCM, and 1.3% ether.

Green-Chemistry *N*-alkylation of Indole Using Ionic Liquid

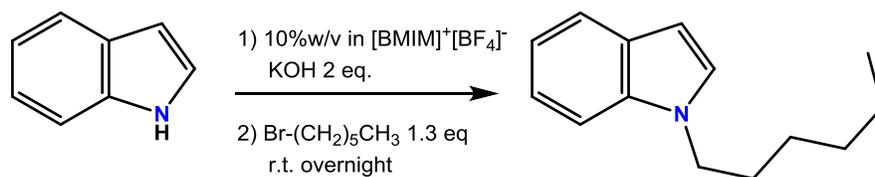


Scheme 2.

1H-indole (4.98 g, 42.5 mmol) at 10% w/v [BMIM]⁺[BF₄]⁻ was stirred with crushed potassium hydroxide (KOH) pellets (2.00 equiv., 4.83 g, 86.2 mmol) and 1-bromo-*n*-pentane (1.30 equiv., 6.85 mL, 55.2 mmol) at room temperature. After 3 h, the mixture was filtered through glass wool to remove undissolved KOH and potassium bromide (KBr), and the filtrate was extracted with 3 × 50 mL Et₂O. The combined organic fractions were dried over Na₂SO₄, filtered, and concentrated *in vacuo*. Silica gel chromatography (30 mm high × 20 mm diam.; 40 mL hexane followed by 11 × 30-mL portions of DCM) was used to separate the reaction products, yielding 1-*n*-pentyl-1H-indole (4.68 g, 58.5% yield) as a colorless oil that contained 7.4% 1-bromo-*n*-pentane and 0.4% water as indicated by NMR.

Purification of the Ionic Liquid

M.J. Earle, et al. reported that after diethyl ether extraction of the reaction products, the ionic liquid only required filtration to be suitable for reuse—as determined by NMR—and that there were only small mechanical losses from process to process.³ Despite this literature precedent, we were concerned about the possibility of batch-to batch-contamination. Therefore, the extra step of ionic liquid [BMIM]⁺[BF₄]⁻ purification before reuse was employed using the method of L.L. Lazarus, et al.⁴ The ionic liquid was diluted with 120 mL DCM and eluted on a column of 1 cm Celite topped with 1 cm silica gel and then with 3 cm Norit. An additional 50 mL DCM was used to completely elute the material. After evaporation *in vacuo*, 67% of the ionic liquid used in the prior reaction was recovered.

Green-Chemistry *N*-alkylation Using Recycled Ionic Liquid, Compared with Using Fresh Ionic Liquid

Scheme 3.

1H-indole (3.58 g, 30.6 mmol) at 10% w/v recycled [BMIM]⁺[BF₄]⁻ was stirred with crushed KOH pellets (2.0 equiv., 3.43 g, 61.2 mmol) and 1-bromo-*n*-hexane (1.3 equiv., 7.36 mL, 59.4 mmol) at room temperature. After stirring overnight, the mixture was filtered through glass wool to remove undissolved KOH and KBr, and the filtrate was extracted with 3 × 50 mL Et₂O. The combined organic fractions were dried over Na₂SO₄ and concentrated *in vacuo*. Silica gel chromatography (30 mm high × 20 mm diam; 40 mL hexane followed by 11 × 30-mL portions of DCM) gave 7.52 g of a yellow oil that, by NMR and GC-MS analysis, consisted of 2.98 g 1-*n*-hexyl-1H-indole (48.4% yield), 0.228 g residual 1-*n*-pentyl-1H-indole carried over from the previous reaction, and unreacted 1-bromohexane as the remainder.

1H-indole (4.98 g, 42.5 mmol) at 10% w/v fresh [BMIM]⁺[BF₄]⁻ was stirred with crushed KOH pellets (2.0 equiv., 4.77 g, 83.0 mmol) and 1-bromo-*n*-hexane (1.30 equiv., 6.86 mL, 55.3 mmol) at room temperature. After stirring overnight, the mixture was filtered through glass wool to remove undissolved KOH and KBr, and the filtrate was extracted with 3 × 50 mL Et₂O. The combined organic fractions were dried over Na₂SO₄ and concentrated *in vacuo*. Silica gel chromatography (30 mm high × 20 mm diam; 40 mL hexane followed by 11 × 30-mL portions of DCM) gave 9.92 g of a yellow oil that, by NMR and GC-MS analysis, consisted of 2.13 g 1-*n*-hexyl-1H-indole (24.9% yield) and unreacted 1-bromohexane as the remainder.

Gas Chromatography – Mass Spectrometry (GC-MS)

Samples of the 1-hexyl-1H-indole products in 1 mL *n*-hexane were prepared for GC-MS analysis using a Hewlett-Packard 6890 Series GC System with Mass Selective Detector and Auto-Injector. These were run on HP-5 5% phenyl methyl silane columns (30.0 m × 0.250 mm). The flow rate of helium was at 1.1 mL/min. Splitless injections of 2 μL were run at an injection temperature of 250°C. The temperature gradient was 70°C for 3 min., then increased to 290°C at a rate of 10° C/min., and held at 290°C for 10 min. The total flow rate was 23.5 mL/min.

Nuclear Magnetic Resonance (NMR)

NMR spectra were recorded on a Varian 400MR spectrometer (proton frequency 399.765 MHz) equipped with an AutoX DB broadband probe. Pulse sequences, acquisition, and data processing were accomplished using Vnmrj software (Vnmrj 4.2, Agilent Technologies, Santa Clara, CA). The spectrometer was locked on to D₂O and spectra were acquired at 28°C without spinning. Water suppression suitability studies were carried out using the presaturation (presat), WET (WET1D), and excitation sculpting (water_ES) pulse sequences (Vnmrj 4.2, Agilent Technologies, Santa Clara, CA) with automatic suppression of the tallest peak (water at δ 4.86 ppm), an observation pulse of 90° (10.8 μs), a spectral width of 6410.3 Hz, a relaxation time of 30 s, and an acquisition time of 5.112 s. Eight scans were taken. Quantitation spectra were recorded using the WET pulse sequence (wet 1D) (Vnmrj 4.2, Agilent Technologies, Santa Clara, CA) with automatic suppression of the tallest peak (water at δ 4.86 ppm) with an observation pulse of 90° (10.8 μs), a spectral width of 6410.3 Hz, a relaxation time of 30 s, and an acquisition time of 5.112 s. Waltz decoupling was used during the acquisition time to suppress ¹³C side-bands. The gain was set to 24 dB and 64 scans. Three replicates were taken for each sample.

RESULTS AND DISCUSSION

Following a conventional synthetic procedure using NaH with DMF, 1-*n*-pentylindole was produced at an 85.2% yield, which is consistent with literature.² The purity level was 90% by weight with the impurities consisting of solvents, which produces acceptable results in subsequent acylation reactions. Following the green synthetic procedure using KOH with [BMIM]⁺[BF₄]⁻, 4.31 g of product was produced at a 54.2% yield. This is significantly lower than similar procedures in literature that have reported yields of around 90%.³ To improve the yield we increased the reaction time from 3 hours to overnight (approx. 16 hours) for the reactions involving 1-*n*-bromohexane; however, subsequent yields were even lower. These lower yields are likely due to factors other than the chain length of the alkyl halide. It seems more likely that either the starting material was insufficiently pure or another contaminant was present that interfered with the reaction. As such, the effect of lengthening the reaction time cannot be determined; however, it is unlikely to be detrimental since the product is stable.

The green method has significantly higher cost due to the price of the ionic liquid. The ionic liquid used for this procedure was 1-butyl-3-methyl-imidazolium tetrafluoroborate ([BMIM]⁺[BF₄]⁻), which cost us \$172/100g. The cost of DMF needed for the same scale reaction was approximately \$12. There is no cost advantage to reusing the ionic liquid at the scales we are exploring since organic solvents are used in the recycling process. This need for purification and additional solvent use runs counter to the environmental ideal of reduced solvent use, as these wash solvents are not recovered at the lab scale. In an industrial process, recycling of these organic solvents may produce some cost efficiencies that are not realized on a lab scale.

Other factors involved in reusing the ionic liquid are its recovery yield and quality after purification. First, our ionic liquid recovery was only 67%. While this recovery may be optimized, it is important to consider these losses during recovery when comparing the differences of the green-chemistry method over the conventional method. Second, performing an alkylation with bromohexane in purified ionic liquid that was used in a previous alkylation with bromopentane gave a product that contained both hexyl and pentyl products in a 13:1 ratio, by GC-MS analysis. We did not determine whether this was due to pentyl-indole product being carried over in the ionic liquid, or whether unreacted bromopentane was left over in the ionic liquid after purification, which would compete with the bromohexane to create the mixture of alkylated products. In either case, the purification scheme was insufficient to ensure the removal of impurities from prior reactions. Therefore, the current purification method would only allow the reuse of the ionic liquid for running the identical reaction; otherwise, the purification method would need significant improvement to ensure complete removal of impurities.

CONCLUSION

In our hands, the conventional synthesis produced higher yields, less waste, and more easily purified material than did the green process using ionic liquids. Our limited trials using the ionic liquid method did not produce the yields expected from the prior literature.³ The ionic liquid method also did not perform as anticipated in terms in its ability to recover, purify, and reuse the ionic liquid. The literature suggested that ether extraction of the product gave an ionic liquid that only required filtration before reuse.³ We experienced significant carryover from a prior reaction of either product or the bromoalkane reagent. Purification of the ionic liquid⁴ required the use of conventional solvents and produced solid waste; hence, the process did not meet the objectives of reducing waste. Purification also resulted in the loss of some of the ionic liquid and was still not adequate to remove impurities that would affect the next reaction for which it was used. Besides not delivering on the expectations of comparable yields and reduced solvent use, the green process was significantly more expensive and required more time due to the extra step of solvent recovery and reprocessing. Finally, while the ionic liquid is less volatile than conventional solvents, and therefore has the possibility of presenting less inhalation hazard risk, this does not mean that the ionic liquid is less toxic for humans or the environment.^{5,6} Based on all these considerations, for our purposes in efficiently creating laboratory scale precursors for the combinatorial synthesis of biologically active materials for structure-activity studies, the conventional synthesis ultimately produces less toxic waste and makes more efficient use of materials than the ionic liquid method and in that sense is more environmentally friendly of two options.

ACKNOWLEDGEMENTS

The authors acknowledge support from The Kenneth T. and Eileen L. Norris Foundation for the Summer Undergraduate Research Experience (SURE) Program and acknowledge Dr. Anastasia Levitin for her supervision of the program at the Keck Graduate Institute. We also thank Darlene Dang for the assistance she has provided in the laboratory. Finally, we thank the Department of Chemistry at Pomona College for the use of their GC-MS and the Keck Graduate Institute School of Pharmacy for supplying the resources and funding necessary to successfully carry out this project.

REFERENCES

1. Blaazer, A.R., Lange, J.H.M., van der Neut, M.A.W., Mulder, A., den Boon, F.S., Werkman, T.R., Kruse, C.G., Wadman, W.J. (2011) Novel indole and azaindole (pyrrolopyridine) cannabinoid (CB) receptor agonists: Design, synthesis, structure-activity relationships, physicochemical properties and biological activity, *Eur J Med Chem* 46, 5086-5098.
2. Earle, M.J., and Seddon, K.R. (2000) Ionic liquids. Green solvents for the future, *Pure Appl Chem* 72, 1391-1398.
3. Earle, M.J., McCormac, P.B., and Seddon, K.R. (1998) Regioselective alkylation in ionic liquids, *Chem Commun* 2245-2246.
4. Lazarus, L.L., Riche, C.T., Malmstadt, N.J., and Brutchey, R.L. (2012) Effect of ionic liquid impurities on the synthesis of silver nanoparticles, *Langmuir* 28, 15987-15993.
5. Romero, A., Santos, A., Tojo, J., and Rodriguez, A. (2008) Toxicity and biodegradability of imidazolium ionic liquids, *J Hazardous Materials* 151, 268-273.
6. Bernot, R.J., Brueske, M.A., Evans-White, M.A., and Lamberti, G.G. (2005) Acute and chronic toxicity of imidazolium-based ionic liquids on *Daphnia magna*, *Environ Toxicol Chem* 24, 87-92.

ABOUT THE STUDENT AUTHOR

Jonathan Redrico worked on this project as part of the 2014 Summer Undergraduate Research Experience program at Keck Graduate Institute. He graduated with a Bachelor of Science degree in Biochemistry and Molecular Biology from Chapman University in May 2015. He is currently pursuing an industrial career in biotechnology.

PRESS SUMMARY

Green chemistry is the use of chemical methods that reduce environmental impact in terms of reducing hazardous waste, reducing workplace hazard, utilizing sustainable materials, or reducing energy requirements. One such approach is to replace conventional solvents with ionic liquids. We examined the indole alkylation reaction routinely done in our labs. In this particular case, we found that at the laboratory scale, the conventional method was more effective and generated less toxic waste than the ionic liquid method.

Use of the Bruker AXS SMART BREEZE™ System for Macromolecular X-ray Data Collection

*Christiana Standler, G. Blake Overson, Cody A. Black, Guizella A. Rocabado, and Bruce R. Howard**

Department of Physical Science, Southern Utah University, Cedar City, UT

Students: christystandler@yahoo.com, goverson4@gmail.com, codyadamblack@gmail.com, guizella.rocabado@gmail.com

*Mentor: howard@suu.edu**

ABSTRACT

The Bruker AXS SMART BREEZE™ system is a single-crystal X-ray diffractometer designed to collect data from crystals of small organic or inorganic compounds. It is typically equipped with a Molybdenum-anode sealed tube to facilitate data collection from small unit cells. We recently acquired this system, but chose to have it installed with a copper-anode sealed tube with the hope of using it to collect data from larger unit cells such as those found in crystals of proteins or other macromolecules. This is the first and only BREEZE™ system installed by Bruker AXS with a copper-anode to date. Here we show that this system is capable of efficiently collecting quality X-ray diffraction data from crystals of the enzymes lysozyme and xylanase. This capability to collect diffraction data from both macromolecular and small-molecule crystals greatly expands the scope of undergraduate research projects that can be addressed using this instrument.

KEYWORDS

X-ray; Diffraction; Crystallography; Diffractometer; Protein; Enzyme; Crystal; Structure

INTRODUCTION

X-rays were discovered in 1895 by Wilhelm Roentgen at Wurzburg University in Germany, resulting in his winning the Nobel Prize in Physics in 1901.¹ The first diffraction of X-rays from crystals was reported in 1912 by Max von Laue,² and a description of the diffraction as reflections from internal planes of electron density was provided soon after by Sir William Lawrence Bragg.³ These contributions resulted in Nobel Prizes in Physics for von Laue and Bragg in 1914 and 1915 respectively.¹ The first crystal structure reported was for sodium chloride,⁴ and since then this technique has become one of the most effective methods for defining the three-dimensional atomic structures of molecules and coordination compounds.

X-ray crystallography involves the collection of diffracted rays and analysis of the resulting patterns to generate an electron density map of the molecules in the crystal. X-ray crystallography allows for the imaging of particles as small as the wavelength of the X-rays used, which is often about one to two Angstroms (Å). The low-resolution data are found at low scattering angles, while the higher resolution data diffract to higher angles. This data must be analyzed, and phase information for the diffracted rays must be determined. This is often provided by the analysis of the influence of heavy atoms within the crystal. For excellent discussions of these processes, the reader is referred to the work of Rhodes and Rupp.^{5,6}

Molecules must be in a crystalline structure for X-rays to be successfully diffracted and molecular models to be determined. This is a challenge in the case of protein molecules because many are difficult to crystallize, as they do not naturally occur in crystalline form. Protein crystals are also fragile because they require a high level of hydration (typically 40%–60% water by volume) in order to maintain a biologically relevant structure in the crystal. Despite these challenges, X-ray crystallography is arguably the most common method used to determine three dimensional models of protein molecules.⁷ Of the 123,021 structures deposited in the Protein Data Bank as of September 2016, 109,963 (89.4%) were solved with X-ray crystallography.

The first X-ray crystal structure of a protein was reported for myoglobin^{8,9}, which introduced the scientific community to the complexities of protein structure and indicated that proteins could adopt specific three-dimensional folds. It also provided insight into how a specific conformation of protein could drive function. The second protein structure solved was hemoglobin, which showed how proteins can work cooperatively to carry out their biological function.^{10,11}

The Physical Science Department at Southern Utah University (SUU) recently purchased a Bruker AXS SMART BREEZE™ System (**Figure 1**). It is comprised of a stand-alone cabinet with leaded glass windows enclosing a sealed fine-focus X-ray tube which is water-cooled. The crystal is mounted on a goniometer head with a fixed-chi geometry. Data are collected on a BREEZE™ CCD detector with an active area of 62 mm x 62 mm, fitted with a Peltier thermoelectric cooling chip maintained at -25 °C to reduce background noise. The system is equipped with an Oxford Cryostream 700 in order to control sample temperature and to allow data collection at cryogenic temperatures (ranges from 80 to 400 Kelvin).

Although this system is designed to collect data from crystals with small unit cells, we wanted to evaluate its effectiveness for use with larger unit cells such as those found in protein crystals. When our system was installed, we selected a copper-anode sealed tube for an X-ray source. This is the first and only SMART BREEZE™ system installed by Bruker AXS with a copper tube.¹² We wanted to use copper-K α radiation because we hoped to be able to use the system for protein crystals in addition to small molecules. To confirm this expectation, we chose hen egg white lysozyme to carry out our crystallization and diffraction experiments. Hen egg white lysozyme was first isolated and crystallized in 1944.¹³ It was the first enzyme to have its structure determined,¹⁴ although the structure wasn't refined until nearly a decade later.¹⁵ Because details of this protein have been studied and very well defined, including crystallization conditions, lysozyme provided a good target to carry out our evaluation experiments. We also tested another available protein standard: xylanase from *Trichoderma longibrachiatum*.^{16, 17}

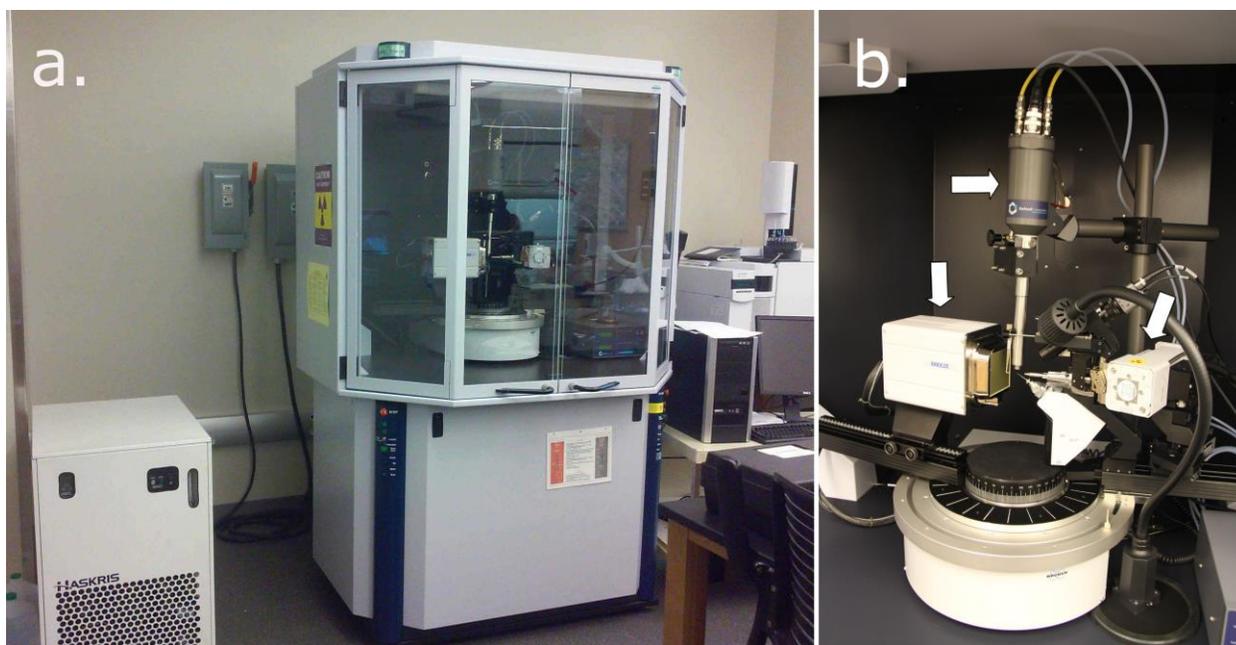


Figure 1. a. SMART BREEZE™ System in the lab with Haskris water-chiller on the left to cool the X-ray tube. b. Close-up view inside of the cabinet. The BREEZE™ CCD detector is on the left, the sealed X-ray tube on the right, and the Cryostream at the top (white arrows).

MATERIALS, METHODS AND RESULTS

Crystallization

Hen egg white (HEW) lysozyme was purchased from Sigma-Aldrich in the form of a lyophilized powder and resuspended in 0.1 M sodium acetate, pH 4.8, at a final concentration of 75 mg/ml. Crystallization was carried out in sitting drops using a reservoir solution consisting of 10% w/v NaCl, 0.1 M sodium acetate pH 4.8, 25% v/v ethylene glycol.^{18, 19} For each drop, 7 μ l of protein solution and 3 μ l of reservoir solution were mixed together and allowed to equilibrate at room temperature with 1 ml of reservoir solution via vapor diffusion (**Figure 2**).

Xylanase was purchased from Hampton Research as a concentrated solution containing the enzyme at 36 mg/ml, 0.18 M Na/K phosphate pH 7, and 43% w/v glycerol. Crystallization was carried out in sitting drops using a reservoir solution containing 1.3 M ammonium sulfate, 200 mM bicine pH 9. The protein solution contained 18 mg/ml xylanase, 21.5% glycerol, 90 mM Na/K phosphate pH 7. For each drop, 2 μ l of protein solution was added to 1.4 μ l of reservoir solution and 0.6 μ l of 1 M NaI, and allowed to equilibrate at room temperature in vapor contact with 1 ml of reservoir solution (**Figure 3**).¹⁷



Figure 2. Tetragonal crystals of Hen egg white lysozyme.

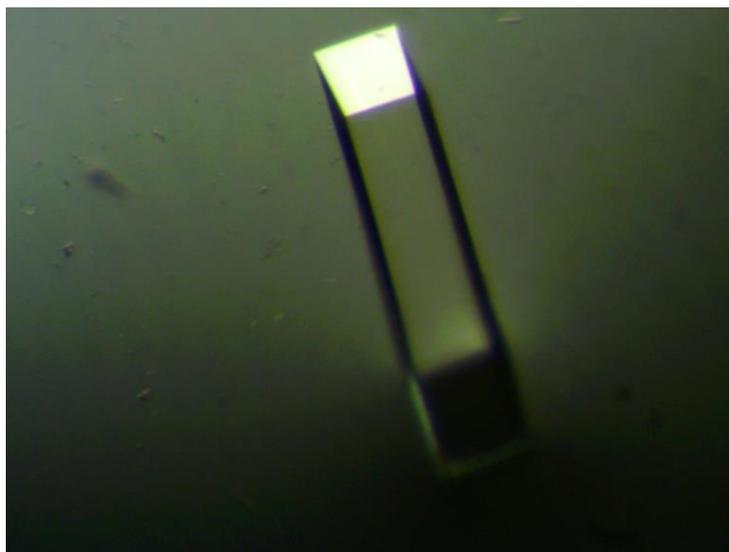


Figure 3. Orthorhombic crystal of Xylanase used for data collection.

X-ray data collection and processing

Crystals of lysozyme were suspended in nylon loops and cooled quickly by placement in a stream of dried nitrogen gas chilled to a temperature of 100 Kelvin (Oxford Cryostream 700, Oxford Cryosystems). X-rays were generated by a 1.5 kW fine-focus copper sealed tube running at 40 kV and 30 mA (Siemens, Germany). A collimated, monochromatic, 1.54-Å X-ray beam 300 μm in diameter was directed through the crystal and into a lead beamstop. Diffraction data were collected on a BREEZE™ CCD detector (Bruker AXS).

An example of the diffraction data from a lysozyme crystal can be seen in **Figure 4**. A total of 2709 images were collected by oscillating the crystal in 0.5° steps with 60-second exposures to generate a highly redundant (~40-fold overall) dataset. The details of the data collection are listed in **Table 1a**. The diffraction data were indexed, integrated, scaled and merged using SAINT²⁰ and SADABS.²¹ The X-ray data are listed in **Table 2a**.

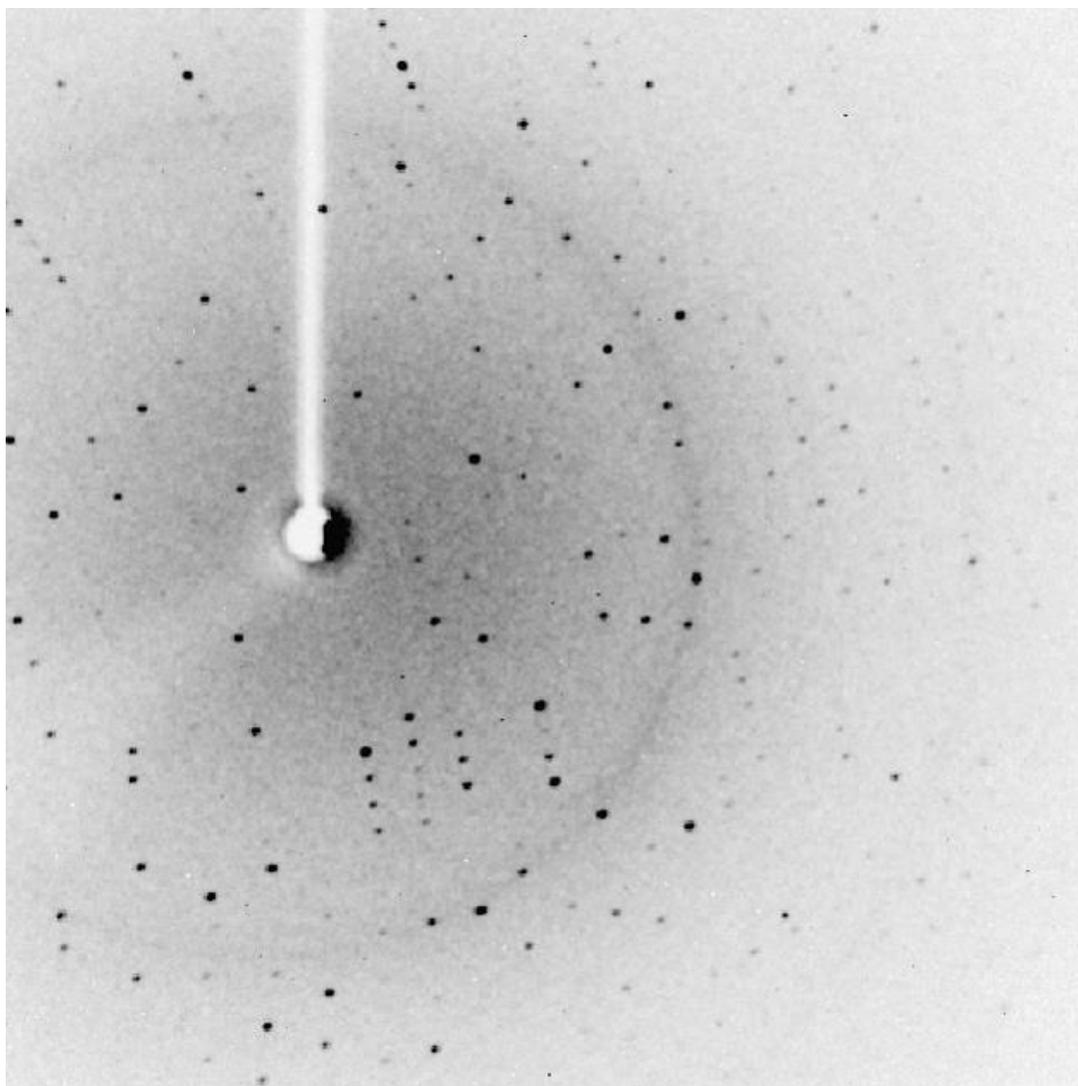


Figure 4. Example of diffraction data from a lysozyme crystal measuring approximately 0.2 mm × 0.2 mm × 0.1 mm in size. The image is a 60 second exposure with an oscillation angle of 0.5°. The right edge of the detector corresponds to a d-spacing of ~1.8 Angstroms.

Table 1a. Lysozyme X-ray Data Collection									
Axis	dist/mm	2θ/°	ω/°	φ/°	χ/°	Width (°)	Sweep (°)	Frames	Time (s)
Omega	50.386	-14.5	-199	257.5	54.7	0.5	188.5	377	60
Omega	50.386	18	-166.5	142.5	54.7	0.5	188.5	377	60
Omega	50.386	-17	-201.5	-100	54.7	0.5	61	122	60
Omega	50.386	18	-166.5	-112.5	54.7	0.5	165.5	331	60
Omega	50.386	-19.5	-202.05	-317.5	54.7	0.5	185.5	371	60
Omega	50.386	-14.5	-199	45	54.7	0.5	188.5	377	60
Omega	50.386	13	-171.5	-27.5	54.7	0.5	188.5	377	60
Omega	50.386	18	-166.5	-347.5	54.7	0.5	188.5	377	60

Table 2a. Lysozyme X-ray Data Statistics				
Resolution (Å)	% Complete	Redundancy	Mean I/ σ	R _(int)
19.75 - 5.11	97.3	78.51	80.48	0.0601
5.11 - 4.00	100	64.11	76.49	0.0581
4.00 - 3.47	100	59.1	67.27	0.0683
3.47 - 3.14	100	57.24	58.03	0.0764
3.14 - 2.90	100	52.84	41.48	0.1051
2.90 - 2.72	100	49.02	29.38	0.1402
2.72 - 2.58	100	46.73	25.49	0.163
2.58 - 2.46	100	44.33	22.33	0.1853
2.46 - 2.36	100	42.58	19.4	0.2122
2.36 - 2.27	100	40.8	18.45	0.2236
2.27 - 2.19	100	39.24	15.02	0.2646
2.19 - 2.12	100	37.42	14.24	0.2878
2.12 - 2.06	100	35	12.07	0.3251
2.06 - 2.01	100	31.81	10.12	0.3791
2.01 - 1.96	100	26.89	7.87	0.4374
1.96 - 1.92	100	23.64	6.15	0.4698
1.92 - 1.88	100	18.89	5.2	0.5097
1.88 - 1.84	100	12.53	3.58	0.5991
1.84 - 1.80	99.6	8.07	2.62	0.6567
19.75 - 1.80	99.8	40.27	26.62	0.128

A xylanase crystal (**Figure 3**) measuring approximately 0.15 mm \times 0.15 mm \times 0.85 mm was transferred to a cryoprotected solution consisting of 1.3 M ammonium sulfate, 200 mM bicine pH 9, 30% v/v glycerol and 175 mM NaI for a few seconds and then flash-cooled by plunging into liquid nitrogen. This crystal was held at 100 K for data collection as described above. For this crystal we wanted to test the efficiency of the system by collecting shorter exposures and limiting the data collection time to five hours (**Table 1b**). A shortened data collection time would allow the collection of multiple data sets per day, and would support the feasibility of providing a data collection opportunity to each student group in an academic laboratory. Such efficient data collection on a protein standard like xylanase would allow the incorporation of a crystallography section into an academic biochemistry laboratory. As shown in **Table 2b**, we were able to collect a complete dataset to 1.78 Å resolution with an overall redundancy of 13.5, an \sim 13-fold mean signal to noise ratio, and reasonable integration statistics.

Table 1b. Xylanase X-ray Data collection

Axis	dist/mm	2θ/°	ω/°	φ/°	χ/°	Width (°)	Sweep (°)	Frames	Time (s)
Omega	50.270	25.31	-158.69	28.00	54.73	0.50	188.00	376	10
Omega	50.270	25.31	-158.69	-23.18	54.73	0.50	188.00	376	10
Phi	50.270	24.21	-159.79	-127.00	54.73	0.50	150.00	300	10
Omega	50.270	25.31	-158.69	102.96	54.73	0.50	188.00	376	10
Omega	50.270	25.31	-158.69	-195.63	54.73	0.50	188.00	376	10

Table 2b. Xylanase X-ray Data statistics

Resolution	% Complete	Redundancy	Mean I/σ	R _(int)
58.78 - 7.74	99.3	18.09	41.28	0.0660
7.74 - 5.03	99.9	17.56	32.92	0.0776
5.03 - 3.94	100.0	17.27	37.21	0.0608
3.94 - 3.42	100.0	17.25	33.77	0.0717
3.42 - 3.09	100.0	17.19	26.45	0.0940
3.09 - 2.86	100.0	16.44	20.68	0.1212
2.86 - 2.68	100.0	15.65	15.89	0.1528
2.68 - 2.55	100.0	14.93	13.16	0.1796
2.55 - 2.43	100.0	14.28	11.09	0.2064
2.43 - 2.34	100.0	13.66	9.55	0.2361
2.34 - 2.25	100.0	13.10	8.46	0.2615
2.25 - 2.18	100.0	12.77	7.41	0.2950
2.18 - 2.11	100.0	12.27	6.88	0.3068
2.11 - 2.06	100.0	11.90	5.68	0.3471
2.06 - 2.01	100.0	11.58	5.40	0.3585
2.01 - 1.96	100.0	11.26	4.35	0.4224
1.96 - 1.92	100.0	11.01	3.79	0.4743
1.92 - 1.88	100.0	10.67	3.08	0.5382
1.88 - 1.84	100.0	10.41	2.55	0.5948
1.84 - 1.81	100.0	10.21	2.27	0.6356
1.81 - 1.78	100.0	9.96	1.96	0.6833
58.78 - 1.78	100.0	13.50	12.81	0.1500

Crystal structure solution and refinement of lysozyme

The unit cell parameters were similar to a tetragonal crystal form of HEW lysozyme with the same space group that was deposited in 1975 by R. Diamond (PDB ID: 1LYZ).⁷ The space group for this entry is P4₃2₁2 with cell dimensions of a = b = 79.10 Å, c = 37.90 Å. We downloaded the atomic coordinates for this structure and removed the solvent molecules. This dehydrated structural model was used to find a molecular replacement solution using the program Phaser in the CCP4 suite.^{22, 23} The model was built, adjusted and solvent molecules placed using Coot.²⁴ Refinement was carried out with REFMAC5.²⁵ The atomic model and electron density maps displayed in **Figure 5** were made with PyMol.²⁶

Space Group	P4 ₃ 2 ₁ 2
Cell dimensions (Å)	a = b = 78.98, c=36.85
Resolution (Å)	20 - 1.80
R _{work}	0.1891
R _{free}	0.2368
rms bond lengths (Å)	0.023
rms bond angles (°)	1.77
$R_{\text{factor}} = \frac{\sum_{hkl} F_o - F_c }{\sum_{hkl} F_o }$ R _{free} is the R _{factor} for 4.8% of the data not used in refinement. R _{work} is the R _{factor} for 95.2% of the data used in refinement.	

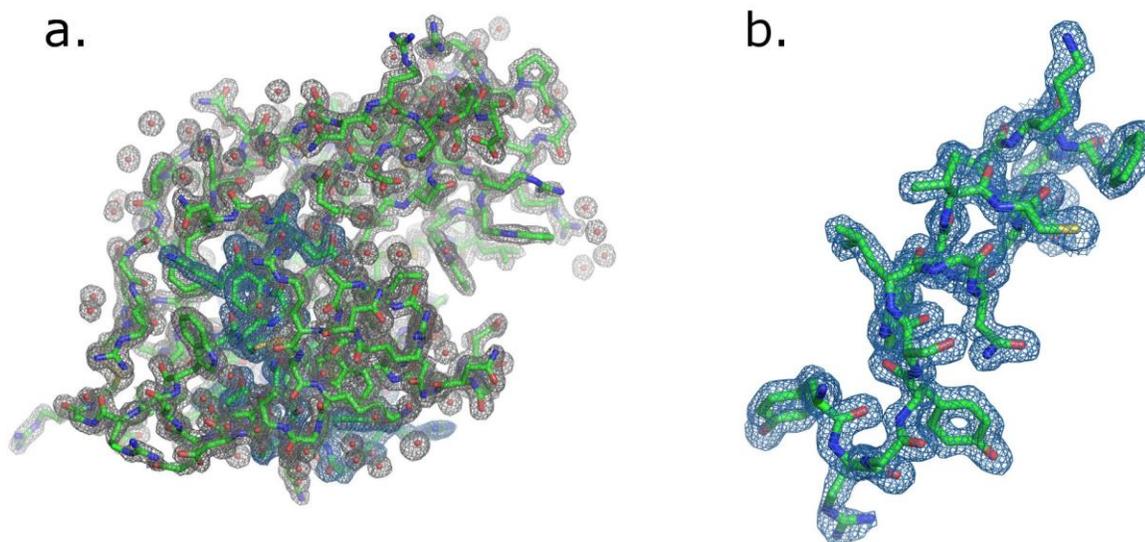


Figure 5. 2Fo-Fc electron density map contoured at 1 σ is superimposed on the refined atomic model of lysozyme in stick form with carbon atoms green, nitrogen blue, sulfur yellow, and oxygen red. **a.** Entire lysozyme molecule including ordered water molecules modeled as single oxygen atoms due to the data resolution. **b.** An α -helix from the model shown individually to better illustrate the quality of the map.

Crystal structure solution and refinement of xylanase

The unit cell parameters of the xylanase crystal were similar to an orthorhombic form of this enzyme in the PDB (ID 4HKL).⁷ The space group for this entry is P2₁2₁2₁ with cell dimensions of a = 48.3, b = 59.0 Å, c = 69.7 Å. The atomic coordinates for this structure with the solvent molecules removed was used as a search model for molecular replacement using Phaser in the CCP4 suite.^{22, 23} The model was adjusted, side chain rotamers were checked and fitted into the electron density map, and solvent molecules were added using Coot.²⁴ Refinement was carried out with REFMAC5.²⁵ The atomic model and electron density maps displayed in **Figure 6** were made with PyMol.²⁶

Table 3b. Xylanase Structure Refinement	
Space Group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions (Å)	a = 48.34, b = 58.78, c = 69.70
Resolution (Å)	45 - 1.78
R _{work}	0.1584
R _{free}	0.1971
rms bond lengths (Å)	0.020
rms bond angles (°)	1.79
$R_{\text{factor}} = \frac{\sum_{\text{hkl}} F_o - F_c }{\sum_{\text{hkl}} F_o }$ R _{free} is the R _{factor} for 5.1% of the data not used in refinement. R _{work} is the R _{factor} for 94.9% of the data used in refinement.	

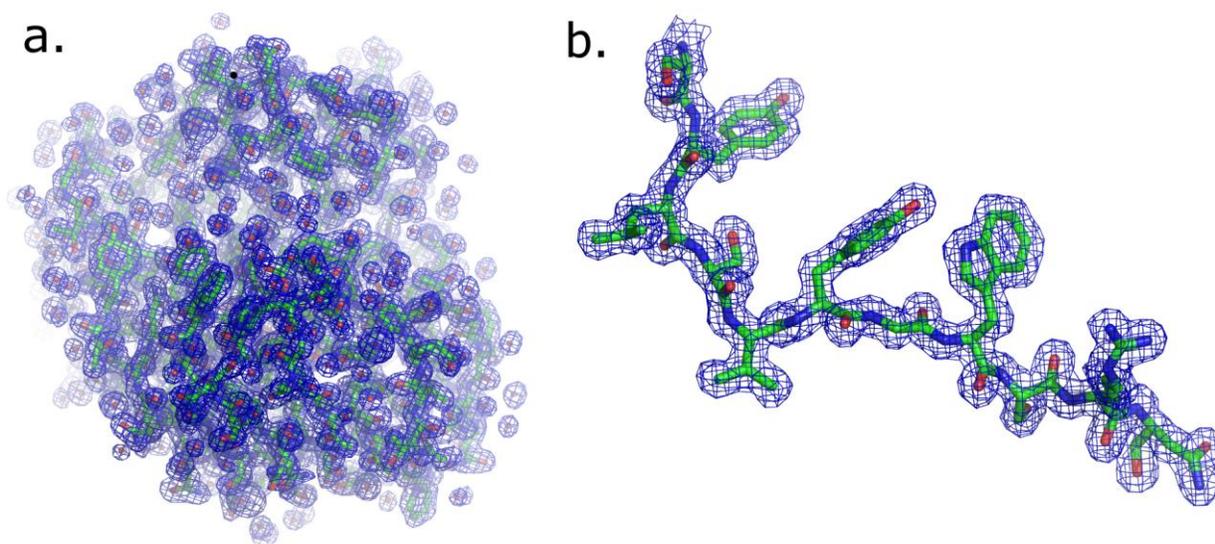


Figure 6. 2Fo-Fc electron density map contoured at 1 σ is superimposed on the refined atomic model of xylanase in stick form with carbon atoms green, nitrogen blue, and oxygen red. **a.** Entire xylanase molecule including ordered water molecules modeled as single oxygen atoms. **b.** A β -strand (residues 72 to 82) from the model shown individually to show map quality.

CONCLUSIONS

Our experiments demonstrate that the Bruker AXS SMART BREEZE™ system equipped with a copper sealed tube and Oxford Cryostream is an effective instrument for data collection from these protein crystals with smaller unit cells. We were able to collect a 40-fold redundant dataset from a lysozyme crystal with an average $I/\sigma I$ of 26.6 in a total time of 53 hours. A good quality 8-fold redundant dataset could have been collected in less than half a day, which underscores the efficiency of this system regarding macromolecular crystals.

Our xylanase project further demonstrates the feasibility of collecting complete, redundant, moderately high-resolution datasets on these protein crystals in a short time period, and that this system has the potential to provide a successful experience in protein crystallography for students in an academic biochemistry laboratory. Although longer exposure times would have likely improved the R_{sym} statistics of the data, our five-hour data collection strategy proved sufficient for a structure solution using the technique of molecular replacement, and resulted in a well-defined electron density map that was easily interpreted (**Figure 6**), and a molecular model with good stereochemistry and agreement with the experimental data. This seems like a very reasonable exercise to incorporate into an academic lab to provide a nice introduction to the field of protein crystallography, with a reasonably high chance of success for students.

Using this system we are able to address a wide range of crystallographic projects, ranging from simple ionic compounds and organic small molecules to larger macromolecules like small proteins. Our work shows that the signal to noise ratio at close distances is sufficient for these protein standards using the air-cooled BREEZE™ detector. X-ray systems for protein crystals have not previously incorporated this detector which is cooled to only -25 °C using an air-cooled Peltier device. Often, X-ray systems are instead equipped with the APEX II™, a more expensive detector with an identical CCD chip, cooled to -50 °C to minimize dark current, and requiring a water/ethylene glycol heat exchange system. In addition to being less expensive, the air-cooled BREEZE™ detector has significantly lower maintenance compared to this water/glycol-cooled detector. The cost of our system is less than half that of a dedicated protein data collection system, yet it still provides an effective platform for analysis of protein crystals with unit cells of up to at least 80 Angstroms. We look forward to future experiments analyzing protein crystals with larger unit cells, to assess the efficiency of data collection when the detector is pushed to greater distances to resolve more tightly-spaced diffraction patterns. The ability to collect data on such a wide range of samples, under a wide range of temperatures provides a relatively cost-effective and powerful analytical tool for academic labs and for a wide variety of undergraduate research projects at Southern Utah University.

ACKNOWLEDGEMENTS

We gratefully acknowledge funding from the Undergraduate Research and Scholarship Program (UGRASP) at SUU. We also thank the Department of Physical Science and The Walter Maxwell Gibson College of Science and Engineering for generous support of the X-ray diffraction facility at SUU. We are also grateful for the L.S. and Aline W. Skaggs Center for Health and Molecular Sciences and the Edward & Shirley Stokes Open Chemistry Laboratory for housing the X-ray system.

REFERENCES

1. The Official Web Site of the Nobel Prize. Physics Prizes and Laureates; Complete List, Nobelprize.org/nobel_prizes/physics/ (accessed Oct 2016)
2. Freidrich, W., Nipping, P.I., and Laue, M. (1912) Interferenz-Erscheinungen bei Röntgenstrahlen. *Bayerische Akad d Wiss zu München, Sitzungsber math-phys Kl* 303–322.
3. Bragg, W.L. (1913) The Diffraction of Short Electromagnetic Waves by a Crystal. *Proc Camb Phil Soc* 17, 43–57.
4. Bragg, W.L. (1914) The Structure of Some Crystals as Indicated by their Diffraction of X-rays. *Proc R Soc Lond A* 89(610), 248–277.
5. Rhodes, G. (2006) *Crystallography Made Crystal Clear*. Burlington, MA: Academic Press Publications.
6. Rupp, B. (2010) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. Garland Science, Taylor & Francis Group.
7. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N, Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) The RCSB Protein Data Bank, www.rcsb.org. *Nucleic Acids Research* 28, 235–242.
8. Kendrew J.C., Bodo G., Dintzis H.M., Parrish RG, Wyckoff, H., and Phillips, D.C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181, 662–666.
9. Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., and Shore, V.C. (1960) Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å. resolution. *Nature* 185, 422–427.
10. Perutz, M.F., Rossman, M.G., Cullis, A.F., Muirhead, H., Will. G., and North, A.C. (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. *Nature* 185, 416–422.
11. Perutz, M.F., Muirhead, H., Cox, J.M., and Goaman, L.C. (1968) Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. *Nature* 219, 131–139.
12. Personal communication, Larry Hanright, Field Service Supervisor, Bruker AXS.
13. Alderton, G., Ward, W.H., and Fevold, H.L. (1944) Isolation of lysozyme from egg white. *J Biol Chem* 157, 43–58.
14. Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C., Phillips, D.C., and Sarma, V.R. (1965) Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* 206, 757–761.
15. Diamond, R. (1974) Real-space refinement of the structure of hen egg-white lysozyme. *J Mol Biol* 82, 371–391.
16. Torronen, A., Rouvinen, J., Ahlgren, M., Harkki, A., and Visuri, K. (1993) Crystallization and preliminary X-ray analysis of two major xylanases from *Trichoderma reesei*. *J Mol Biol* 233, 313–316.
17. Watanabe, N., Akiba, T., Kanai, R., and Harata, K. (2006). Structure of an orthorhombic form of xylanase II from *Trichoderma reesei* and analysis of thermal displacement. *Acta Cryst D Biol Cryst* 62, 784–92.
18. Alderton, G. and Fevold, H.L. (1946). Direct crystallization of lysozyme from egg white and some crystalline salts of lysozyme. *J Biol Chem* 164, 1–5.
19. LYSOZYME.CO.UK (2006) <http://lysozyme.co.uk/crystallization.php> (accessed Jan 2012)

20. Bruker AXS (2013) SAINT, Version 8.32B, Bruker AXS Inc., Madison, Wisconsin, USA.
21. Bruker AXS (2012) SADABS, Version 2012/1, Bruker AXS Inc., Madison, Wisconsin, USA.
22. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007) Phaser crystallographic software. *J Appl Cryst* 40(4), 658–674.
23. The CCP4 suite: programs for protein crystallography (1994) *Acta Cryst D Biol Cryst* 50(5), 760–763.
24. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Cryst D Biol Cryst* 60, 2126–2132.
25. Murshudov, G.N., Vagin, A.A., Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst D Biol Cryst* 53(P3), 240–255.
26. The PyMOL Molecular Graphics System, Version 0.99 Schrödinger, LLC.

ABOUT THE STUDENT AUTHORS

Christiana Standler graduated from Brigham Young University-Hawaii (BYU-H) in April of 2014 with a B.S. degree in Pre-Professional Biology. In March 2015, she presented her multivariate statistics research from BYU-H at the International Convention of Psychological Science in Amsterdam. Christiana completed further undergraduate coursework at Southern Utah University (2014-2015) where she continued her research interests. Christiana began her MSc degree in Human Biology at the University of Copenhagen in September of 2015 and is expected to graduate in August 2017. She is currently completing her master's thesis at Bioneer A/S in Hørsholm, Denmark, with a focus on hepatocyte differentiation of human induced pluripotent stem cells.

Gary Blake Overson graduated from SUU in the spring of 2016 and is currently applying to matriculate into medical school for the fall of 2017. He is presently employed through The Classical Academies in Southern California, working as a high school chemistry instructor/tutor. He enjoys helping his students learn and understand the basics of chemistry while building their foundation in this subject. He enjoys spending his free time relaxing and taking care of his family. He desires to attend and complete medical school, and to help bridge the gap between the research being completed and the treatment of individual people.

Cody Black graduated from Dixie State University in May of 2009 with an A.S. then continued his education at SUU. He is currently working on his Pharm.D. at the University of New England in Maine, where he is honored to be recognized as a Rho Chi member, as well as fellowshipping on several research topics funded by NSF, such as the mechanism of antimicrobial drug-efflux activity across bacterial inner-membranes. He is also a member/leader of various organizations focused on addressing current health disparities and public education. He enjoys educating the community and colleges, and plans on pursuing a post-graduate education after graduating from UNE in May of 2018. Outside of his various educational and community outreaching activities, he relishes his time with his wife and children by exploring the beauty of the outdoors.

Guizella Rocabado graduated from Aims Community College in May of 2009 with an A.S. degree and was awarded the President's Medallion. She continued her education at SUU where she earned a B.S. degree in biology and a B.S. degree in chemistry in May of 2012. She continued her studies and earned a M.S. degree in medicinal chemistry from the State University of New York at Buffalo (UB) in September of 2014 where she researched the effect of tetraurea macrocycles on G-quadruplex DNA. She was awarded the Schomburg Fellowship for underrepresented students as well as membership in the Collaborative Learning and Integrated Mentoring in the Biosciences (CLIMB) program at UB during her graduate work. She is currently working as an assistant professor of chemistry at SUU and is expected to start her doctorate work in chemical education at the University of Southern Florida in the Fall of 2017.

PRESS SUMMARY

Our experiments demonstrate that the Bruker AXS SMART BREEZE™ system equipped with a copper-anode sealed tube and Oxford Cryostream is an effective instrument for data collection from protein crystals of cell dimensions of at least up to 80 Angstroms. The cost of this system is less than half that of dedicated protein data collection systems, yet still provides an effective platform for protein crystal analysis in addition to its primary role in chemical crystallography.

Ebola Impact and Quarantine in a Network Model

Anca Rădulescu*, Joanna Herron

Department of Mathematics, State University of New York at New Paltz, New Paltz, NY

Student: herronj1@hawkmail.newpaltz.edu

Mentor: radulesa@newpaltz.edu

ABSTRACT

Much effort has been directed towards using mathematical models to understand and predict contagious disease, in particular Ebola outbreaks. Classical SIR (susceptible-infected-recovered) compartmental models capture well the dynamics of the outbreak in certain communities, and accurately describe the differences between them based on a variety of parameters. However, repeated resurgence of Ebola contagions suggests that there are components of the global disease dynamics that we don't yet fully understand and can't effectively control.

In order to understand the dynamics of a more widespread contagion, we placed SIR models within the framework of dynamic networks, with the communities at risk of contracting the virus acting as nonlinear systems, coupled based on a connectivity graph. We study how the effects of the disease (measured as the outbreak impact and duration) change with respect to local parameters, but also with changes in both short-range and long-range connectivity patterns in the graph. We discuss the implications of optimizing both these measures in increasingly realistic models of coupled communities.

KEYWORDS

Epidemic Spread; Network Dynamics; Network Connectivity; Coupled Differential Equations; Compartmental Model; Information Transfer; Outbreak Impact; Outbreak Duration

INTRODUCTION

Background to Ebola and modeling

The evolution, prognosis and spreading of contagious diseases has been studied for a long time, with a variety of approaches.^{5,3,23} While huge progress has been made in efficiently applying containment methods^{22,38} and treatment⁵⁰ in most cases, Ebola remains the 21st century's taunting example of a disease which mankind does not seem prepared to handle, even at small scales.^{12,18}

While the course of the illness is quite drastic and fatality rates have ranged in past outbreaks from 45-90%⁴⁶ (even with prompt clinical intervention), it has been argued that Ebola, with a relatively small reproduction number $1 \leq R_0 \leq 7$ (varying among reports of different outbreaks),^{12,32} poses a lesser contagion threat than faster spreading diseases such as small pox ($R_0 > 7$). However, the explosive Ebola contagions that resurge periodically, and in particular the recent simultaneous developments in a few countries around the world³¹ – seem to suggest that there are components of the disease dynamics that we don't yet fully grasp,¹⁶ as well as social,⁴¹ cultural and economic²⁴ key parameters that we cannot yet fully control.

Much effort has been recently directed towards understanding and predicting Ebola outbreaks with the help of mathematical modeling.^{32,42,34,Siettos et al.} A large volume of the existing modeling work addresses disease dynamics via low dimensional compartmental models,^{45,25,44} deterministic or stochastic.³³ These models describe well

outbreak dynamics in specific communities, and often explain the differences between outbreaks that occurred under different parameter regimes (e.g., geographical location, promptitude of clinical measures, timely removal of infected individuals from the community^{4,12,44}). The most traditional and well-known modeling approaches to disease spread remain the SIR-type models, which describe the progress of the contagion in a single community through transfers between three main compartments: the susceptible (S), infected (I) and recovered (R) individuals. Variations of the SIR model have been vastly explored, in conjunction with parameters estimated empirically, and in increasingly complex contexts, including factors like hospitalization, treatment plans, quarantine. However, these analyses – based on localized factors and data from remote rural areas – may all be irrelevant in the context of utmost concern, that of a global outbreak,^{20,9} affecting urban areas as well as small communities, acting at multiple simultaneous foci. In order to understand the dynamics of a global contagion, predict its potential effects and most efficiently alter its course – one rather needs to place the problem in the mathematical framework of complex systems.^{37,36}

The huge oriented graph that represents physical connectivity between individuals is highly modular, has hubs, multiple-scale communities (e.g., from families to towns to states), and is constantly changing due to a variety of factors (e.g. personal, economic, commuting patterns, long distance travel). To capture the global dynamics of epidemics in such a system, one must consider simultaneously the state of each node as well as its interconnections with other nodes.¹⁷ An accurate approach should bridge looking at single node dynamics to studying the effects of connectivity patterns on the systemic behavior⁶ and on the outbreak aftermath (asymptotic behavior). Altering dynamics in an optimal way requires knowledge of architecture,⁷ of the location and density of hubs and of rich clubs,⁵² of the robust versus vulnerable network points¹⁰ (where local edge changes produce large effects).

Networked dynamic systems⁵³ have been used in a variety of fields with a focus on understanding the behavior of an ensemble of coupled dynamic nodes, be they cells, web servers, people, or nodes in an electric grid. However, in recent years, there has been an emerging interest in *network* modeling of epidemics,^{17,6,26,40} trying to quantify and better understand the effect of dynamic human interactions on the spread of disease between, as much as within communities, and looking to identify the factors (both deterministic and stochastic²⁸) most determinant of the long term outcome.

It was noted, when studying the effects of contact heterogeneity on the dynamics of communicable disease,⁴⁹ that small differences in the contact networks (e.g., taking into account restructuring at a time resolution of minutes), are typically not essential when attempting to describe disease spread on a longer timescale (of several weeks, or months). On the other hand, the same studies^{49,48} emphasize the importance of including detailed information about the heterogeneity of contact duration, the rate of new contacts being identified, etc. Relatively new work, including our own, is using a combination of dynamical systems and graph theoretical methods to understand how relatively small, local perturbations in connectivity patterns may produce, if targeted at the right vulnerability points, very large effects on the state of the entire network. The avenue of modeling mathematically the dynamic epidemic interactions in an infected network has been pursued in a variety of contexts, including overlay³⁵ and adaptable networks.²¹ Along these lines, we have developed methods and measures of how likely the system is to undergo a sharp transition when perturbing local connectivity.⁴³

The emergence of a contagion (of Ebola, or of any other viral or bacterial infection) can be then viewed as the propagation of a perturbation in a complex network of coupled nodes.²⁹ One can then study how adapting the structure of the network may contribute to minimizing the global effects of the contagion. Experience has shown that, along with other important factors, timely placement of temporary travel restrictions or quarantines may be critical in isolating and extinguishing the contagion before it reaches catastrophic levels.¹⁹ One can then search for the locations where a small change in connectivity is likely to produce the largest overall effects.¹⁵ This goes along the lines of other new approaches in the field, which introduce the network structure as a system parameter,^{13,27}

and search for optimal quarantine measures to efficiently isolate the epidemic.

In this paper, we consider, in conjunction with our model, two measures of outbreak effects on population: the *impact*, defined as the total fraction of individuals affected by the outbreak (eventually either recovered or dead), and the outbreak *duration*, defined as the time from the start of the outbreak (presence of the first Ebola case), until the outbreak extinguishes (less than one infected case in each community). Starting with a basic SIR model in a single, relatively small (1000 individuals) community, we construct incrementally a plausible model for Ebola spread within a network of such communities. We focus in particular on studying how our two measures of outbreak dynamics depend on the density of potential connections between these communities and on the stochastic likelihood of people to travel along these connections. The network architecture is allowed to change along the process, controlled by factors related to the current state of the outbreak, accounting for travel interdictions and quarantines typically introduced in such circumstances. Our model aims to inform on the most efficient quarantining strategies and optimal timing that would minimize life loss as well as outbreak duration.

MODELING METHODS AND RESULTS

Basic model

We build upon a basic compartmental model of Ebola due to Astacio et al.,⁴ which was originally assembled as a classical three-variable SIR (Susceptible-Infectious-Recovery) model. In the later iterations of the model construction, the authors introduced incubation in the model, by means of a new (fourth) variable E , representing the “latent” (exposed) population (during the virus incubation, before development of symptoms).

$$\begin{aligned} \frac{dS}{dt} &= -\beta S(I + qE)/N \\ \frac{dE}{dt} &= \beta S(I + qE)/N - \delta I \\ \frac{dI}{dt} &= \delta I - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \quad \text{Equation 1.}$$

where $S(t)$ is the susceptible population at time t (i.e., everyone who had not yet contracted the disease); $E(t)$ is the latent population (individuals who have contracted the virus, but are still asymptomatic); $I(t)$ represents the infected population (showing the signs and symptoms of Ebola); $R(t)$ is the number of dead or recovered individuals (i.e., in an oversimplified view, individuals who can no longer infect others with the Ebola virus). The infection rate (i.e., the rate $dE/dt = -dS/dt = \beta S(I + qE)/N$ at which people become infected) is proportional with the product of the number of susceptible individuals and the number of individuals carrying the virus (both latent and infected, with the latent individuals having a lesser impact, represented by the smaller weight $q < 1$). The proportionality constant β is the product of the per capita contact rate and the probability of infection after contact with an infected individual. The infection rate was normalized by the factor $N(t)$, representing the total population at time t : $N(t) = S(t) + E(t) + I(t) + R(t)$. The rate of transfer from the latent to the infectious stage is a fraction δ of the number of latent individuals – where $1/\delta$ is the average time for a latent individual to become infectious. The rate of death/recovery is a fraction γ of the infectious population, where $1/\gamma$ is the average time it takes a person to die or recover once in the infectious stage.

The parameters were based on the disease dynamics described in a relatively well localized 1976 contagion in Yambuku, Zaire. The ranges of values, as per our reference, are listed in Table 1. In our study, we worked primarily with these values; changes and extensions, based on information on post-mortem and post-recovery potential for

contagion, are explored in the Discussion. In this study, however, we work under the same simplifying assumption as the one made by Astacio et al.:⁴ that neither deceased nor recovered individuals can contribute any further to the perpetuation of the disease cycle, hence are represented by a common variable, whose asymptotic value $R(t)$, whose asymptotic value as t approaches ∞ can be viewed as measuring the *impact* of the outbreak (total fraction of people affected).

Parameter	Range	Value	Units
$1/\gamma$	4-10	7	days
$1/\delta$	2-21	12	days
q	-	0.25	-
β	-	0.567	-

Table 1. Parameter values for the Yumbuku outbreak, as per the Astacio et al. reference.⁴

Simple network model

We first studied contagion propagation in a small, unstructured network of interconnected communities. Although this preliminary model makes a few coarse and rather unrealistic simplifying assumptions, this first stage helps understand some very basic problems and questions to address.

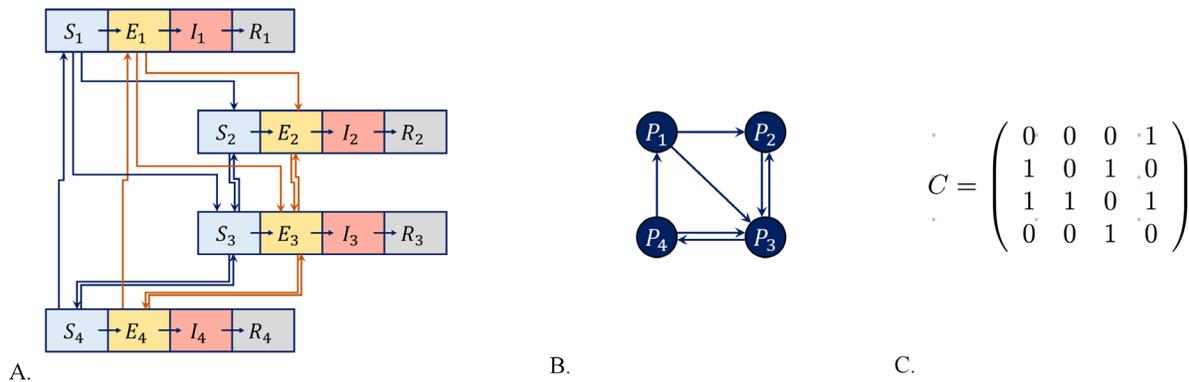


Figure 1. Model network, for $n = 4$ communities, (as used in some of our later simulations). **A.** Illustration of the network, showing each population as a compartmental node, with its 4 coupled SEIR variables, and showing transfer between populations as oriented arrows from the original population to the target population. **B.** Schematic representation of the same network, in which each population is viewed as a node, and the connections between populations are viewed as oriented edges. **C.** Adjacency matrix C corresponding to the connectivity graph described in (B).

We considered a graph in which the n nodes represent the interconnected communities, and the oriented edges connecting node pairs represent one-way communication between the two respective nodes. Each node/population P_k , $1 \leq k \leq n$, represents a standard SEIR unit (as described in the previous section), characterized by a 4-dimensional variable (S_k, E_k, I_k, R_k) . The communication between nodes was set up, for consistency, also in a compartmental fashion, so that a fraction of individuals travels through each existing outgoing edge from a specific node to the corresponding adjacent nodes. For simplicity, each node was assumed to have originally $N_0 = 1000$ individuals. The total number of individuals $N(t) = \sum[S_k(t) + E_k(t) + I_k(t) + R_k(t)]$ is constant throughout the

outbreak (that is, external effects such as travel in and out of the network, birth rate and death rate due to factors independent of the disease were ignored). The corresponding coupled dynamics can then be extended from the system Equation 1. to the following system:

$$\begin{aligned}
 \frac{dS_k}{dt} &= -\beta S_k(I_k + qE_k)/N_0 + q_c[CS]_k - q_c \sum_j C_{j,k} S_k \\
 \frac{dE_k}{dt} &= \beta S_k(I_k + qE_k)/N_0 + q_c[CE]_k - q_c \sum_j C_{j,k} E_k - \delta E_k \\
 \frac{dI_k}{dt} &= \delta E_k - \gamma I_k \\
 \frac{dR_k}{dt} &= \gamma I_k
 \end{aligned}
 \tag{Equation 2.}$$

where $[CS]_k$ and $[CE]_k$ represent the k th component of the vectors CS and CE , respectively. The $4n$ dimensional system Equation 2. reflects the inner SEIR dynamics of each particular node $P_k = (S_k, E_k, I_k, R_k)$, for $1 \leq k \leq n$, but also captures, in a compartmental way, the population flow between the nodes. The adjacency matrix C is defined so that $C_{j,k} = 1$ if there is directed travel (an oriented edge) from the node P_k into the node P_j . It is zero diagonal (the graph has no self loops, since travel within one’s own community is not relevant).

The number of healthy and, respectively, latent individuals leaving node P_k at time t , headed towards each of the adjacent nodes P_j is $q_c C_{j,k} S_k$ and, respectively, $q_c C_{j,k} E_k$. If there is an edge between nodes P_k and P_j , these numbers are proportional with the existing healthy and, respectively, latent population in node P_k . For simplicity, the proportion of travelers q_c between adjacent nodes was taken to be the same for all connected node pairs (and was fixed to $q_c = 2.5\%$ in our simulations of this system). The rate out of S_k due to travel to all efferent nodes is therefore $q_c \sum_j C_{j,k} S_k$, and the rate out of E_k due to travel is $q_c \sum_j C_{j,k} E_k$.

Subsequently, each node will receive an incoming flow of travelers from the nodes connected to it. More specifically, the rate due to travel of healthy people from every node P_j into node P_k is $q_c C_{k,j} S_j$, so that the total rate into S_k due to travel from other nodes is $q_c \sum_j C_{k,j} S_j = q_c [CS]_k$. Similarly, the total rate into E_k due to travel from other nodes is $q_c \sum_j C_{k,j} E_j = q_c [CE]_k$. We based this scheme on the assumption that latent individuals are able to travel out of their own community: being asymptomatic, they have not yet been diagnosed, although they pose a risk in spreading the disease, by adding themselves to the existing latent population at a different node.

Infected individuals cannot travel in our model. In addition, circulation of recovered individuals was ignored, since it would have no effect on the disease dynamics (under the assumption that they can no longer contract or spread the disease, they would only permute between the R_k compartments of different nodes).

The model was conceived having in mind typical rural structures, where people travel along established routes between specific locations (for daily or other periodic needs), with the return not necessarily occurring immediately or along the same path (hence the oriented edges). The graph adjacency matrix C delivers a complete description of the communication patterns within the network, and induces instantaneous spread of contagion. Throughout our analysis, we kept all parameters fixed and we focused primarily on how the connection density and patterns affect viral diffusion through the network.

As a start, we studied the effect of edge density on the disease dynamics, in particular on the outbreak impact and

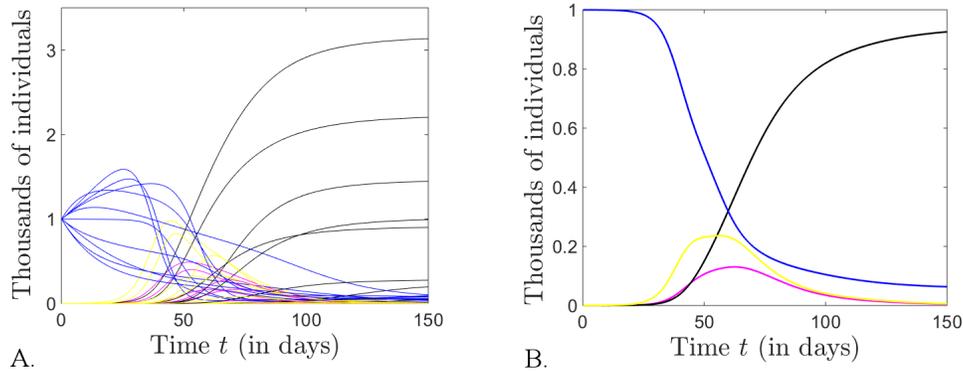


Figure 2. Simulation of interactive dynamics in a network with $n = 10$ communities, in which the contagion was triggered by only one latent individual in P_1 . A. The time evolution of all SEIR variables, obtained by numerically solving system Equation 2., is illustrated simultaneously for all $n = 10$ nodes, with the S_k shown in pink, the E_k shown in yellow, the I_k shown in red and the R_k shown in black. B. The average populations $\bar{S}(t) = \frac{1}{n} \sum_{k=1}^n S(t)$, $\bar{E}(t)$, $\bar{I}(t)$, $\bar{R}(t)$ are shown, with the same color coding.

duration. For our network model, we define the impact as:

$$R_\infty = \lim_{t \rightarrow \infty} \bar{R}(t), \text{ where } \bar{R}(t) \text{ is the recovered average per community: } \bar{R}(t) = \frac{1}{n} \sum_{k=1}^n R_k(t)$$

For this first model, we assume a constant connectivity matrix throughout the process, and we investigate the cases of a single focal point (triggered by two infections), and of two focal points (with one infection each). Fixing the network size, we ran numerical simulations for different connectivity patterns, with the primary aim of investigating the effect of the connectivity density on the outbreak coupled dynamics. For small network sizes ($n \leq 4$), we considered all possible matrix configurations for each fixed edge density $0 \leq \Delta \leq n(n - 1)$, and we computed the mean impact over all such configurations (see Figures 3). For larger n 's, we computed a sample-based mean, over a subset of 50 configurations for each edge density Δ (Figure 4a and b). This sample approach was preferred for larger networks, since the number of configurations increases extremely fast with n , making a numerical inspection of all configurations for each fixed density very expensive and impractical. In both cases, we plotted the outbreak impact and duration as functions of the edge density Δ , showing both mean value and error bars (over the distribution of adjacency configurations).

Our computations suggest that, for both one and two foci, the outbreak impact increases with the density of connections. This is because the outbreak impact depends, as one might expect, on the average number of nodes connected to the foci, which in turn depends on the graph's edge density (as shown in Figure 4c). For both one and two foci, the duration has a unimodal shape (with a peak in the intermediate connectivity range). This is not a surprise either, and can be explained by the fact that for very low connectivity the contagion is quickly contained, with minimal casualties, and for very high connectivity - it produces a fast diffusion, with a high number of casualties. It is for intermediate connectivity patterns that the dynamics takes longer to develop.

Clearly, one desires to reduce duration, as well as impact (since longer infection times can be as well detrimental for the community). The two measures don't assume their minima in the same conditions. Given that most parameters of the system (such as R_0 value, number of individuals, connectivity graph) are hard-wired into the network, one may consider imposing conditions on travel along the existing edges (quarantines), in an attempt to optimize a function of both impact and duration, and assess possible regulations that would render the outcome least destructive. For example, some studies have documented the dramatic impact of long outbreaks on economy.^{9,1} However,

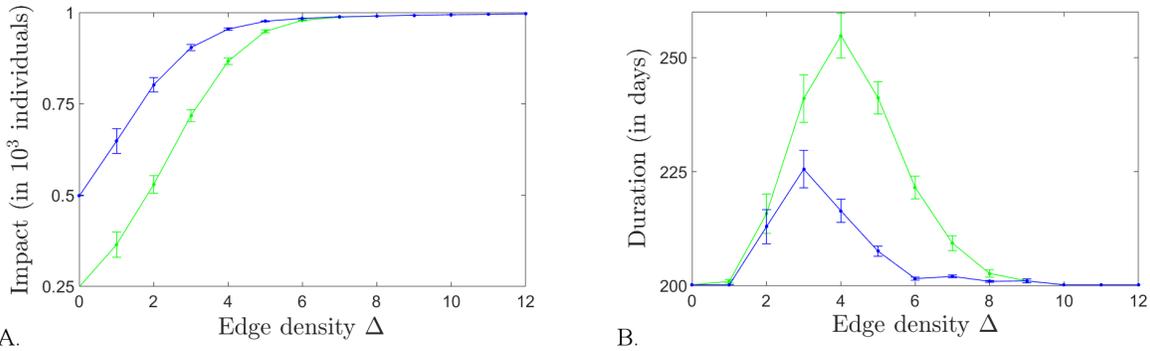


Figure 3. Dependence of the impact R_∞ and of the outbreak duration L on edge density Δ . For a network of $n = 4$ populations each originally with $N_0 = 1000$ individuals, we illustrate R_∞ and L as the edge density is varied for $\Delta = 0$ edges to $\Delta = 12$ edges. For each density value, the average impact and duration were computed over all possible $C_{n(n-1)}^\Delta$ adjacency configurations with the respective density. We considered the case of one focal point consisting of two latent cases, both in the same node (green curve), and the case of two focal points, each consisting of one latent case, starting simultaneously at two distinct nodes (blue curve).

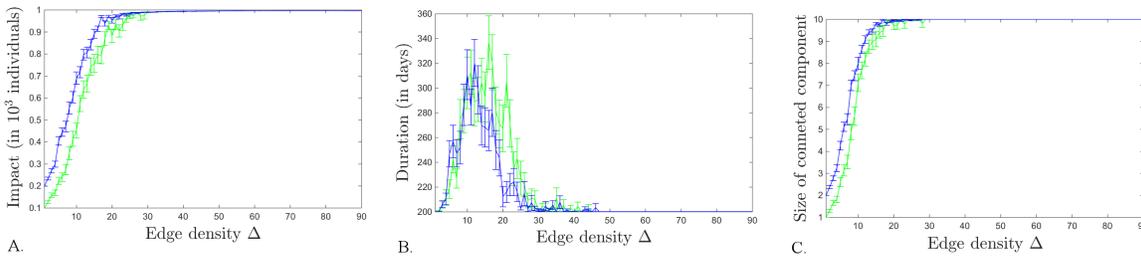


Figure 4. Dependence of the impact R_∞ and of the outbreak duration L on edge density Δ . For a network of $n = 10$ populations each originally with $N_0 = 1000$ individuals, we illustrate R_∞ (in panel **A**) and L (in panel **B**) as the edge density is varied. Panel **C**, shows the average number of nodes connected (via oriented graph paths) to the initially infection focus or foci as a function of edge density. For each density value, the average impact, duration and size of the infected connected components were computed over a sample of 50 adjacency matrices chosen at random, with the respective density. We considered the case of one focal point consisting of two latent cases, both in the same node (green curve), and the case of two focal points, each consisting of one latent case, starting simultaneously at two distinct nodes (blue curve).

although lifting quarantines and increasing travel may shorten duration (as suggested by our model), this is not a feasible policy, since it would lead to an unacceptable increase in casualties. Some of these aspects will be further discussed in the next section.

It is also worth noticing that the impact has a slower pick-up in the case of one original focus, while the duration is comparable for one or multiple foci in the low connectivity range, then significantly longer for one focus in the intermediate connectivity range. In other words: for intermediate internode communication, the impact is milder if the original infection starts in one single focus, but this scenario loses its advantage in the context of duration.

More precisely: in this preliminary model, R_∞ is relatively low in average only for values of the edge density up to $\sim 20\%$ (as illustrated for $n = 4$ in Figure 3 and for $n = 10$ in Figure 4). After crossing a window of very high sensitivity (in our figures, density $15 - 20\%$), R_∞ increases to its asymptotic value $N_0 = 1000$ (the total number of individuals per node). This predicts a devastating effect, describing an outbreak that would affect virtually everyone

within the linked communities, even at relatively low communication levels.

Of course, however feared and somber the perspective of a global pandemic, this is not the prediction we expect in reality, based on our existing experience with Ebola outbreaks; hence the model should be revised to reflect more realistic trends. One major flaw of this preliminary model is the assumption of instantaneous and continuous disease spread whenever communication means exist between two nodes. This is clearly not a realistic condition, greatly overestimating the contagion spread, which further reflects in the unrealistically large values of R_∞ . Below, we update the model to refine this condition.

Existence of an oriented edge from k to j should only signify that direct communication is *possible* between the two nodes (e.g., there is a road connecting two villages), but not necessarily that this connection is used continuously, or maintained at the same level at all times. We introduce a simple way of accounting for this variability, while keeping the model compartmental (i.e., counting overall transfer rates, rather than keeping track of dynamics of individuals). At each time step t , each existing connection from some k to some j is used with a fixed probability r_{kj} . For each node, the probability of inward or outward travel can be tuned according to the local or global situation of the network. For example, the flow can be temporarily diminished or cut completely if the node needs to be quarantined to prevent infection spread. For our first analysis, we take the rate travel r for simplicity to be constant throughout the outbreak process, and identical for all outgoing edges. In the following section, we will allow r to adapt, producing a variable distribution of values over all the network edges, changing according to the implementation measures typically taken to minimize the impact of the outbreak.

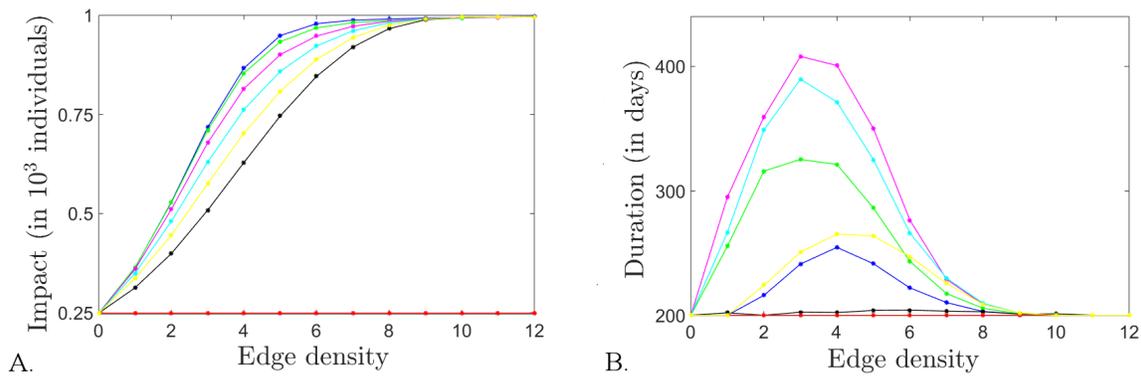


Figure 5. Dependence of outbreak aftermath on edge density Δ and travel likelihood r , for a network of $n = 4$ communities. The impact R_∞ (A) and outbreak duration (B) are computed, for each fixed edge density (number of 1s in the adjacency matrix) as averages over all the possible configurations of the adjacency matrix, with that fixed density. The different color plots correspond to different values of the probability to travel along an available edge: $r = 1$ (blue), $r = 0.5$ (green), $r = 0.2$ (magenta), $r = 0.1$ (cyan), $r = 0.05$ (yellow), $r = 0.01$ (black), $r = 0$ (red).

Figure 5 illustrates in the case of probabilistic travel (described above), and for a single focal point of one individual, the dependence of impact and duration on both the density of edges, and on the probability r to travel out of each node along the prescribed edges. Notice that impact and duration, while having the same qualitative dependence on density, exhibit subtle differences when changing the overall likelihood to travel r . For example, in the case of $n = 4$ coupled nodes, the impact is $R_\infty = \frac{1}{4}N_0 = 250$ for $r = 0$, since no travel implies that only the focal community will be affected (there is no dependence on edge density, since the edges are never used to transport the virus). Moving r to any small positive value acts as a bifurcation, since the effects of the density Δ appear immediately, causing the impact to increase, more quickly and steeply for larger values of r . These being said, however, the dependence on travel likelihood in the $r > 0$ range is not as dramatic in this model as one may

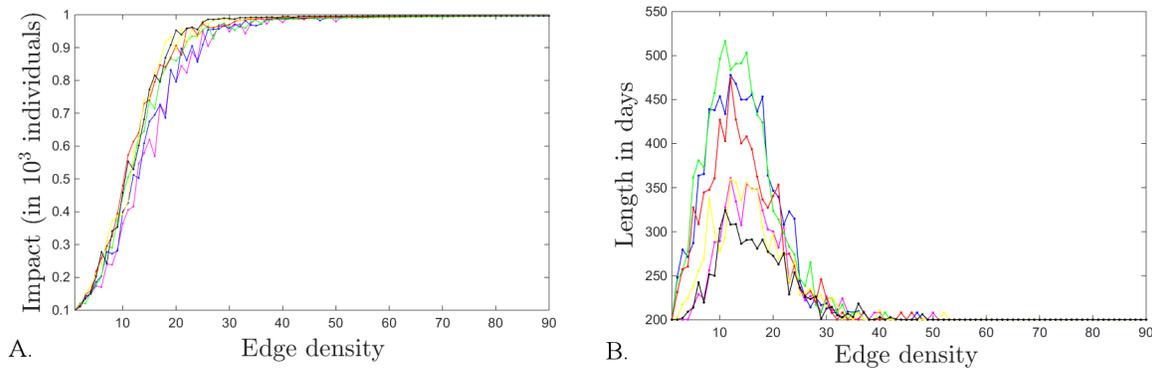


Figure 6. Dependence of outbreak aftermath on edge density Δ and travel likelihood r , for a network of $n = 10$ communities. The impact R_∞ (A) and of the outbreak duration (B) are computed, for each fixed edge density, as averages over a sample of $S = 50$ adjacency configurations with that fixed density. The different color plots correspond to different values of the probability to travel along an available edge: $r = 0.05$ (magenta), $r = 0.1$ (blue), $r = 0.2$ (green), $r = 0.5$ (red), $r = 0.8$ (yellow) and $r = 1$ (black).

expect; it suggests that indiscriminately diminishing travel in the network can't single-handedly accomplish a major decrease in impact, unless the travel is altogether prohibited.

An even more surprising effect appears when studying the dependence on r of the unimodal curve representing the outbreak duration (Figures 5b and 6b). While the critical *point* does not vary much – remaining, for each curve, broadly in the same intermediate density range for each n (a more precise localization could be obtained at the expense of lengthier computations), the behavior of the critical *value* differs both qualitatively and quantitatively between different r values, first increasing with r , and then decreasing. This suggests that indiscriminately lowering r may in fact be detrimental, by increasing the duration of the outbreak, without substantially lowering the impact (especially in the region of intermediate densities). Only a dramatic and implausible shift of r to a value close to zero throughout the network would result in improving both impact and duration. These effects can be observed in both Figures 5 and 6, for $n = 4$ and $n = 10$, respectively.

Altogether it seems clear that, in order to control the outbreak in our model, one cannot just treat the system as a whole, but instead has to target more specific network sites, based on (1) the network architecture and (2) the outbreak's current state throughout the network, starting from the source of contagion.

Modular adaptable network and effects of quarantine

In this section, we will simulate an outbreak in the more realistic context of two interacting subnetworks (or modules), and introduce more structured quarantine measures, with the aim of reducing both impact and duration. As a plausible, but simplified scenario, we consider the modules to be organized as *hubs*, in the sense that each has a central node connected bidirectionally with all other nodes in the respective subnetwork, also allowing a specific number of additional random oriented edges between the other nodes of the respective module (as in the example pictured in Figure 7). This aims to represent the interaction of two large structures (e.g., countries), each organized as a network of smaller communities (e.g., towns, or villages). Such a multi-modular graph contains local, intra-modular connections between nodes (e.g., roads between villages), and long-distance, inter-modular connections. In our case we considered, for simplicity, a single inter-modular, bidirectional edge, running between the two central hub nodes (which could be seen as the only significant communication means between the two countries, e.g. – airports).

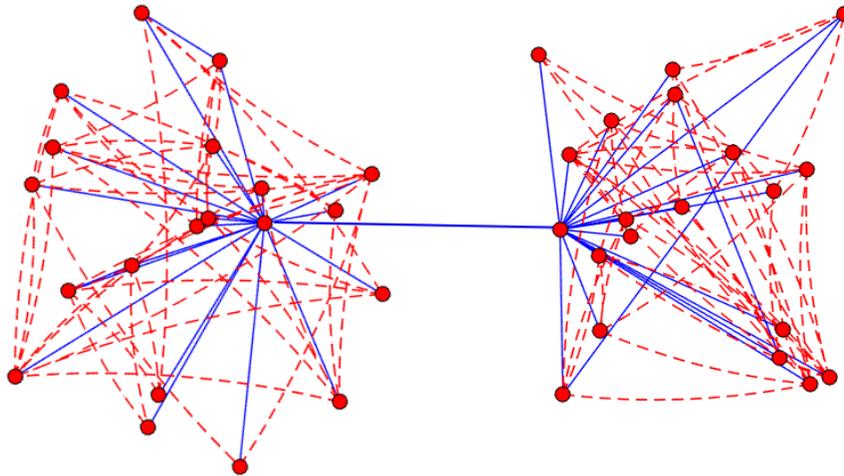


Figure 7. Model network architecture with two interconnected modules. Both modules (subgraphs), each with $n = 20$ nodes and $\Delta = 100$ oriented edges, are organized as hubs, each with one central node connected bidirectionally with all other nodes in the corresponding hub. The two hubs are connected by one single, bidirectional edge between their two central nodes. In this figure, bidirectional edges are represented in blue, and unidirectional edges are represented by dotted red curves, curved counter clock-wise. This type of architecture was used for the simulations, e.g., in Figure 8.

We discuss infection transmission and impact in this type of network, and the efficiency of introducing timely quarantines to control the outbreak. We distinguish between two types of quarantine, local (intra-modular) and global (inter-modular), as follows. If Ebola infection is detected for a period longer than θ days in a population/node (i.e., $I_k(t) > 0$ for the specific node P_k), a local quarantine is introduced by cutting all in and out connections with the node P_k . If infection persists at any node within either hub for longer than τ days after the local quarantine (with $\tau > 0$), then the two hubs are immediately disconnected (the connecting edge is cut off). We study how the timing of these isolation measures affects long term dynamics, towards finding a scheme that would minimize the inconvenience of lengthy quarantine, while still delivering efficient outbreak containment. While one naturally expects a prompter quarantine response to lead to a better outcome, our study looks in more detail at the extent to which the timing matters at both local and global levels.

We considered different architectures, edge densities and travel likelihoods (r_{int} along the active intra-modular edges and r_{ext} along the inter-modular edge), at the start of the outbreak (time of the original infection in the network), and we measured in each case the dependence of the outbreak impact R_{∞} on the time delays θ and τ of the two quarantines. In Figure 8 we illustrate, for modules of size $n = 10$, the dependence of R_{∞} on θ and τ (with each curve showing the dependence on τ for a different value of θ , increasing from blue to green to red). The critical range (of maximum sensitivity) for θ is less than one day; the critical range for τ is a little longer. The top panels reproduce the outcome for a lower connectivity network (30 out of the maximum of 90 possible edges in each module); the bottom panels consider a higher connectivity network (50 edges). The left panels present results for lower travel likelihoods before the infection (intra-modular $r_{\text{int}} = 0.4$ and $r_{\text{ext}} = 0.2$ along the inter-modular edge); the right panels show higher pre-infection travel likelihoods ($r_{\text{int}} = r_{\text{ext}} = 0.5$).

Generally, as one may expect, for any fixed local delay θ , the impact increases with the inter-modular delay τ (the longer one waits to completely separate the two modules, the larger the infection-produced damage). Similarly, for each fixed τ , the impact increases with the local delay θ (with the same wait time before separating the modules, a longer local delay increases the global average damage). However, these qualitative effects present wide quantitative

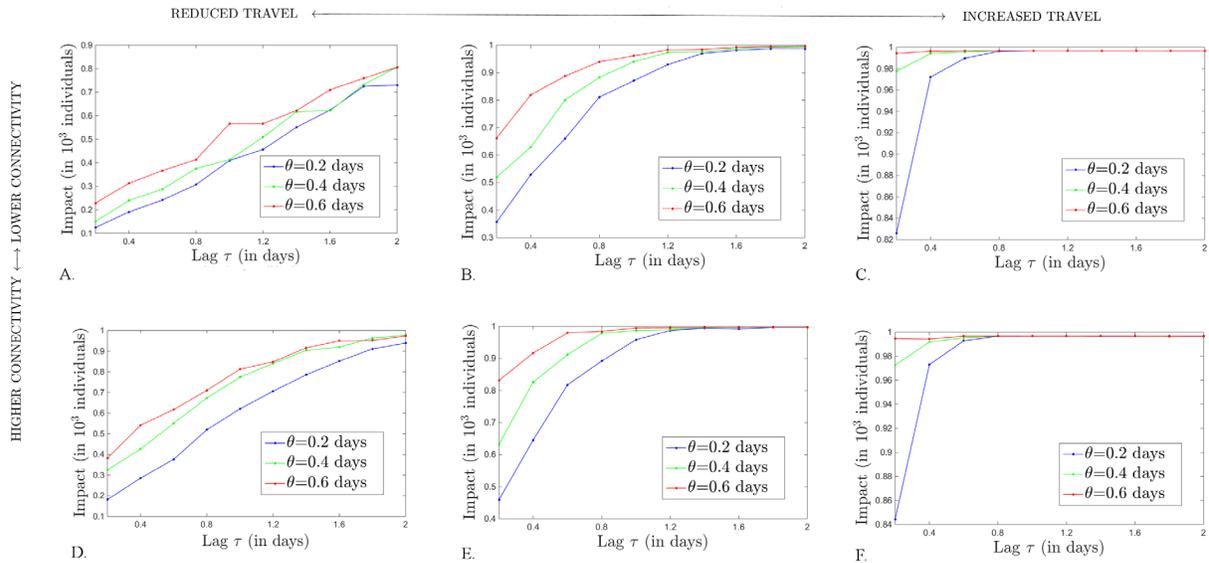


Figure 8. Dependence of impact R_∞ on quarantine time lags θ and τ for a network composed of two hubs, each with $n = 10$ nodes, connectivity density $\Delta = 30$ and (top panels) and $\Delta = 50$ (bottom panels) A. and D. travel likelihood $r_{int} = 0.2$ along intra-modular edges and $r_{ext} = 0.1$ along the inter-modular edges; B. and E. travel likelihood $r_{int} = 0.4$ along intra-modular edges and $r_{ext} = 0.2$ along the inter-modular edges; C. and F. travel likelihoods $r_{int} = r_{ext} = 0.5$. Each curve represents the dependence of R_∞ on the inter-modular lag τ , for a fixed intra-modular lag θ , as follows: $\theta = 0.2$ days (in blue), $\theta = 0.4$ days (in green) and $\theta = 0.6$ days (in red).

variations with the structure and connectivity of the network.

One may note that, within the realistic connectivity range, there are no wide differences determined by the edge density *per se* (panels A and D, as well as B and E have only subtle quantitative differences, and panels C and F look almost identical); the more substantial differences are induced by people’s likelihood to travel along the existing edges before the quarantines are imposed.

For very low original travel rates along edges (left panels), local and global quarantines both drastically decrease impact in an almost linear fashion, with a slightly higher efficiency when shortening the global, rather than the local quarantine response.

For a network with medium travel along edges (center panels), the scenario is different: setting early quarantines has a very strong effect (the curves increase steeply for low values of τ), effect which tapers off asymptotically as τ is relaxed. The impact also differs a lot with θ , especially for values of τ less than one day (when the impact gets close to the common asymptotic value). In this case, it seems important to strive for a short local wait θ . However, comparing the rates of change along each curve and across curves, it appears that shortening θ and τ has in this case comparable effects on the impact.

For highly traveled networks (right panels), a slightly large τ may lead to a catastrophic impact even for very low θ values. On the other hand, a large θ single-handedly leads to a dramatic impact, even with very short τ . In such networks, only a combination of both small θ and small τ can substantially reduce the impact. An efficient control of the impact requires very quick local intervention, followed tightly, possibly even before observing success or failure of local measures, by global separation.

DISCUSSION

In this paper, we investigated how the effects of an Ebola outbreak depend on the mobility of individuals in a network of connected populations exposed to the virus. This falls in the realm of a widely studied theoretical problem: that of dependence of coupled nonlinear dynamics on the architecture of the coupling. For our illustration, we worked with system parameters measured in a historic Ebola outbreak, but the same compartmental construction, the same concepts and methods can be used to study other viral transmissions, or any information diffusion process in a network of coupled nodes.

Aside from common sense conclusions on the necessity of prompt quarantines at early signs of a viral outbreak, our study brings interesting quantitative highlights, with potential applications when customizing and tuning quarantine placement and timing. There are a few particular observations that may be of general value when assessing the importance of quarantine and other measures to control an outbreak. First, we saw that conditions which minimize the outbreak impact may not be ideal for other aspects (e.g., duration), as important for the affected population. Second, a prompt local quarantine is generally helpful, but cannot efficiently lower the impact in and of itself, if not paired with a prompt global separation. A delay in global separation may make the efficiency of local quarantine irrelevant. Third, fast global separation is optimal if paired with a prompt local quarantine, and in some cases this can efficiently and substantially lower the impact.

This suggests that a helpful approach to infection control in a hypothetical outbreak would require authorities to assess the type of network and travel flow that is at risk, and tune the quarantine timing to occur within the respective optimal ranges. For example, in the case of a highly traveled modular network, the global separation may have to be imposed faster than common sense suggests, even if this may involve additional resources in putting in motion the slow global machinery.

Our study is only one step, among a few others,^{30,51,Siettos et al.} towards understanding the importance of the architecture and hardwiring in a network exposed to an epidemic outbreak. It has clear limitations, but also a wide potential for extensions (to other disease dynamics, other population networks or even more general dynamic networks), and for more elaborate analyses of the underlying mathematical phenomena. Some of these aspects are discussed below.

One limitation introduced conceptually in the model relates to the strict assumptions made on Ebola spread and immunity. For example, we worked from the start under the hypothesis that the individuals who had contracted Ebola once, cannot have the disease again. This was based on a corresponding assumption in our original reference paper,⁴ which in turn was supported by the lack of evidence of any individuals with more than one Ebola infection within their life span. This idea is currently considered controversial in the disease dynamics literature, especially with the known variety of Ebola virus strains which may make a prior infection with one strain irrelevant immunologically to a new infection with a different strain. Immunity built-up aspects, such as duration and effectiveness to multiple viral strains, have been also explored mathematically.³⁹

Another simplifying assumption we made was that infected people can no longer spread the virus after the infection clears, whether this occurs through recovery or death of the individual. In reality, this may not be accurate. While some studies show that the risk of transmission from bodily fluids of convalescent patients is low (when infection control guidelines for the viral hemorrhagic fevers are followed⁸), other studies have shown that the Ebola virus can be spread through the sperm of recovered individuals for up to six months after infection.¹⁴ It is also well known at this point that dead bodies can remain contagious for up to 60 days,² with the potential of infection spread through contact with the dead body (e.g., during ritual funerals¹¹). Further iterations of the model may consider introducing these effects into the coupled equations.

Finally, one important aspect to explore in future studies is the extent to which details of the network configuration

can flip the optimal quarantine circumstances. The current paper shows that the quarantine measures required for maximal control may differ with measures such as edge density, or travel likelihood. Other studies have explored the impact of the small-worldness, or assortativity of the network on the overall dynamics.³⁰ There is, however, the likely possibility that there is no canonically optimal response based only on global network measures, and that an adequate quarantine systems has to be customized in response to the local details of the network architecture. Returning to our original analogy between viral diffusion and brain dynamics – in the same fashion in which clinical neuroscience is evolving towards “brain profiling,” and personalized clinical assessments, in the same way the response to a global Ebola outbreak may have to consider a “population network profiling” in order to deliver optimal results.

Our small values for optimal quarantine time lags suggest that full preparedness for a global outbreak may involve having a pre-established plan of action, to avoid fatal computation and decision-related delays during the spreading of the outbreak. This would require constantly updated knowledge of local and global travel patterns, a dynamic “global connectivity map” that could be implemented directly, when necessary, into simulations, and produce immediate predictions and provide efficient choices for quarantines. The cost of maintaining online global information may be a well placed investment in view of a potential pandemic.

REFERENCES

1. Adegun, O. (2014) The effects of Ebola virus on the economy of West Africa through the trade channel. *IOSR Journal of Humanities and Social Science (IOSR-JHSS)*(19) 10, 48–56.
2. Ahmad, K. (2014) Ebola virus infection: An overview. *Journal of Islamabad Medical & Dental College (JIMDC)* 3(2), 76–78.
3. Anderson, R., May, R., and Anderson, B. (1992) *Infectious diseases of humans: dynamics and control* 28. Wiley Online Library.
4. Astacio, J., Briere, D., Guilléon, M., Martinez, J., Rodriguez, F., and Valenzuela-Campos, N. (1996) Mathematical models to study the outbreaks of Ebola. *Biometrics Unit Technical Report*, 1-18.
5. Bailey, N. (1975) *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.
6. Barthélemy, M., Barrat, A., Pastor-Satorras, R., and Vespignani, A. (2004) Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Physical Review Letters* 92(17), 178701.
7. Bassett, D. and Bullmore, E. (2006) Small-world brain networks. *The Neuroscientist* 12(6), 512–523.
8. Bausch, D., Towner, J., Dowell, S., Kaducu, F., Lukwiya, M., Sanchez, A., Nichol, S., Ksiazek, T., and Rollin, P. (2007) Assessment of the risk of Ebola virus transmission from bodily fluids and fomites. *Journal of Infectious Diseases* 196(Supplement 2), S142–S147.
9. Briand, S., Bertherat, E., Cox, P., Formenty, P., Kieny, M.-P., Myhre, J., Roth, C., Shindo, N., and Dye, C. (2014) The international Ebola emergency. *New England Journal of Medicine* 371(13), 1180–1183.
10. Bullmore, E. and Sporns, O. (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10(3), 186–198.
11. Chippaux, J.-P. (2014) Outbreaks of Ebola virus disease in Africa: the beginnings of a tragic saga. *J Venom Anim Toxins Incl Trop Dis* 20(1), 44.
12. Chowell, G., Hengartner, N., Castillo-Chavez, C., Fenimore, P., and Hyman, J. (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(1), 119–126.
13. Danon, L., Ford, A., House, T., Jewell, C., Keeling, M., Roberts, G., Ross, J., and Vernon, M. (2011) Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, 284909.
14. Deen, G., Knust, B., Broutet, N., Sesay, F., Formenty, P., Ross, C., Thorson, A., Massaquoi, T., Marrinan, J., Ervin, E., et al. (2015) Ebola RNA persistence in semen of Ebola virus disease survivors—a preliminary report.

New England Journal of Medicine.

15. Enns, E., Mounzer, J., and Brandeau, M. (2012) Optimal link removal for epidemic mitigation: A two-way partitioning approach. *Mathematical Biosciences* 235(2),138–147.
16. Erickson, B., Widener, A., et al. (2014) Preparedness for Ebola questioned. *Chemical & Engineering News* 92(42), 7–7.
17. Eubank, S., Guclu, H., Kumar, V., Marathe, M., Srinivasan, A., Toroczkai, Z., and Wang, N. (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* 429(6988),180–184.
18. Feldmann, H. (2014) Ebola—a growing threat? *New England Journal of Medicine* 371(15), 1375–1378.
19. Frieden, T., Damon, I., Bell, B., Kenyon, T., and Nichol, S. (2014) Ebola 2014 – new challenges, new global response and responsibility. *New England Journal of Medicine* 371(13), 1177–1180.
20. Gostin, L., Lucey, D., and Phelan, A. (2014) The Ebola epidemic: a global health emergency. *JAMA* 312(11), 1095–1096.
21. Gross, T., D’Lima, C., and Blasius, B. (2006) Epidemic dynamics on an adaptive network. *Physical Review Letters* 96(20), 208701.
22. Haas, C. (2014) On the quarantine period for Ebola virus. *PLoS Currents* 6.
23. Heesterbeek, J. (2000) *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation* 5. John Wiley & Sons.
24. Hewlett, B. S. and Hewlett, B. L. (2007) *Ebola, culture and politics: the anthropology of an emerging disease*. Cengage Learning.
25. Hu, Z., Bi, P., Ma, W., Ruan, S., et al. (2011) Bifurcations of an SIRS epidemic model with nonlinear incidence rate. *Discrete and Continuous Dynamical Systems B* 15(1), 93–112.
26. Keeling, M. and Eames, K. (2005) Networks and epidemic models. *Journal of the Royal Society Interface* 2(4), 295–307.
27. Kenah, E. and Miller, J. (2011) Epidemic percolation networks, epidemic outcomes, and interventions. *Interdisciplinary Perspectives on Infectious Diseases [P]*, 1–13.
28. Kenah, E. and Robins, J. (2007) Network-based analysis of stochastic SIR epidemic models with random and proportionate mixing. *Journal of Theoretical Biology* 249(4), 706–722.
29. Kiskowski, M. (2013) A three-scale network model for the early growth dynamics of 2014 West Africa Ebola epidemic. *PLoS currents* 6.
30. Kopman, S., Akbas, M., and Turgut, D. (2012) Epidemicsim: Epidemic simulation system with realistic mobility. In *Local Computer Networks Workshops (LCN Workshops), IEEE 37th Conference on*, 659–665.
31. Labouba, I. and Leroy, E. (2015) Ebola outbreaks in 2014. *Journal of Clinical Virology* 64, 109–110.
32. Legrand, J., Grais, R., Boelle, P.-Y., Valleron, A.-J., and Flahault, A. (2007) Understanding the dynamics of Ebola epidemics. *Epidemiology and Infection* 135(4), 610–621.
33. Lekone, P. and Finkenstädt, B. (2006) Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics* 62(4), 1170–1177.
34. Lewnard, J., Mbah, M., Alfaro-Murillo, J., Altice, F., Bawo, L., Nyenswah, T., and Galvani, A. (2014) Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis. *The Lancet Infectious Diseases* 14(12), 1189–1195.
35. Marceau, V., Noël, P.-A., Hébert-Dufresne, L., Allard, A., and Dubé, L. (2011) Modeling the dynamical interaction between epidemics on overlay networks. *Physical Review E* 84(2), 026105.
36. Moreno, Y., Pastor-Satorras, R., and Vespignani, A. (2002) Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems* 26(4), 521–529.
37. Newman, M. (2003) The structure and function of complex networks. *SIAM Review* 45(2),167–256.
38. Pandey, A., Atkins, K., Medlock, J., Wenzel, N., Townsend, J., Childs, J., Nyenswah, T., Ndeffo-Mbah, M., and Galvani, A. (2014) Strategies for containing Ebola in West Africa. *Science* 346(6212), 991–995.
39. Piazza, N. and Wang, H. (2013) Bifurcation and sensitivity analysis of immunity duration in an epidemic

- model. *Int. J. Numer. Anal. And Modeling Series B Computing and Information* 4, 179–202.
40. Read, J., Eames, K., and Edmunds, W. (2008) Dynamic social networks and the implications for the spread of infectious disease. *Journal of The Royal Society Interface* 5(26),1001–1007.
 41. Richards, P., Amara, J., Ferme, M., Kamara, P., Mokuwa, E., Sheriff, A., Suluku, R., and Voors, M. (2015) Social pathways for Ebola virus disease in rural Sierra Leone, and some implications for containment. *PLoS Negl Trop Dis* 9(4), e0003567.
 42. Rivers, C., Lofgren, E., Marathe, M., Eubank, S., and Lewis, B. (2014) Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia. *PLoS currents* 6.
 43. Rădulescu, A. and Verduzco-Flores, S. (2015) Nonlinear network dynamics under perturbations of the underlying graph. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25(1), 013116.
 44. Safi, M., Imran, M., and Gumel, A. (2012) Threshold dynamics of a non-autonomous SEIRS model with quarantine and isolation. *Theory in Biosciences* 131(1),19–30.
 45. Satsuma, J., Willox, R., Ramani, A., Grammaticos, B., and Carstea, A. (2004) Extending the SIR epidemic model. *Physica A: Statistical Mechanics and its Applications* 336(3), 369–375.
 46. Schieffelin, J., Shaffer, J., Goba, A., Gbakie, M., Gire, S., Colubri, A., Sealfon, R., Kanneh, L., Moigboi, A., Momoh, M., et al. (2014) Clinical illness and outcomes in patients with Ebola in Sierra Leone. *New England Journal of Medicine* 371(22), 2092–2100.
- Siettos et al.. Siettos, C., Anastassopoulou, C., Russo, L., Grigoras, C., and Mylonakis, E. (2015) Modeling the 2014 Ebola virus epidemic—agent-based simulations, temporal analysis and future predictions for Liberia and Sierra Leone. *PLoS Currents* 7.
48. Smieszek, T., Fiebig, L., and Scholz, R. (2009) Models of epidemics: when contact repetition and clustering should be included. *Theoretical Biology and Medical Modelling* 6(1), 11.
 49. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Colizza, V., Isella, L., Régis, C., Pinton, J.-F., Khanafer, N., Van den Broeck, W., et al. (2011) Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine* 9(1), 87.
 50. Ströher, U. and Feldmann, H. (2006) Progress towards the treatment of Ebola haemorrhagic fever. *Expert Opinion on Investigational Drugs* 15(12), 1523–1535.
 51. Valdano, E., Ferreri, L., Poletto, C., and Colizza, V. (2015) Analytical computation of the epidemic threshold on temporal networks. *Physical Review X* 5(2), 021005.
 52. van den Heuvel, M. and Sporns, O. (2011) Rich-club organization of the human connectome. *The Journal of Neuroscience* 31(44), 15775–15786.
 53. Zelazo, D. and Mesbahi, M. (2008) On the observability properties of homogeneous and heterogeneous networked dynamic systems. In *47th IEEE Conference on Decision and Control*, 2997–3002.

ABOUT THE STUDENT AUTHOR

Joanna Herron is currently a senior at SUNY New Paltz, pursuing a Bachelor of Science in mathematics and psychology. She has had the opportunity to help conduct a variety of research in areas from game studies to computational neuroscience during her undergraduate career, and hopes to work in a behavioral laboratory after graduation.

PRESS SUMMARY

We define the impact and the duration of a viral outbreak in a network model of an infected population. We study the dependence of these measures on local and global network connectivity, and on the travel likelihood within the network. We perform simulations of the outbreak course, first in a network with fixed travel rates, then in an adaptable network with two connected hubs. In particular, we study the effects that quarantine can have on the impact and duration, for different possibilities of network hardwiring. We find that quarantine options need to be specific in order to alleviate outbreak effects.

Analysis of Vortex Pool-and-Chute Fishway

Mathew Nyberg, Brian Draeger, Brian Weekly, Eileen Cashman*, Michael Love

Department of Environmental Resources Engineering, Humboldt State University, Arcata, CA

Students: mjn190@humboldt.edu, bd539@humboldt.edu, brw167@humboldt.edu

Mentors: eileen.cashman@humboldt.edu*, mlove@h2odesigns.com

ABSTRACT

Fishways are constructed in riverine habitats where structures such as culverts, dams, and flood channels have negatively impacted flow conditions suitable for the movement of native and migratory fish species. These auxiliary channels are engineered to resist gravitational force with frictional force, resulting in sustained depth and reduced velocity over a range of design flow rates. The Chézy hydraulic resistance coefficient accounts for such forces and provides a metric useful for determining the effectiveness of a fishway to alter flow conditions prohibitive to the passage of fish. The objective of this analysis was to use a scale model of an innovative vortex pool-and-chute fishway, that operates with both plunging and streaming flow simultaneously, designed by Michael Love and Associates, to determine the Chézy resistance coefficients over a range of flow rates under controlled hydraulic conditions. Using dimensional analysis to ensure proper scaling allowed laboratory measurements of the model to be translated into a real-world prototype design. The conceptual prototype fishway is a 144-foot-long by 30-foot-wide channel with an 8% slope. A 1:15 scale model was constructed to evaluate the design at prototype equivalent flow rates between 58 and 283 cubic feet per second (cfs). Chézy coefficients were estimated by two different calculation methods; the *streaming flow* method and the *streaming and plunging flow* method. Coefficients ranging between 22.3 and 39.2 $\text{ft}^{1/2}/\text{s}$ were determined by the *streaming flow* calculation method, whereas the *streaming and plunging flow* calculation method yielded estimates from 18.9 to 25.0 $\text{ft}^{1/2}/\text{s}$ at corresponding flow rates. For flows that were exclusively *plunging*, values of 32.2 to 41.9 $\text{ft}^{1/2}/\text{s}$ were found. In general, Chézy coefficient estimates were observed to decrease with increasing discharge and values were found to be comparable to those calculated for fishways implemented at similar slopes. The preliminary model fishway results indicated that implementation of a prototype fishway could effectively alter flows for adequate fish passage under the given conditions.

KEYWORDS

Hydraulics; Fish Passage; Fishway; Chézy Coefficient; Geometric Scaling; Froude Scaling; Streaming Flow; Plunging Flow; Dimensional Analysis; Similitude

INTRODUCTION

The development of man-made structures within natural waterways has resulted in the decline of native fish species due to the ensuing unfavorable flow conditions. Consequently, engineers have designed secondary structures to create environments suitable for the movement and migration of fish within affected stream areas. Several types of fish passage structures have been developed, including fish elevators and various fish ladder designs. The objective of this project was to provide an analysis of a vortex pool-and-chute (VPC) fish ladder designed by Michael Love and Associates (MLA) proposed for Alameda Creek in Alameda County, California. Due to urbanization in the Alameda Creek watershed, channelization and man-made barriers designed for flood control have made the creek prohibitive for native steelhead trout migration. The fishway analyzed is a scaled model of the prototype designed by MLA. The prototype design is a 144-foot-long by 30-foot-wide VPC fishway, positioned on an 8% slope. Eleven chutes with vortex weir shoulders, chamfered on the downstream edge to minimize adhesion, are evenly distributed along the fishway with the goal of adequately accommodating fish passage over a range of flow rates between 20 cfs and 200 cfs. To ensure similitude, geometric and Froude scaling were conducted to convert geometric and flow parameters between the model and prototype fishways. The scale model of the prototype was built in The Humboldt State University (HSU) Environmental Resources Engineering Construction Laboratory. The fishway was constructed using half-inch thick Plexiglas® due to the similar roughness characteristics of the concrete prototype.

The HSU sediment transport flume was utilized to evaluate the scale model's performance over a range of discharges. The flume is equipped with two pumps capable of supplying up to 700 gallons per minute (gpm) to a two-foot-deep, two-and-a-half-foot-wide by forty-foot-long channel. Fluid enters the channel and supply line gate-valves enable regulated flow. A mounted point depth gauge with latitudinal and longitudinal adjustments provides for three-dimensional measurement within the channel. A hinged mount at the flume head and pneumatic jack at the tail of the channel allow for slope modification. Together these flume characteristics enable collection of data under controlled hydraulic conditions.

While it is essential that the fishway reduces flow velocity and maintains adequate depth for fish passage, it is also important that the flow rate is sufficient to attract fish to the fishway entrance. An analysis of the collected data was carried out to estimate the Chézy coefficient of the fishway at various flow rates. The Chézy coefficient is a measure of resistance to flow and is a function of gravity and the roughness of the material over which uniform flow occurs in an open channel. With a known Chézy coefficient and fishway geometry, flow rates and associated velocities may be estimated for the fishway. Results from the scaled model analysis will assist MLA in determining whether the fishway will allow sufficient flow to attract fish while providing adequate resistance to create a favorable environment for the successful passage of native migratory fish species.

Fishway Basics

The standard pool-and-weir is the basic fishway design, where stream flow is altered by channel spanning weirs that effectively divide the stream section into pools. Depending on the magnitude of discharge, one of two flow regimes exists: plunging or streaming. These regimes are characterized by the manner in which the water circulates within the pools. At lower rates, flow plunges over the weir crests into the downstream pool and follows a path downward to the channel bottom, along the bed in the downstream direction, and then rises at the upstream side of the subsequent weir face. A portion of the flow rises and spills over the weir crest into the next pool while some of the flow travels back upstream, causing a circulation pattern that is distinct from the streaming flow regime.¹ The described circulation pattern for the plunging flow regime is illustrated by the flow lines shown in **Figure 1**. Given that pool sizes are sufficient to dissipate turbulence caused by plunging flows, conditions in this regime are favorable for fish passage. The inclusion of orifices at the bottom of the weirs further supports this plunging circulation pattern and can extend the range of flows over which the plunging regime persists.¹

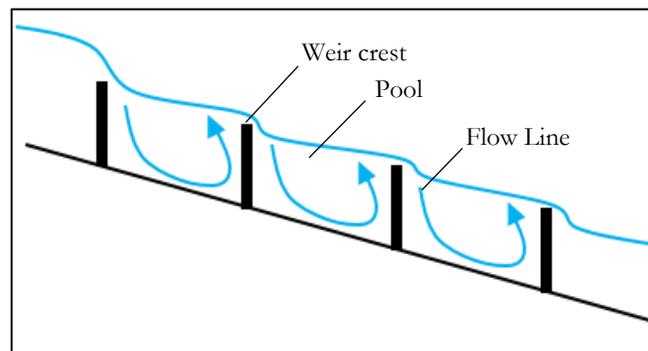


Figure 1. Plunging flow regime (Draeger 2015).

Higher flow rates result in streaming flow conditions, characterized by higher velocity flows that skim over the pool water surfaces and above the weir crests. A circulation pattern in the opposite direction of that described for plunging flow results due to the effects of shear stresses propagating through the water. The streaming flow regime circulation pattern is shown by the flow lines in **Figure 2** and is not conducive to fish passage.

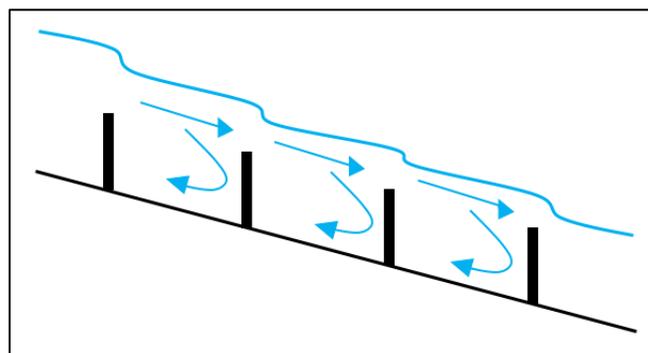


Figure 2. Streaming flow regime (Draeger 2015).

The relative elevations of the weir crests and the downstream pool water surface distinguish the difference between plunging flow and streaming flow. In plunging flow conditions, the downstream pool water surface is below the upstream weir crest.² When the downstream pool elevation surpasses the weir height, the influence of the weir diminishes, ultimately leading to the change in circulation pattern where streaming flow begins. Higher flow rates result in streaming flow conditions in which the downstream pool water surface elevation is higher than the upstream weir crest.² Once streaming flow begins in the pool-and-weir design it occurs across the entire width of the fishway, resulting in velocities and hydraulics not suitable for fish passage.

Pool-and-Chute Fishways

Pool and chute style fishways are commonly used for passage of steelhead trout and rely on the leaping ability of the fish to ascend the fishway. These types of fishways are typically designed for a specific range of flows and to meet the following objectives: 1) to create a fishway entry that is suitable for fish attraction, 2) to establish small water surface variances between neighboring fishway pools, 3) to develop sufficient pool volumes based on proper geometries to disperse energy and permit fish resting, 4) to establish plunging flows near the fishway walls, 5) to satisfy minimum water depth constraints for fish jumping, and 6) to satisfy maximum water velocity constraints in the fishway to ensure passage.³ A pool-and-chute fishway, as shown in **Figure 3** has a notched “chute” in the center and shoulder weirs. They operate with streaming flow down the center of the fishway, in the chute, and plunging flows along the edges of the wetted shoulder weirs. A fish passage corridor is provided along the edges of the fishway, where flows are plunging, while the majority of the flow in the fishway is streaming down the center. This allows the fish to have a wider range of operating flows, and the higher velocities of the streaming flows discharging from the fishway assist fish in locating the fishway entrance. Orifices may be included near the bottom of the structure and can positively affect the hydraulics of each successive downstream pool. Properly sized orifices, with a combined capacity that does not exceed the low flow of a stream, can provide fish passage throughout the dry season.³ These orifices also provide a route for passage of lamprey and other non-leaping fish species, and affect circulation patterns within the pool that improve sediment transport capacity through the structure.⁴



Figure 3. Downstream view of a partial-spanning pool-and-chute fishway in San Anselmo Creek, Marin County, California.⁵

Vortex Pool-and-Chute Fishway Design

The VPC fishway designed by Michael Love is a relatively distinct design. The vortex weirs are not only V-shaped downward to the channel center, but are also V-shaped longitudinally, with the apexes located in the upstream direction. The current design being analyzed in this document incorporates the pool-and-chute design and includes orifices within the vortex weirs (**Figure 4** and **Figure 5**).

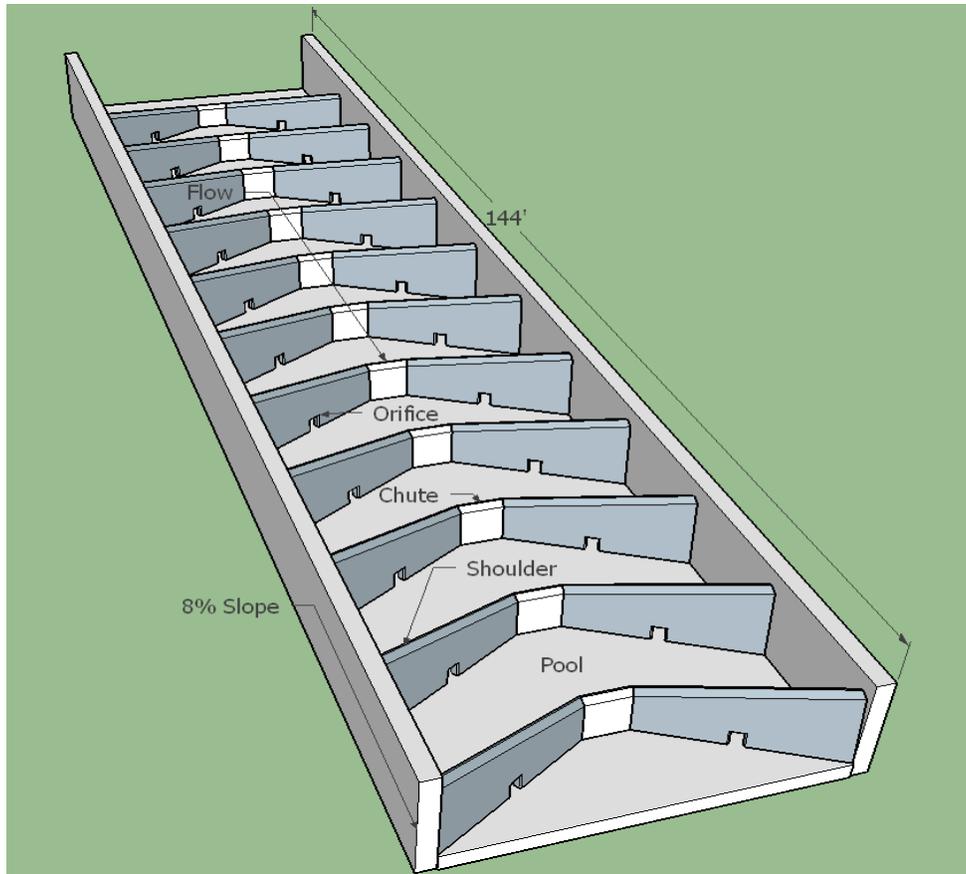


Figure 4. Downstream view of the Vortex Weir Pool-and-Chute Fishway Design prototype (Love and Associates 2015).

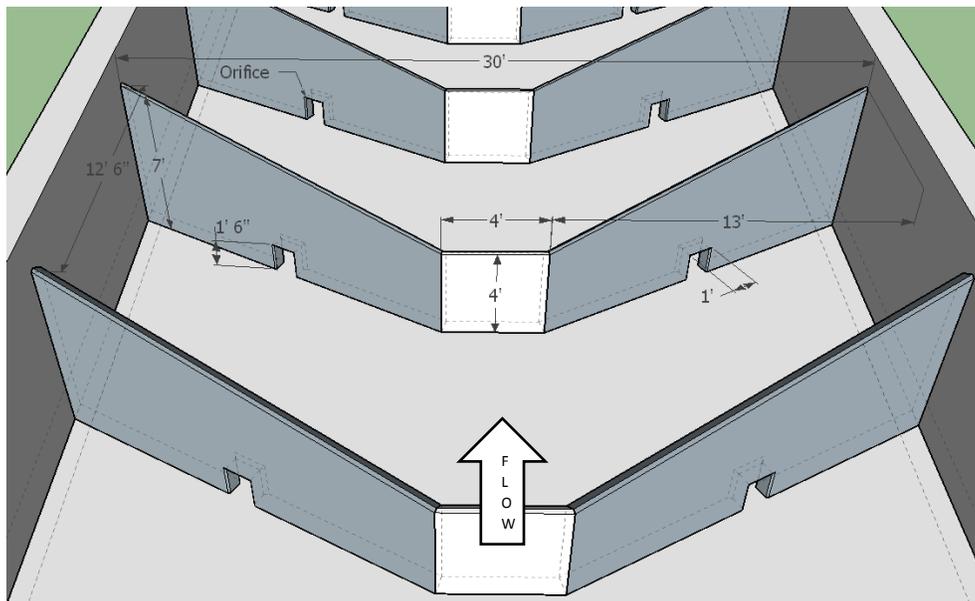


Figure 5. Section dimensions of the Vortex Weir Pool-and-Chute Fishway Design prototype (Love and Associates 2015).

In general, the vortex weir configuration serves to provide better flow conditions for fish passage relative to straight weirs by directing flow towards the channel center, thus reducing turbulence along the edges of the corridor. The VPC fish passage design allows for conveyance under both streaming and plunging flow regimes which may occur simultaneously across the fishway.⁶ At lower discharges, the design permits fish passage under plunging flow conditions.⁷ Flows are concentrated within the center of the

fishway to promote scour, which effectively maintains pool volumes and dissipates flow energies while maintaining a slower and less turbulent corridor, along the margins of the fishway, suitable for fish passage. At higher discharges, steaming flow is focused to the center of the fishway while the V-shaped weirs reduce velocity by extending crest lengths and maintaining plunging zones along the flow boundaries.⁷ Fish are able to pass through the lower velocity plunging flows that occur near the sidewalls of the structure.¹ These zones provide jump heights of one foot or less, as required by the National Marine Fisheries Service design criteria for the target species (steelhead trout).⁸

Scaling

The geometric similarity between a model and a prototype must be preserved to determine potential prototype performance. All dimensions of the model and prototype must have an equivalent linear scale ratio in the x-, y-, and z-dimensions to be considered geometrically similar. Adhering to geometric similarity ensures that all angles within the prototype design are maintained in the scale model.⁹ The model fishway used in this analysis was constructed with a scaling ratio of 1:15 from the prototype design schematics to ensure geometric similarity. In addition to proper geometric scaling, Froude scaling was carried out so that the model and prototype had the same time scale ratio. This is referred to as kinematic similarity.⁹ Froude scaling provides a means to resolve the nonlinear relationship between model and prototype flow rates and velocities.

Uniform Flow and the Chézy Equation

When water depth remains constant throughout a specified reach of an open channel, the flow is classified as uniform flow. Uniform flow commonly occurs in prismatic channels of constant slope and cross-sectional area.¹⁰ In addition to a constant flow depth of y_n , velocity within the channel reach also remains constant. Considering equivalent velocities along the channel reach, the velocity heads drop out of the energy equation.⁹

$$h_f = z_1 - z_2 = S_0 L. \tag{Equation 1}$$

Therefore, the headloss (h_f) over the channel reach is equal to the loss in channel elevation ($z_1 - z_2$) or slope (S_0) times the length (L). Headloss is also given by the Darcy-Weisbach equation:⁹

$$h_f = f \frac{L}{4R_h} \frac{V_0^2}{2g}, \tag{Equation 2}$$

where:

- f = friction factor
- R_h = hydraulic radius (ft)
- g = gravitational acceleration (ft²/s).

By setting **Equation 1** and **Equation 2** equal, the following relationship results for uniform flow channel velocity.⁹

$$V_0 = \left(\frac{8g}{f}\right)^{1/2} R_h^{1/2} S_0^{1/2}. \tag{Equation 3}$$

Under uniform flow conditions with a constant water depth, the value $\left(\frac{8g}{f}\right)^{1/2}$ is constant for a given channel shape and roughness⁹ and is often denoted as C , the Chézy resistance coefficient.¹⁰

The following equations are known as the Chézy formulas:⁹

$$V_0 = C(R_h S_0)^{1/2}, \tag{Equation 4}$$

$$Q = CA(R_h S_0)^{1/2}, \tag{Equation 5}$$

where A = area of flow in the cross section (ft²). Given the channel geometry and flow rate, the Chézy coefficient can be solved as⁹

$$C = \frac{Q}{A(R_h S_0)^{1/2}}. \tag{Equation 6}$$

METHODOLOGY

To determine the Chézy coefficient for the VPC fishway, a number of flow calculations were required. The water depth above the weir at which flow transitions from plunging to streaming was estimated so that a distribution of the two regimes could be estimated for a given flow rate. The assumption was made that shallower depths along the weir crest were plunging. Based on empirical data from previous work, the following is a step-by-step set of instructions used to calculate the Chézy coefficient at a given flow rate.¹¹ With an established Chézy coefficient, the mean velocity at various flow rates was predicted for the prototype fishway.

Plunging-Streaming Transition

To estimate the depth over the weir where flow initially transitions from the plunging regime to the streaming regime (H_{PT}), the transition flow rate must be calculated. This is done by first using a dimensionless relationship to approximate the flow over the horizontal section of the fishway weir where transition begins. A plot developed by Ead et al. and published in 2004 was used to determine the dimensionless flow rate based on the ratio of the on-center weir spacing and the weir crest height (L/p) (Figure 6).¹²

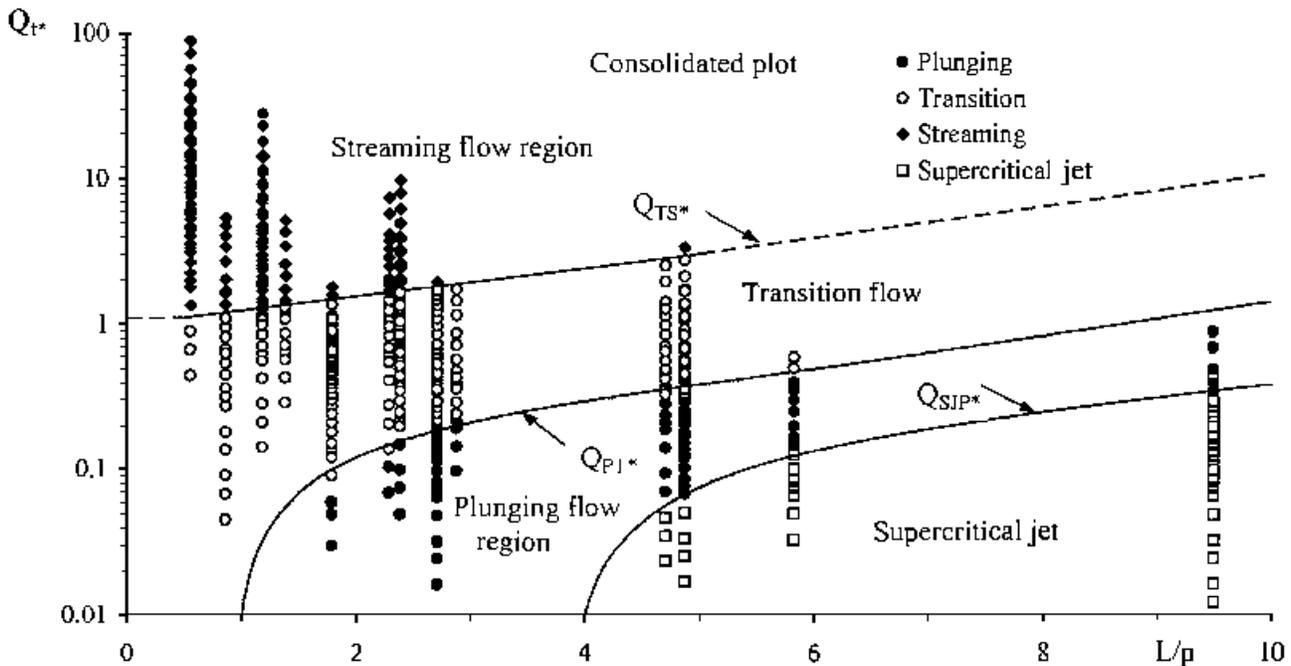


Figure 6. Relationship between Q_{t^*} and L/p in pool and weir fishways.¹²

The values of L and p for the scale model fishway are 0.833 ft and 0.267 ft, respectively. From the plot, the dimensionless plunging transition flow (Q_{t^*}) for the fishway being analyzed was determined to be 0.22 and remains constant regardless of flow rate. The transition flow over the horizontal weir (Q_{PT}) was then calculated to be ¹²

$$Q_{PT} = Q_{t^*} \sqrt{g} b_0 S_0 L^{3/2}, \tag{Equation 7.}$$

where:

- Q_{PT} = transition flow over horizontal weir crest (cfs)
- Q_{t^*} = dimensionless transition flow rate
- b_0 = horizontal weir width (ft)
- S_0 = fishway slope (ft/ft)
- L = the on-center spacing of the weirs (ft)
- g = gravitational acceleration (ft²/s).

The horizontal weir width (b_0) for the model fishway is 0.267 ft. The transition flow rate (Q_{PT}) for the model fishway was approximated to be 0.0203 cfs, and remains constant throughout this analysis. The depth of transition flow (H_{PT}) was then calculated over the chute, at the flow rate found with Equation 7, utilizing a horizontal weir equation (Equation 8).¹³ The bracketed portion of Equation 8 is a correction factor to account for submergence.¹⁴

The corrected weir equation was rearranged and H_{PT} was solved for using a numerical solver:

$$Q_{PT} = \frac{2}{3} C_d \sqrt{2g} b_0 H_{PT}^{3/2} \left[1 - \frac{(H_{PT}-h)^n}{H_{PT}} \right]^{0.385}, \tag{Equation 8}$$

where:

- C_d = weir coefficient ¹³ = $0.602 + 0.075 [H_{PT}/p]$
- H_{PT} = transition water depth over weir (ft)
- p = weir crest height = 0.267 (ft)
- h = water drop over each weir (ft)
- n = horizontal weir submergence correction exponent¹⁴ = 1.5.

The fishway was designed such that the drop over each weir (h) remains constant at 0.067 ft (1/15 ft). Since the transitional water depth (H_{PT}) is a function of only constants, this depth remains constant within the fishway regardless of the overall flow rate. The value of H_{PT} was determined to be 0.984 inch (0.082 ft) for the model fishway.

Orifice Flow

The fishway being analyzed includes orifices located at the bottom center of each shoulder weir; therefore, flow through these ports must be determined. This flow can be calculated using the standard submerged orifice equation:¹⁵

$$Q_{Orifice} = C_d A \sqrt{2gh}, \tag{Equation 9}$$

where:

- $Q_{Orifice}$ = the flow rate through each orifice (cfs)
- C_d = a coefficient used for the weir orifice
- A = the cross-sectional area of the weir orifice (ft²)
- h = water drops over each weir = 0.067 ft.

The common coefficient (C_d) value used for chamfered orifices such as those found in the model fishway is 0.61.¹¹ The cross-sectional area of each orifice in the model fishway was determined to be 5.56E-3 ft². The water surface drop across the weir was constant at 1/15 ft, so the flow rate through each orifice ($Q_{Orifice}$) was approximated to be 0.007 cfs and assumed constant regardless of the overall fishway flow rate.

Plunging Flow

The flow over the outer ends of the shoulder weirs, where depth is less than H_{PT} , is in the plunging regime. Utilizing the same submergence correction as used in **Equation 8**, a separate calculation was made using a standard sharp-crested triangular weir equation to approximate the plunging flow (Q_{PL}):¹⁶

$$Q_{PL} = \frac{8}{15} C_{dt} \sqrt{2g} \tan\left(\frac{\theta}{2}\right) H_{v-weir}^{2.5} \left[1 - \frac{(H_{PT}-h)^n}{H_{PT}} \right]^{0.385}, \tag{Equation 10}$$

where:

- Q_{PL} = plunging flow rate (cfs)
- C_{dt} = discharge coefficient for a triangular weir¹⁷
= $0.6072 - 0.000874(\theta) + 6.1 \times 10^{-6}(\theta)^2$
- θ = angle of the opening in the triangular weir (degrees)
- H_{v-weir} = upstream depth above the triangular weir crest (ft)
- H_{PT} = transition water depth over weir (ft)
- n = triangular weir submergence correction exponent = 2.5.

Streaming Flow

The overall flow rate within the fishway ($Q_{Fishway}$) was determined with the recorded stage in the flume head tank and the calibration curve provided for the HSU ERE sediment transport flume. The streaming flow rate (Q_{ST}) was considered to be the portion of the overall flow rate that is not travelling through the orifices nor within the plunging flows near the fishway walls.

The streaming flow rate was calculated as¹¹

$$Q_{ST} = Q_{Fishway} - Q_{PL} - 2Q_{Orifice}. \quad \text{Equation 11.}$$

Estimating the Chézy Coefficient (C)

The Chézy resistance coefficient was determined using two different methods over a range of flow rates. The first method used only the streaming portion of the overall flow to estimate the Chézy resistance coefficient. This method will be referred to as the *streaming flow* calculation method. The streaming flow rate (Q_{ST}) calculated by **Equation 11** was used in conjunction with the projected streaming flow hydraulic geometry over the chute and weir crest to calculate C_{ST} :¹¹

$$C_{ST} = \frac{Q_{ST}}{A_{ST}} \sqrt{\frac{P_{ST}}{A_{ST}S_0}}, \quad \text{Equation 12.}$$

where:

$$\begin{aligned} C_{ST} &= \text{Chézy coefficient calculated using } Q_{ST} \text{ (ft}^{1/2}/\text{s)} \\ A_{ST} &= \text{wetted area over the weir within the streaming flow region (ft}^2\text{)} \\ P_{ST} &= \text{wetted perimeter over the weir within the streaming flow region (ft)}. \end{aligned}$$

The second method used the sum of the streaming and plunging portions of the overall flow to estimate the Chézy resistance coefficient. This method will be referred to as the *streaming and plunging flow* calculation method. The sum of the streaming and plunging flow rates ($Q_{S/P}$) was determined as the overall flume flow rate minus the flow rate travelling through the two orifices at each weir. The projected hydraulic geometry of the entire wetted chute and weir were used in the Chézy equation to calculate¹¹

$$C_{S/P} = \frac{Q_{S/P}}{A_{S/P}} \sqrt{\frac{P_{S/P}}{A_{S/P}S_0}}, \quad \text{Equation 13.}$$

where:

$$\begin{aligned} C_{S/P} &= \text{Chézy coefficient calculated using } Q_{S/P} \text{ (ft}^{1/2}/\text{s)}, \\ Q_{S/P} &= \text{sum of } Q_{ST} \text{ and } Q_{PL} \text{ (cfs)}, \\ A_{S/P} &= \text{wetted area over the entire weir (ft}^2\text{)}, \\ P_{S/P} &= \text{wetted perimeter over the entire weir (ft)}. \end{aligned}$$

Experimental Methodology

The model fishway was designed with tapered legs attached to the bottom of the structure to provide an 8% slope. The flume head tank was filled to the desired stage height that corresponded to a flow rate determined from a calibration curve. Flume flow rates were selected based on Froude scaling of the design range of prototype flow rates. Once the flume head tank stage remained steady, the tail water height was adjusted such that the drop over the lowermost weir was similar to the drops over the upstream weirs. To develop a water surface profile and determine the Chézy coefficient, the x-, y-, and z-coordinates along the fishway were recorded. Starting at the furthest downstream weir, measurements were taken in the middle of the fishway over the center of the weir. Two measurements were recorded directly over the weir: the water surface elevation and the height of the weir crest. Moving upstream, these measurements were repeated for each weir. When the fourth weir was reached, additional water surface elevation measurements were taken along the centerline in the middle five fishway pools to create a water surface profile depicting the hydraulic jumps that occurred between weirs. Successive measurements continued until the third weir from the top of the fishway was reached. Intermittent measurements were not taken for the first three or final two weirs to eliminate tailwater and headwater influences on hydraulic modeling.

Measurements were then made on the left and right sides of the fishway at the wetted edge of the weir shoulders. Since there was no water depth at these points, only the heights of the weir crests were measured. These measurements provided the lateral boundary of flow necessary for calculating wetted perimeters and wetted areas. This process was repeated twelve times for nine separate flow rates. For three runs (Run #5, #8 and #11), clay was inserted into the weir orifices to determine how orifice obstruction would affect water surface profiles and Chézy coefficient estimates.

Estimating Wetted Perimeter and Wetted Area

The projection of plunging and streaming portions of flow traveling over each weir within the fishway and experimental measurement locations is illustrated in **Figure 7**.

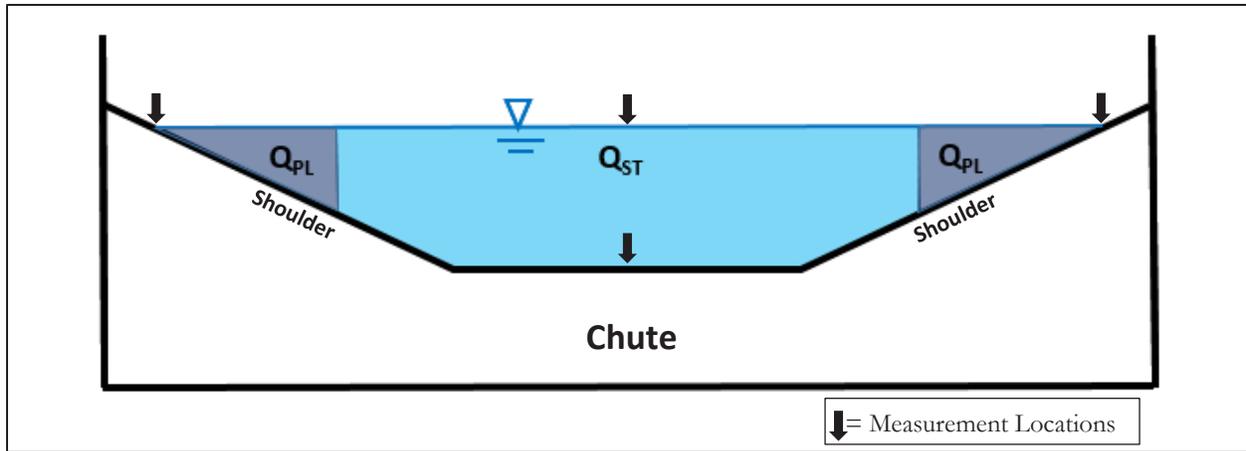


Figure 7. Projection of plunging (Q_{PL}) and streaming (Q_{ST}) portions of flow over weir crests with experimental measurement locations (Draeger, 2015).

This two-dimensional geometry was used to estimate wetted perimeters and wetted areas of flow. For the streaming flow method, calculations neglected the boundary between Q_{PL} and Q_{ST} as part of the streaming wetted perimeter. Streaming area (A_{ST}) was calculated based on the depth of flow measured directly over the center of the horizontal weir crest. The area was bounded laterally by the geometric location in the projection of H_{PT} , the depth at the theoretical boundary between Q_{PL} and Q_{ST} . Plunging area was estimated by subtracting A_{ST} from the total projected area, $A_{S/P}$. Chézy coefficient and subsequent velocity estimates were calculated using wetted perimeters and wetted areas determined in this manner.

ANALYSIS RESULTS AND DISCUSSION

The design range of flow rates for the prototype fishway is 20 to 200 cfs.⁴ Froude scaling of prototype flow rates was carried out to determine a range of flow rates to be used in the model fishway. A total of nine flow rates were examined, with three experimental runs analyzed with the VPC fishway orifices obstructed. Though it was found that the flume slope for run #9 had been inadvertently set at 7.73%, the data was included as the variation was considered small. For run #12, a flow rate exceeding the design range was used to determine the fishway performance under higher discharge conditions. The model and corresponding prototype flow rates used in this analysis are shown in **Table 1**.

Run #	Fishway Slope	Q_{Model} (cfs)	$Q_{Prototype}$ (cfs)
1	8.01%	0.068	58.8
2	8.01%	0.085	73.9
3	8.01%	0.103	89.8
4	8.02%	0.122	106.5
5 (Plugged Orifices)	8.02%	0.122	106.5
6	8.02%	0.137	119.8
7	8.02%	0.165	143.5
8 (Plugged Orifices)	8.02%	0.165	143.5
9	7.73%	0.208	181.2
10	8.01%	0.253	220.2
11 (Plugged Orifices)	8.02%	0.253	220.2
12	8.02%	0.325	283.1

Table 1: Fishway slopes along with model and prototype flow rates used in each of the 12 experimental runs.

The Chézy coefficient was estimated for each dataset using the entire projected cross-sectional flow area over the weirs, given by the *streaming and plunging flow* calculation method described in the methodology section. Additionally, the Chézy coefficients were estimated using the *streaming flow* calculation method, described in the methodology, for runs with sufficient H_{PT} depths (runs #4 through #12). Results of these analyses are presented and discussed in the following subsections.

Chézy Estimates: Streaming Flow Calculation Method

For runs #4 through #12, prototype flow rates ranged from 106.5 to 283.1 cfs. Corresponding Chézy coefficient estimates calculated by the *streaming flow* method varied from 39.2 to 22.3 $ft^{1/2}/s$, respectively. Chézy coefficient estimates were plotted for graphical presentation in **Figure 8**.

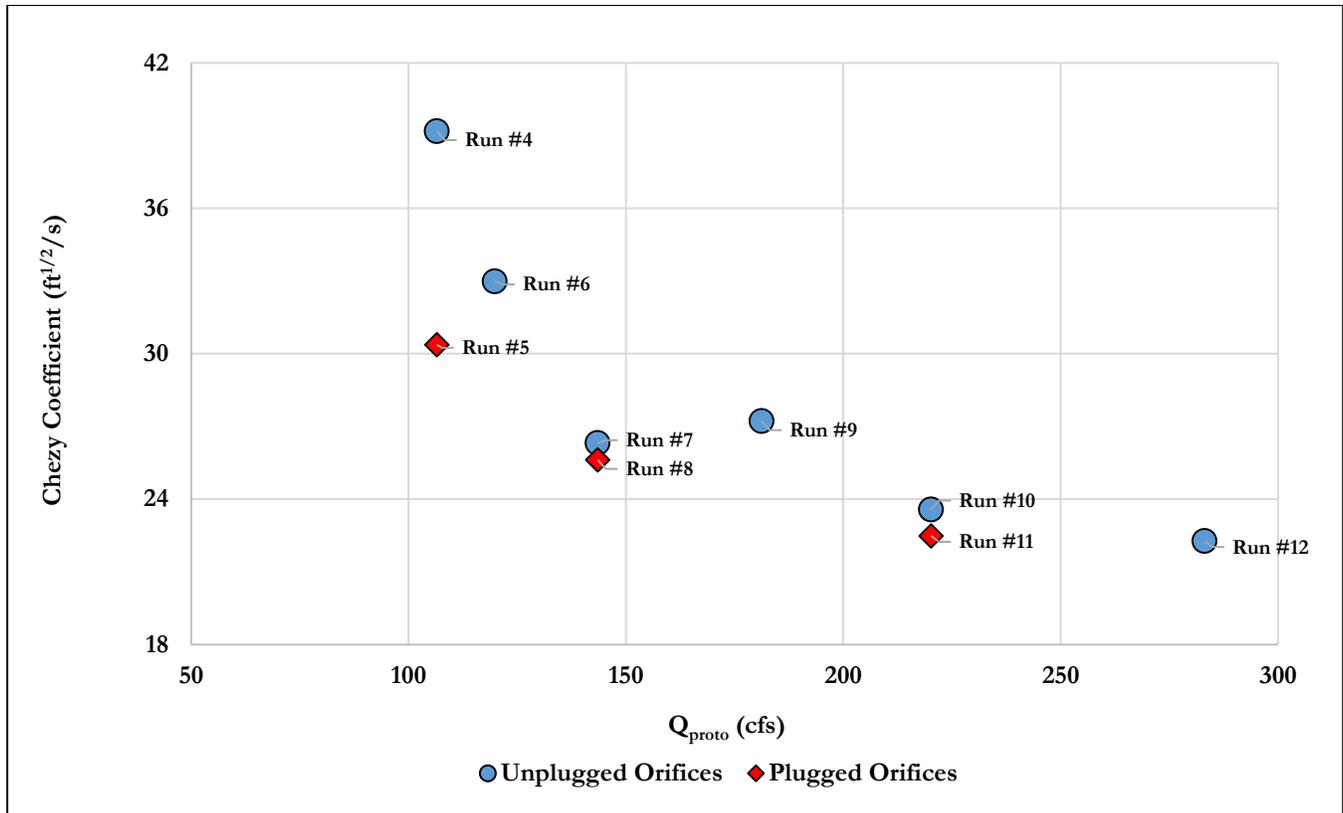


Figure 8. Chézy coefficient estimates using the streaming flow calculation method.

The most apparent observation to be made from the plot shown above is that as flow rates increase, flow resistance also increases and the resulting Chézy coefficient values decrease in magnitude (Figure 8). Also seen is that for equivalent flows, in general, the Chézy approximation is lower when the fishway orifices are plugged. Each of these outcomes suggests that the influence of the weirs increases as the water depth rises. Therefore, the fishway contributes more resistance as discharge increases when both streaming and plunging flow regimes exist.

An unexpected result from this portion of the analysis is the large Chézy coefficient value of $39.2 \text{ ft}^{1/2}/\text{s}$ determined for Run #4, with a prototype flow rate of 106.5 cfs. A Chézy coefficient value of $30.4 \text{ ft}^{1/2}/\text{s}$ was estimated at the same flow rate with plugged orifices, resulting in a 29% difference between the two Chézy coefficients. Conversely, changes of only 2-3% were observed between plugged and unplugged Chézy coefficient estimates for the 143.5 cfs and 206.3 cfs flow rates.

Despite careful review of measurement data entry and calculations, no conclusive justification was established for the Run #4 Chézy coefficient estimates to be considered erroneous. The second set of measurements collected at the 106.5 cfs flow rate yielded similar results. It is noteworthy to point out that at this flow, the average water surface depth over the horizontal weir sections was 0.086 ft, compared to the theoretically calculated transition depth (H_{PT}) of 0.082 ft. The authors suspect that the narrow margin between the theoretical H_{PT} value and the average measured flow depth over the horizontal weirs contributed to the extreme Chézy coefficient estimate for run #4 using the *streaming flow* calculation method. It is also noted that the theoretical H_{PT} value was calculated using Equations 15 and 16, assuming transition flow to occur over horizontal weirs. Furthermore, since run #4 corresponded to the lowest rate with streaming flow; and thus, lowest water depth, error in depth measurements may have had greater influence on the resulting Chézy coefficient estimates. All other Chézy coefficients estimated using the *streaming flow* calculation method fell within the range of 22–30 $\text{ft}^{1/2}/\text{s}$ found in literature for fishway slopes between 4.6 % and 11.1%.¹¹

Chézy Estimates: Streaming and Plunging Flow Calculation Method

Chézy coefficients determined by the *streaming and plunging flow* calculation method ranged from 18.9 to 41.9 $\text{ft}^{1/2}/\text{s}$. The results are shown for visual presentation.

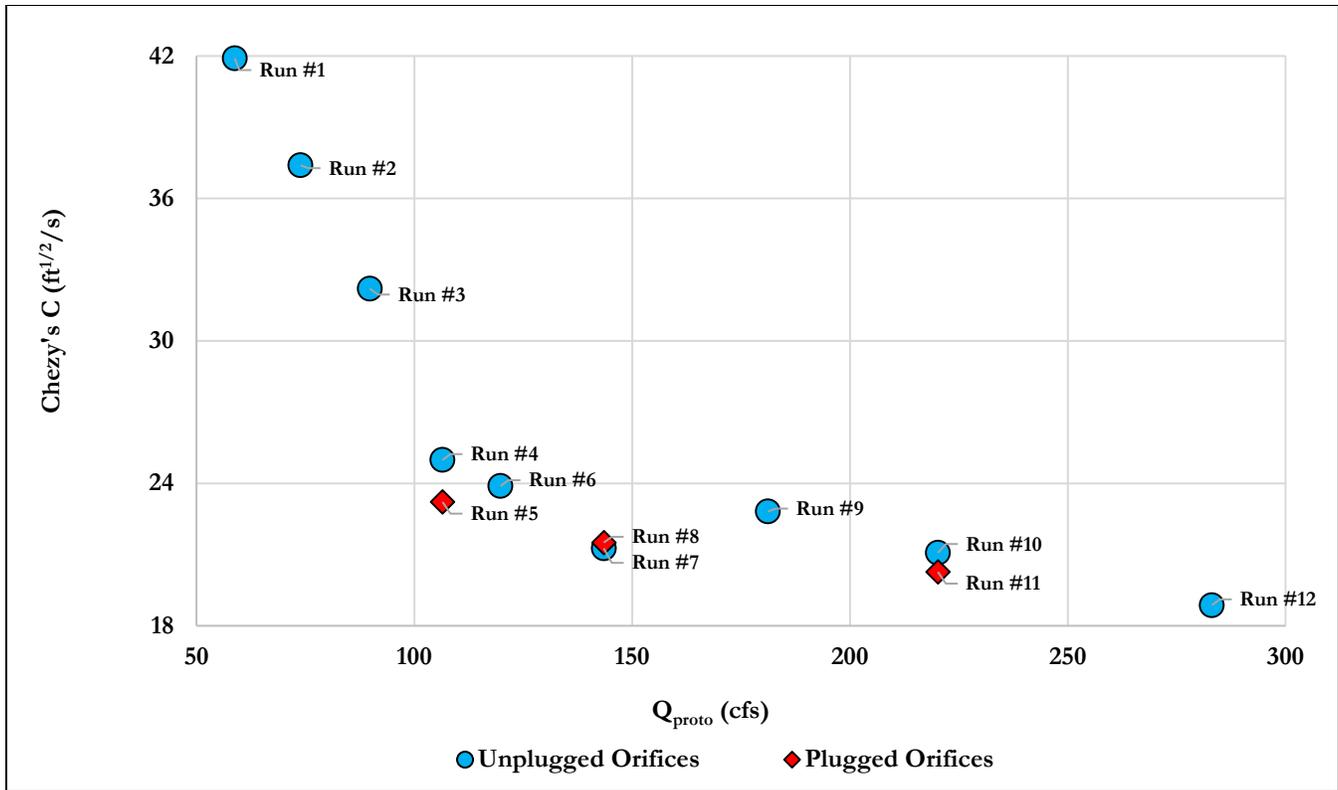


Figure 9. Chézy coefficient estimates using streaming and plunging flow calculation method.

As with the previous calculation method, a general trend is observed in which Chézy estimates decrease with increasing flow rate (Figure 8 and Figure 9). Figure 9 shows the Chézy coefficient decreases much more rapidly in response to increased plunging flows, before streaming flow begins. A fairly linear relationship appears to exist between Chézy estimates and flow rate with an abrupt transition zone as streaming flow begins at approximately 105 cfs. The most noticeable difference between the estimates shown in Figure 9 and those found using only the streaming cross-sections is that there is far less variance within corresponding flow rates, largely due to the outlier in the preceding set. Chézy coefficients were found to be systematically lower when estimated using the *streaming and plunging flow* calculation method versus those computed by the *streaming flow* method. The *streaming and plunging flow* calculation method does not require approximations of Q_{PT} , Q_{PL} , and H_{PT} to estimate Chézy coefficients. Neglecting these theoretical values in the calculations may have contributed to the lower variance between estimates, and the lower Chézy coefficient estimates in general.

The Chézy estimates for flow rates where plugged versus unplugged orifices were compared only vary by 1–7%. As with *streaming flow* calculations, the largest difference was observed at the lowest flow rate. However, considering the relatively small change in flow that resulted, 7% is a reasonable difference relative to 29%, supporting suspicions of the validity of the 106.5 cfs *streaming flow* calculation.

Water Surface Profiles

Water surface elevations were collected at six intermediate locations along the centerline of the pools between the middle six weirs. With elevations measured directly over the center of each weir, information was provided about the change in elevation for consecutive pools, as well as the location, magnitude, and distribution of hydraulic jumps. Visual representations of unplugged orifice profiles for prototype equivalent flow rates of 106.5, 143.5, 181.2 and 283.1 cfs are shown below.

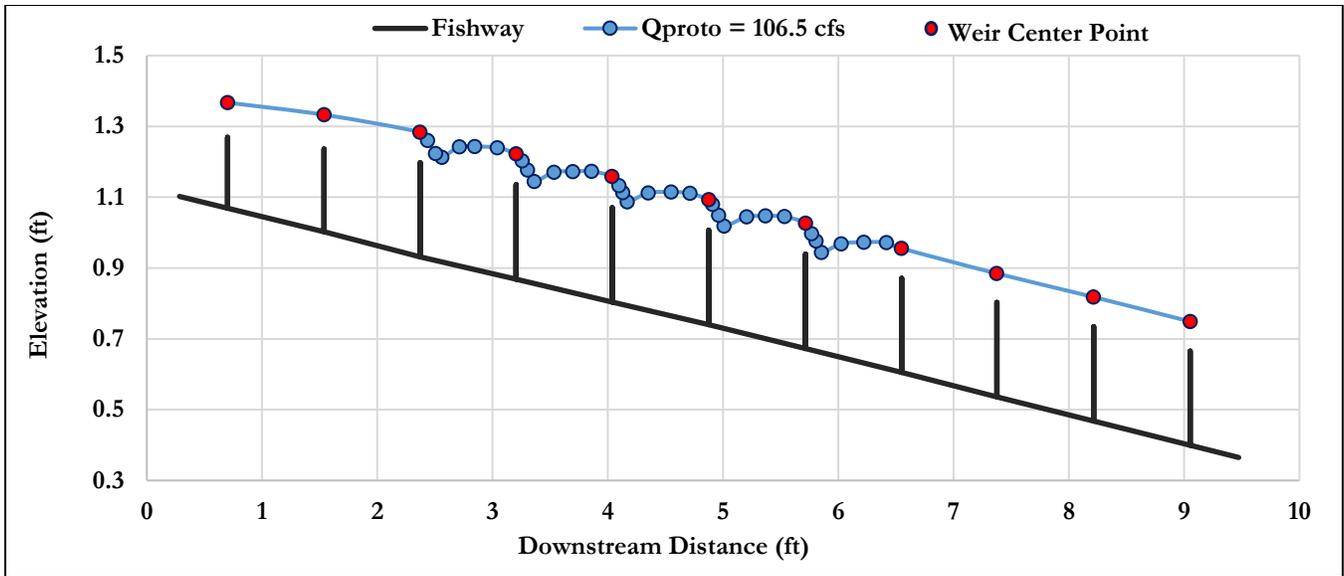


Figure 10. Water surface profile for 106.5 cfs prototype equivalent.

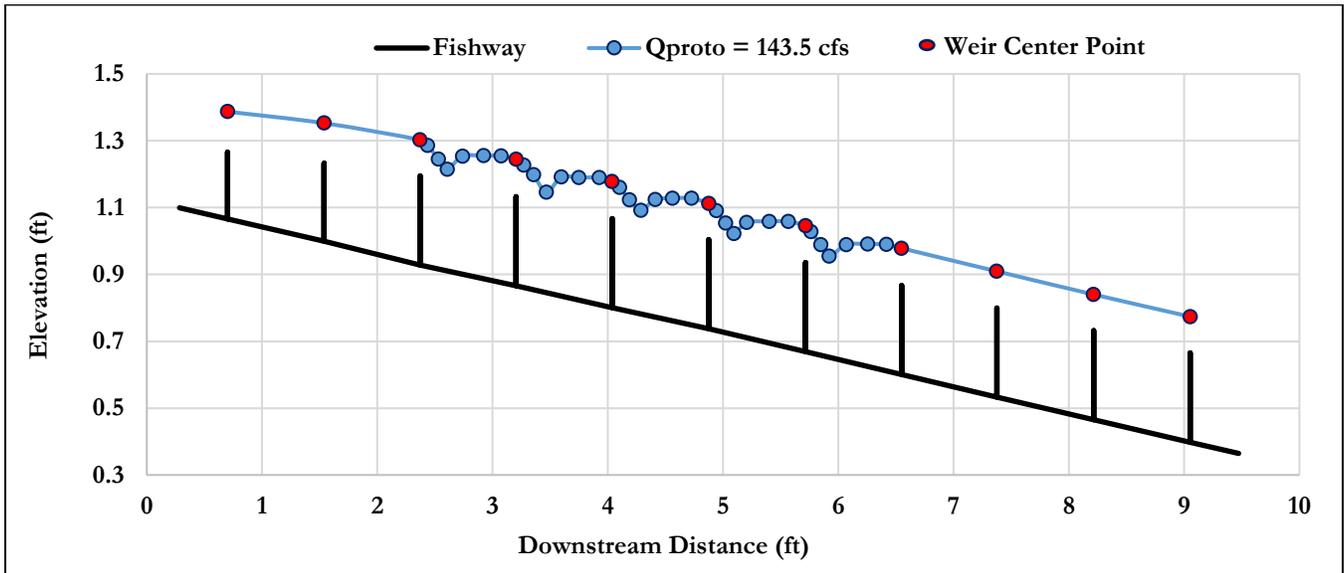


Figure 11. Water surface profile for 143.5 cfs prototype equivalent.

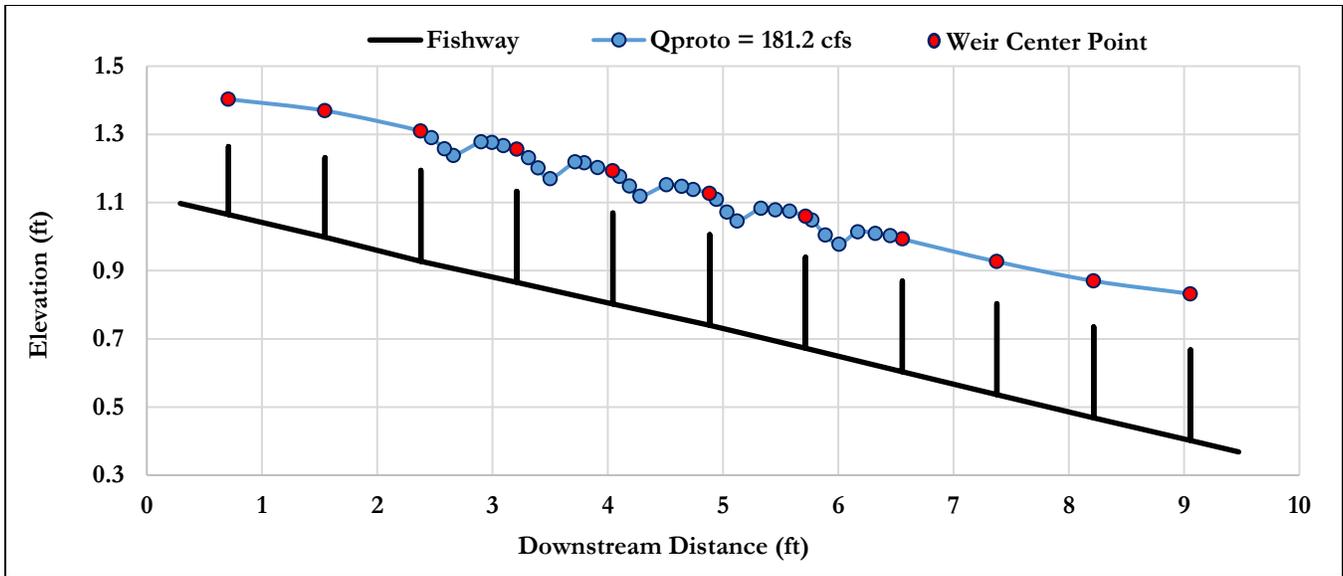


Figure 12. Water surface profile for 181.2 cfs prototype equivalent.

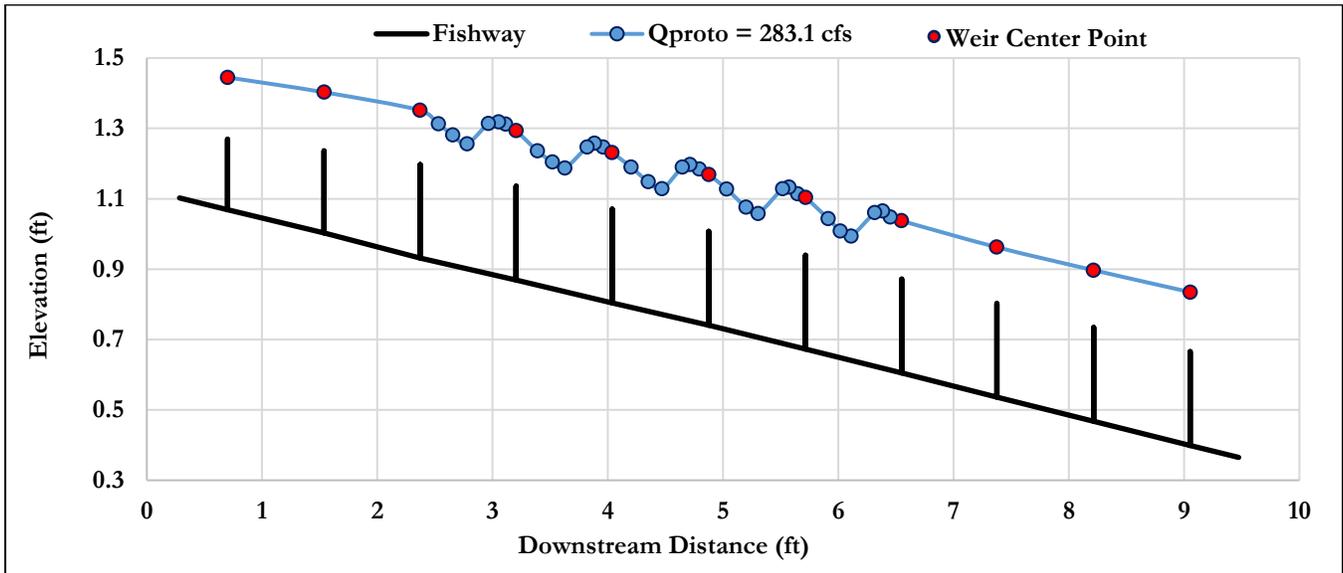


Figure 13. Water surface profile for 283.1 cfs prototype equivalent.

Examination of the water surface profiles shows the effects of momentum with increasing flow. This can be seen by the increasing magnitude of the hydraulic jump heights with increasing flow rates (Figure 10 -Figure 13). The water is also observed to drop over the weirs more abruptly for the lower flow rate runs, shown by the steeper downward slopes in the water surface profiles immediately downstream of the weirs. For the highest flow rate run, shown in Figure 13, the hydraulic jumps are shown to attain heights well above the water surface elevations located immediately above the weir crests.

During experimentation, it was observed that the water surface elevations near the walls of the fishway, where velocities were low, were substantially greater than those near the center of the channel, where velocities were high. Measurements were taken at the wetted edge of shoulder weirs where the water first started to plunge over the weir crests. Plots were generated to illustrate this effect for run #9 and run #12.

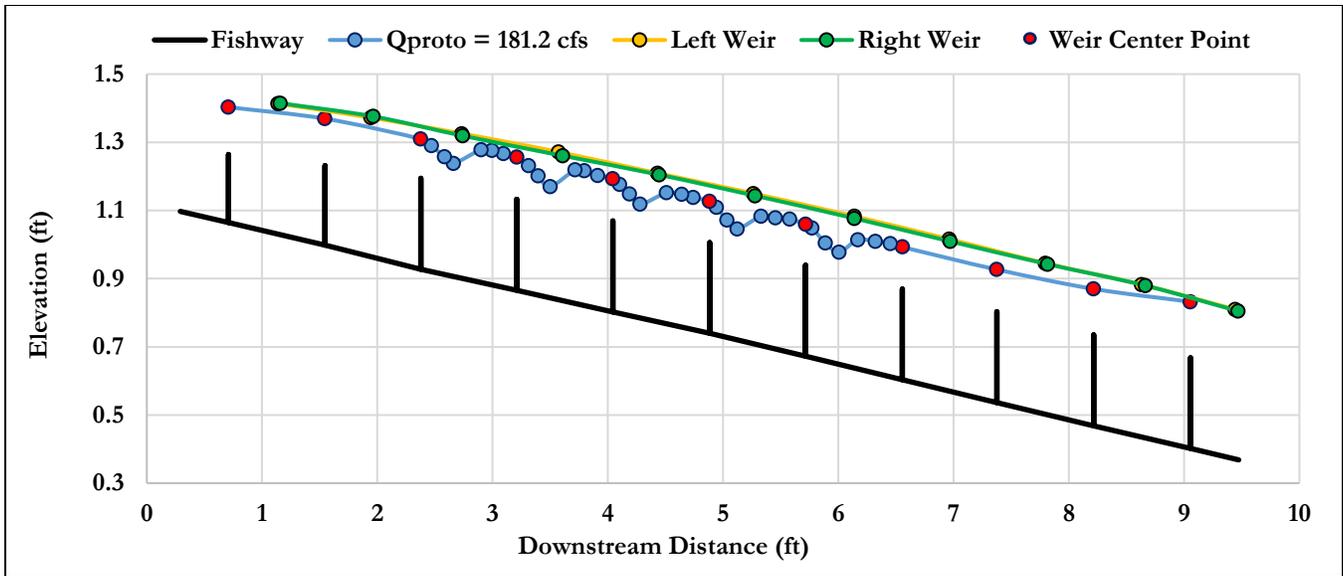


Figure 14. Water surface elevations at left, right, and center weirs for run #9 ($Q_{proto} = 181.2$ cfs).

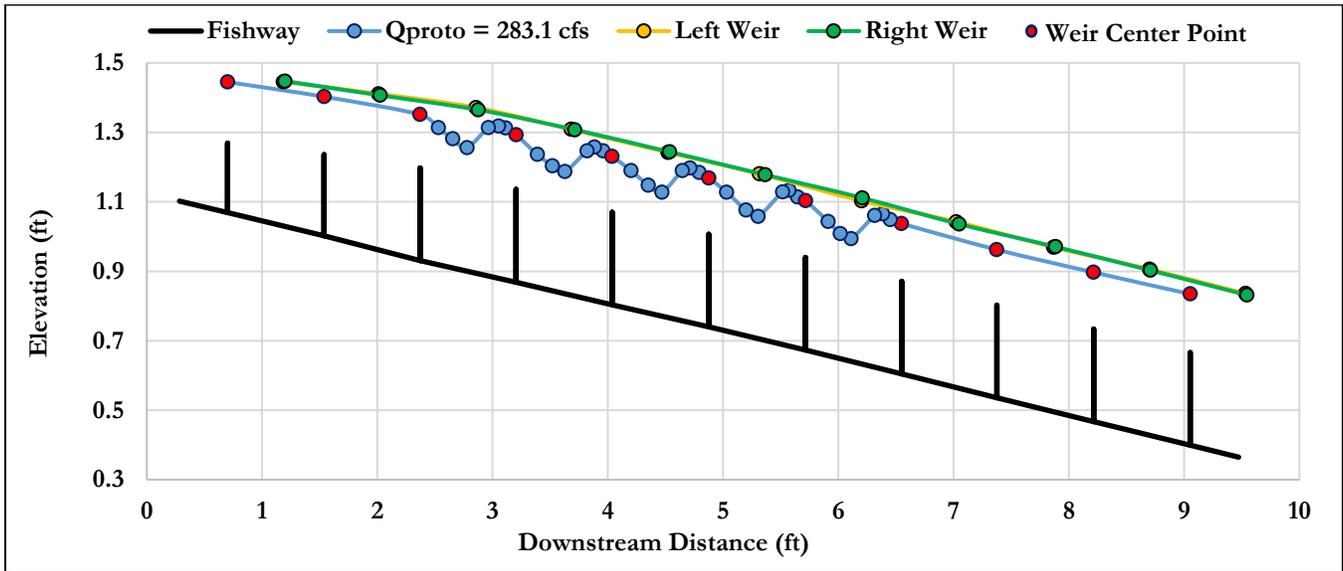


Figure 15. Water surface elevation at left, right, and center weirs for run #12 ($Q_{proto} = 283.1$ cfs).

Intermediate water surface elevations were not measured between the left and right shoulder weirs. Therefore, the water surface profiles for the shoulder weirs do not depict the plunging flow over weir crests, but merely illustrate the difference in water surface elevations between the center weirs and the shoulder weirs. For run #12, the highest flow rate run (283.1 cfs), the left and right weir profiles were in contact with the walls of the fishway (Figure 15).

Plugged Orifices

To test the effect of orifice obstructions within the fish passage, the orifices were plugged with clay for three of the flow rates analyzed (106.5, 143.5 and 220.2 cfs). Though obstructing the orifices had measurable effects when compared to the unplugged profiles, the differences were relatively small and diminished with increased flows.

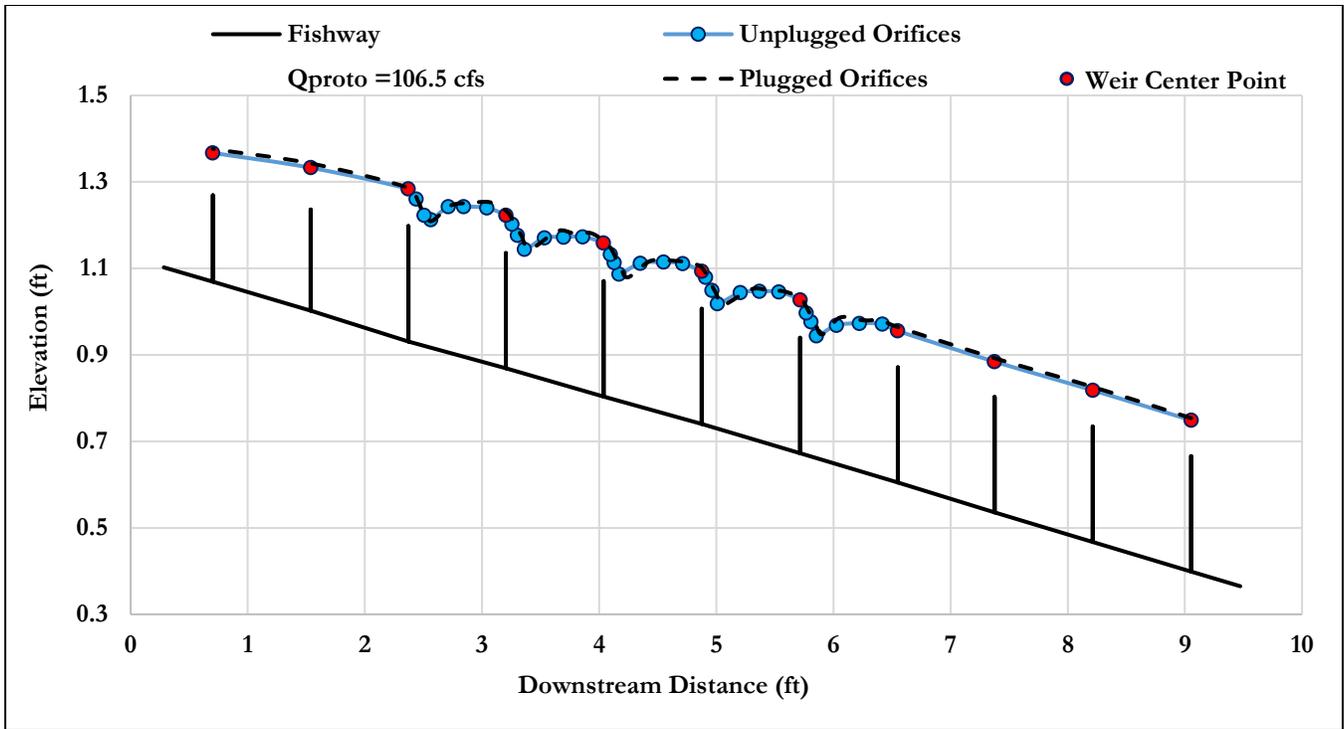


Figure 16. Plugged versus unplugged water surface profiles for 106.5 cfs prototype equivalent (runs #4 and #5).

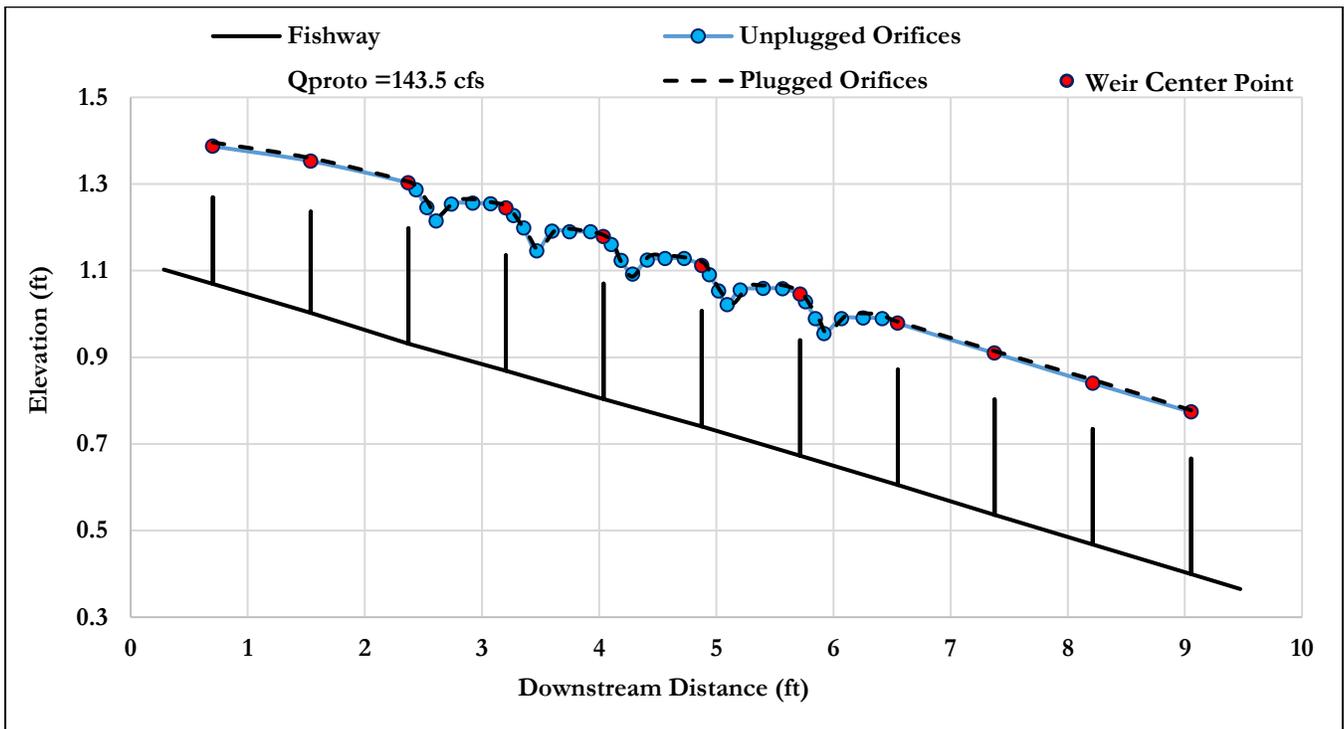


Figure 17. Plugged versus unplugged water surface profiles for 143.5 cfs prototype equivalent (runs #7 and #8).

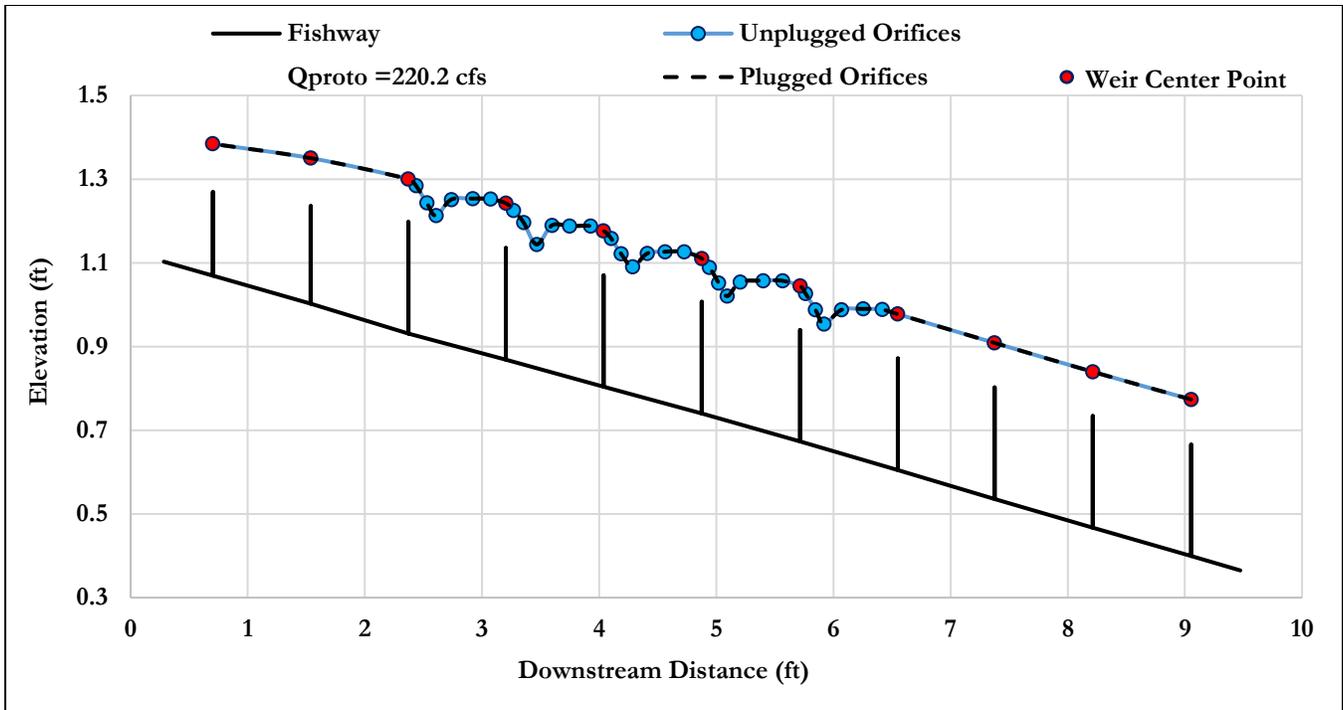


Figure 18. Plugged versus unplugged water surface profiles for 220.2 cfs prototype equivalent (runs #10 and #11).

The influence that plugged orifices had on elevating the water surface profile and the hydraulic jumps is indistinguishable at the highest flow rate. This is because orifice flow is held constant, assuming the water surface drop across the shoulder weirs remains the same with varying flows and plugging orifices adds a proportionally smaller amount of flow over the shoulders at larger flow rates.

Flow velocities

To confirm that flow conditions within the structure were favorable for fish attraction and passage, streaming and plunging prototype fishway average velocities were calculated for each flow rate, by dividing out projected areas.

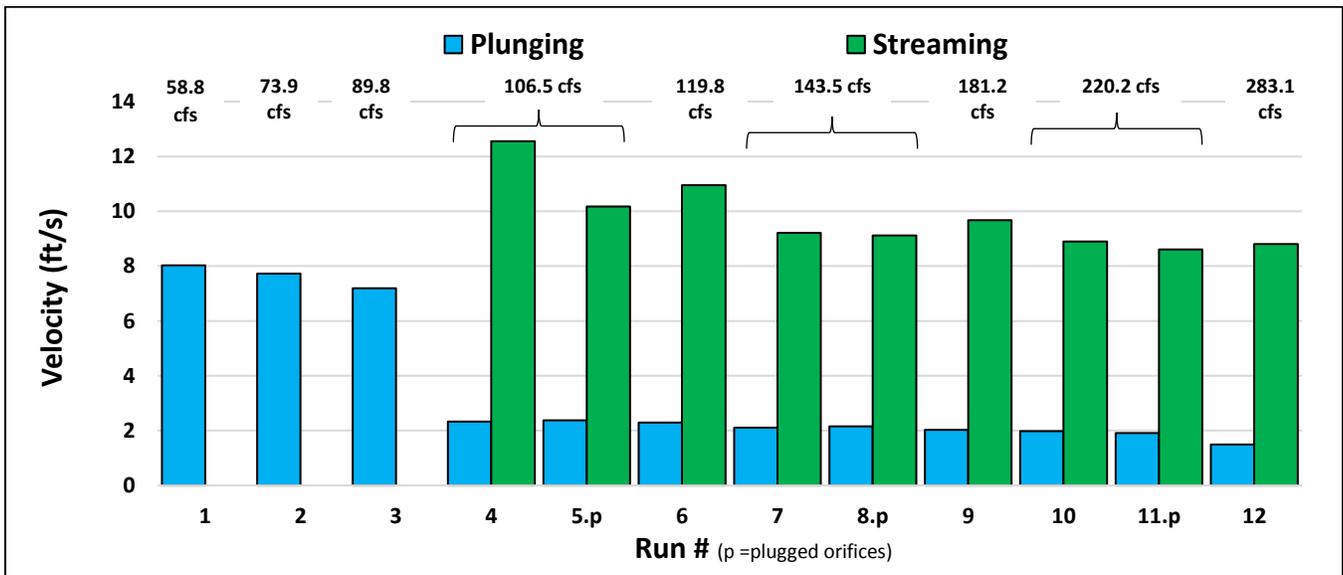


Figure 19. Calculated average plunging and streaming flow velocities at the weir crests of the prototype fishway for each experimental run.

For runs with plunging flow only, velocity was observed to decrease with increasing flows. Due to greater downstream pool depths at increasing flow rates, plunging flow energy was more effectively dissipated by larger pool volumes. A similar trend was

observed for streaming flow velocities. As flows increased, the location of the hydraulic jump, where flow slows from supercritical to subcritical, moved further into the downstream pools, which reduced the acceleration distance to the subsequent weir. The magnitude of the hydraulic jump height was also observed to increase, indicating greater energy dissipation due to turbulence in the pools at higher flow rates (Figure 10-Figure 13). Though the streaming flow velocities varied with flow rate, the plunging flow velocities remained fairly constant with the exception of run #12, where the plunging regime geometry may have been influenced by the sidewalls of the fishway. The plunging velocities were calculated to be approximately two ft/s for all runs where both streaming and plunging flows were observed. It is noted that these velocities are for the flows over the weir crests; velocities within the pools of the fishway appeared slower by visual inspection but were not quantified.

Sensitivity Analyses

A sensitivity analysis was conducted to determine the effects that changes in the recorded flume flow rates would have on the estimated Chezy coefficients. Imprecise readings of the stage in the flume head tank, the error introduced by the use of the discharge calibration curve, and small amounts of leakage bypassing the fishway may have contributed to the use of inaccurate discharge values in the estimation of Chezy coefficients. For example, an error of one-eighth inch in the stage reading would result in a 14% error in discharge for run #1 (58.8 cfs) and a 7% error for run #12 (283.1 cfs). Considering this possibility, the recorded flume flow rates were varied by $\pm 5\%$, 10% , and 15% . The results of the sensitivity analyses are graphically presented below for both the *streaming flow* and the combined *streaming and plunging flow* calculation methods.

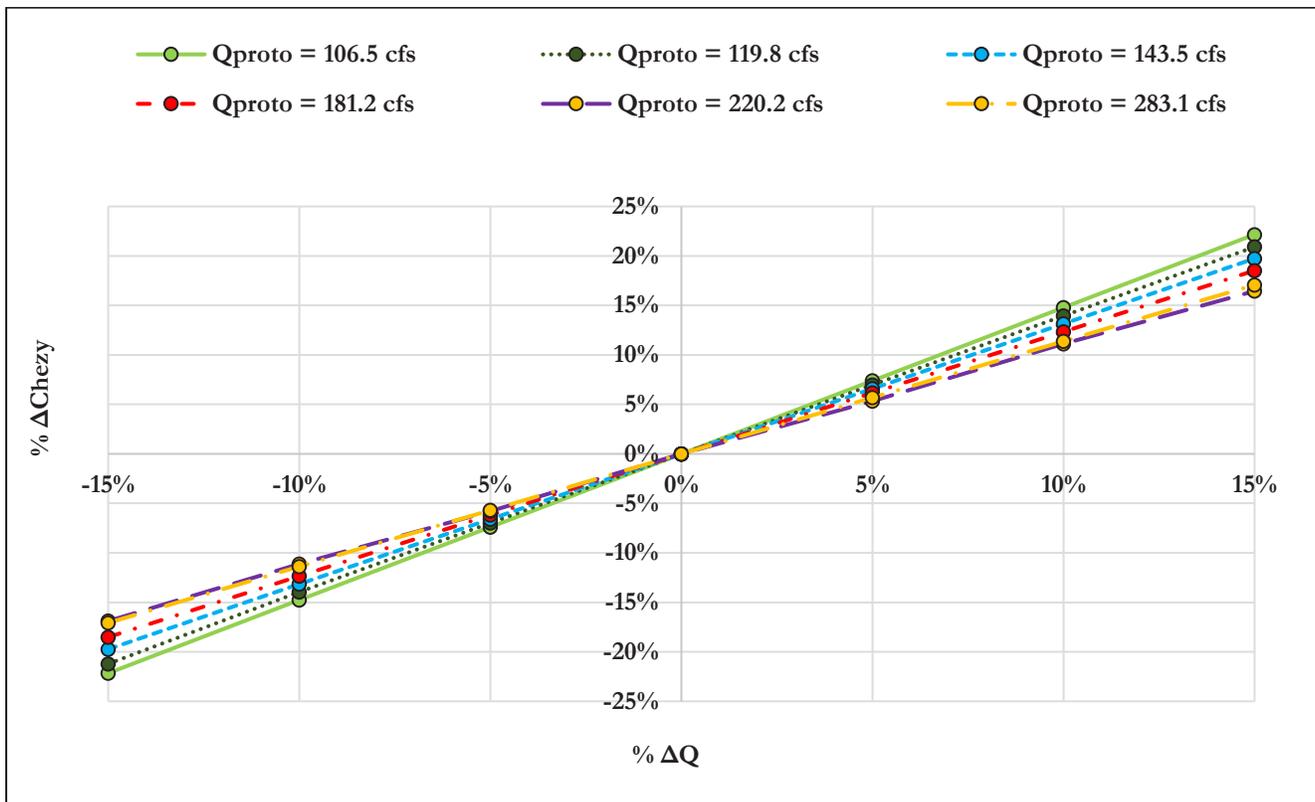


Figure 20. Sensitivity of Chézy coefficient estimates with changes in flume discharge (Streaming Flow Method).

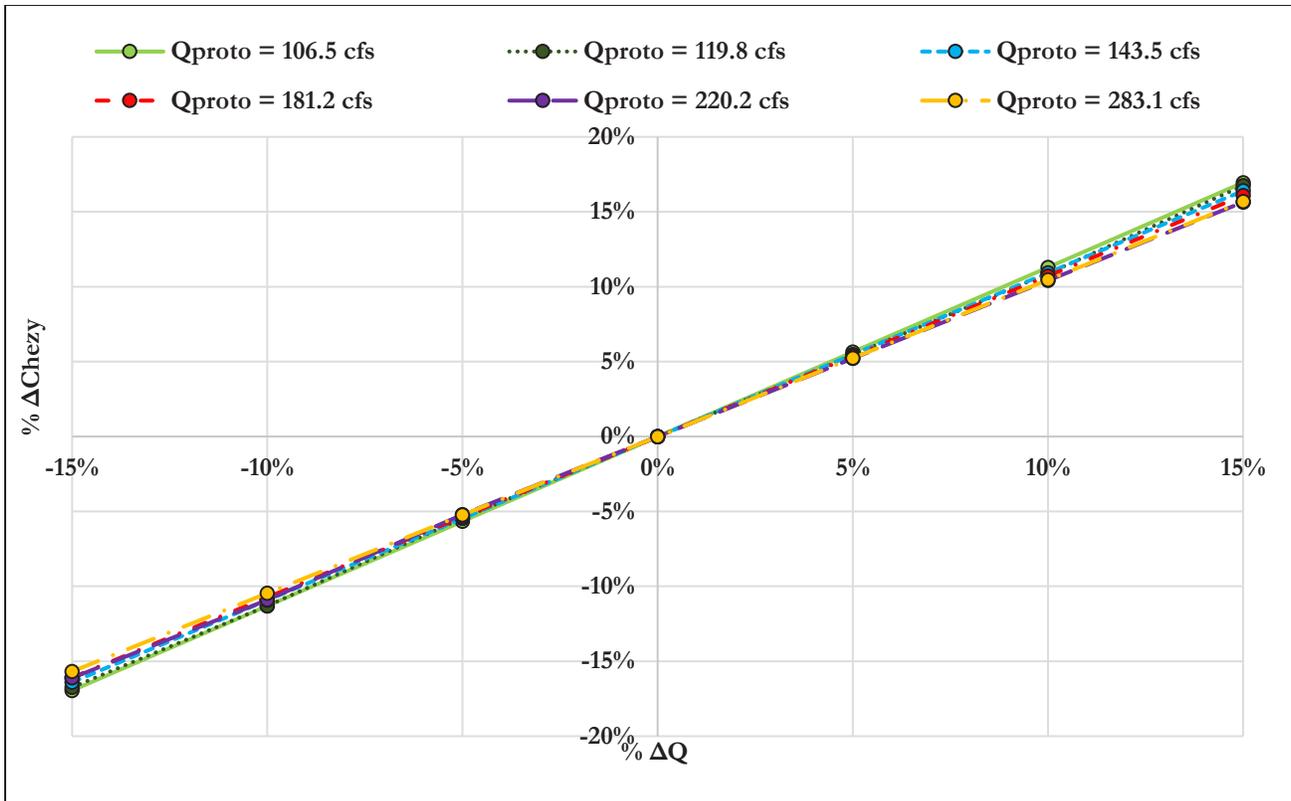


Figure 21. Sensitivity of Chézy coefficient estimates with changes in flume discharge (Streaming and Plunging Flow Method).

From Figure 20 and Figure 21, positive linear relationships are observed between increases in flume discharge rates and Chézy coefficient estimates. This relationship is similar for all unplugged flow rates with streaming and plunging regimes using both calculation methods. The effect is slightly more pronounced for the streaming calculation method as shown by the steeper slopes in Figure 20. The results indicate that if the actual flume discharge was less than the discharge recorded and subsequently used for estimating the Chézy coefficients, then the Chézy values would be less than those reported in this analysis. The opposite would be true for higher discharge rates. For a systematic error in stage reading, the impact on Chézy values would be more pronounced at lower flow rates. For example, a one-eighth inch stage reading discrepancy for run #1 would result in a 14% difference in flow rate translating to a 20% difference in the Chézy coefficient, while the same discrepancy for Run #12 would result in a 7% difference in flow rate, impacting the Chézy estimate by approximately 8%.

As previously mentioned, the *streaming flow* calculation method required values calculated with the theoretical relationships presented by Equations 7 thru 10. The water depth at which the flow regime transitioned from plunging to streaming flow (H_{PT}) was determined using Equations 7 and 8. These equations assumed the transition depth to occur over the horizontal portion of the weir crests; however, the transition between streaming and plunging flows occurred over the sloping shoulder portion of the weirs. The authors neglected to incorporate a correction factor to account for this difference. Therefore, actual transition flow depths may have been less than the theoretically calculated value of 0.082 ft (model fishway), which was subsequently used to determine the streaming flow cross-sectional areas used in the Chézy equation. A sensitivity analysis was conducted to determine the effects that decreases in the theoretically determined H_{PT} value would have on the Chézy coefficients estimated by the *streaming flow* calculation method. The H_{PT} value was reduced by up to 30%, in 5% increments. The results are graphically presented below.

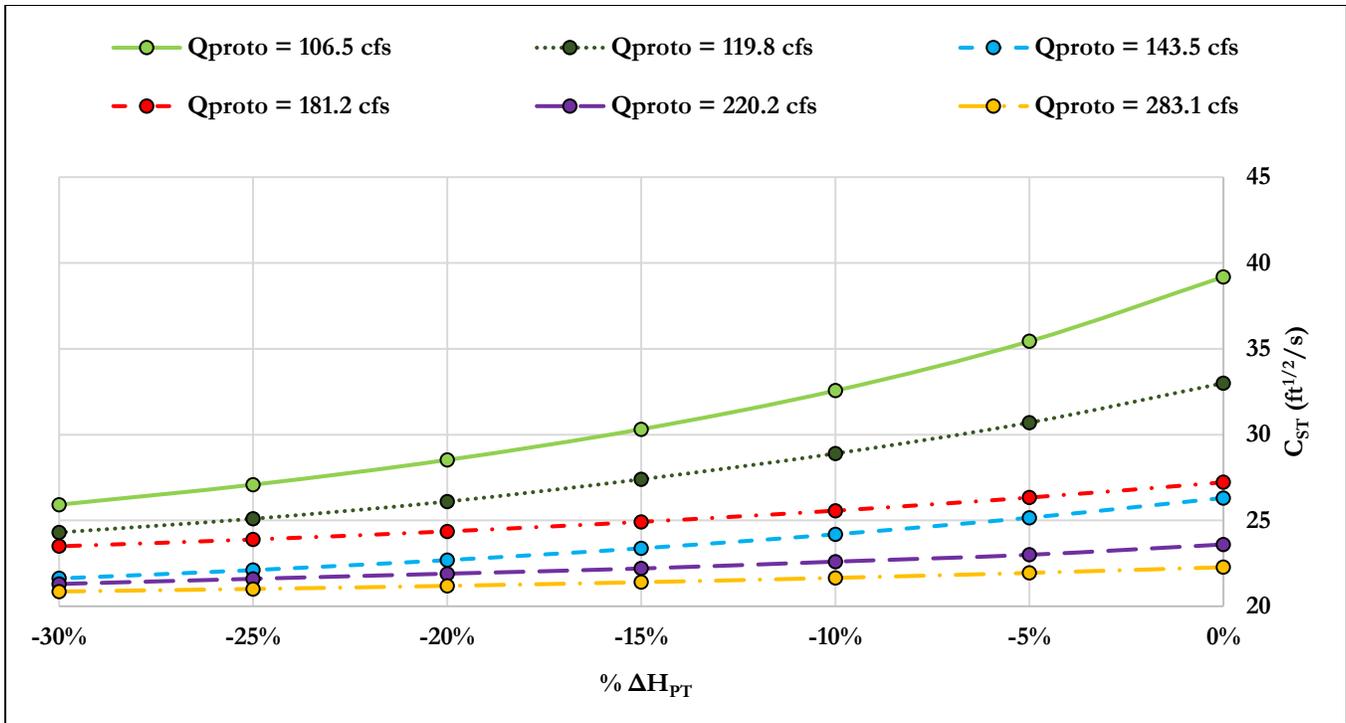


Figure 22. Sensitivity of Chézy coefficient estimates with changes in H_{PT} (Streaming Method).

From Figure 22, decreases in the theoretically calculated transitional flow depth (H_{PT}) are shown to result in decreases in the Chézy coefficient estimates calculated by the *streaming flow* method. The effect is observed to be more dramatic at lower flow rates where equivalent changes have a proportionally greater impact on streaming flow. Run #4 and run #5 had the lowest flow rates, Figure 22 shows that the Chézy estimates for these runs are substantially smaller with reductions in H_{PT} as little as 10%. Conversely, run #12 was the run with the greatest flow rate ($Q_{model} = 0.325$ cfs), and a reduction in H_{PT} as great as 30% only resulted in a 6.4% decrease in the Chézy coefficient estimate. The results indicate that if error was introduced into the Chézy coefficient estimations using the *streaming flow* calculation method, then the estimates reported in this analysis would be overestimates of the true values. It is also noted that a comparison between Figure 8 and Figure 9 reveals that the *streaming flow* calculation method did indeed result in greater Chézy coefficient estimates relative to those computed using the *streaming and plunging flow* method.

Run #	$Q_{Prototype}$ (cfs)	$C_{Streaming}$ (ft ^{1/2} /s)	$C_{Streaming\ and\ Plunging}$ (ft ^{1/2} /s)
1	58.8	No streaming flow	41.9
2	73.9	No streaming flow	37.4
3	89.8	No streaming flow	32.2
4	106.5	39.2	25.0
5 (Plugged Orifices)	106.5	30.4	23.2
6	119.8	33.0	23.9
7	143.5	26.3	21.3
8 (Plugged Orifices)	143.5	25.6	21.5
9	181.2	27.2	22.8
10	220.2	23.6	21.1
11 (Plugged Orifices)	220.2	22.5	20.3
12	283.1	22.3	18.9

Table 2. Comparison of Chézy coefficient estimates for streaming flow method and streaming and plunging flow method.

Therefore, the authors recommend that a correction factor be incorporated into future analyses since horizontal weir equations were used to approximate H_{PT} , when in fact transition from plunging to streaming flows occurred over the sloped shoulder portion of the weirs.

CONCLUSIONS

The results of this analysis include Chézy coefficient estimates for the model vortex pool-and-chute fishway, designed by Michael Love and Associates. The coefficients were calculated by two different methods, the *streaming flow* method and the *streaming and plunging flow* method, each based on the portion of flow used as the discharge in the Chézy equation. Chézy coefficient estimates ranging between 22.3 and 30.4 ft^{1/2}/s, with one outlier value of 39.2 ft^{1/2}/s, were determined by the *streaming flow* calculation method. For the same flows, Chézy coefficient estimates ranging between 18.9 and 25.0 ft^{1/2}/s were calculated by the *streaming and plunging flow* calculation method and values of 32.2 to 41.9 were calculated for flows that were exclusively *plunging*. The *streaming and plunging flow* calculation method resulted in lower Chézy coefficient estimates and less variance between estimates compared to those computed with the *streaming flow* calculation method. In both cases, results indicate that the fishway provides greater resistance as flows increase throughout the range tested. However, the composite Chézy coefficients based on the two flow regimes not only provide better and more stable results but also afford the ability to estimate the velocity in each regime. The plunging zones along the boundaries of flow over the weirs were found to maintain lower velocities, providing a thin nappe for fish to swim or leap over with little difficulty.

Orifices within the fishway weirs were plugged for 3 of the 12 experimental runs examined in this analysis. The plugged orifice runs were observed to result in mildly elevated water surface profiles relative to unplugged runs at the same flow rates. Furthermore, slightly lower Chézy coefficient estimates were determined when the orifices were plugged, which may offer a slight advantage to fishway performance at higher flow rates. The ability of the orifices to provide passage at lower flows and sediment conveyance at higher flows should be further examined to ensure that such functionality outweighs the reduction in flow resistance.

Based on velocities over the weir crests, derived from Chézy coefficient estimates, preliminary VPC model fishway results indicate that implementation of a prototype fishway would alter creek flow velocity effectively and sufficiently. Streaming flows through the center of the fishway appear to be adequate for fish attraction, while plunging flow along the edges provides for fish passage. Though pool velocities visually appeared to be lower, this observation should be quantitatively confirmed. Turbulence in the plunging regime corridor is also an important consideration that impacts fish passage effectiveness. The energy dissipation factor (EDF) is a metric used to assess turbulence and should be determined for the fishway pools. Further research is recommended to determine pool velocities and turbulence within the model fishway to ensure that they do not conflict with these findings.

REFERENCES

- California Department of Fish and Wildlife (2010) California Salmonid Stream Habitat Restoration Manual, 4th ed., Vol. Two, Part VII, 117, <http://www.dfg.ca.gov/fish/resources/habitatmanual.asp> (accessed May 2015)
- Caltrans (2009) Fish Passage Design for Road Crossings, Chapter 9: Fishways, <http://www.dot.ca.gov/hq/oppd/fishPassage/Chapter%209%20Fishways.pdf> (accessed April 2015)
- Bates, K. (1991) Pool-and-Chute Fishways, American Fisheries Society Symposium, Washington: American Fisheries Society, 268-277
- Love, M. (2015, April) Email Correspondence with Eileen Cashman, Arcata, CA, U.S.
- Love, M. (2015) Home: Michael Love & Associates, <http://h2odesigns.com/> (accessed April 2015)
- Powers, P. (n.d.). Pool and Chute Fishways: Discussion and Design Process.
- Allen, S., Love, M., and Llanos, A. (2004) Peacock Creek Pool and Weir Fishway, <http://www.stream.fs.fed.us/fishxing/case/Peacock/PeacockPoster.pdf> (accessed April 2015)
- National Marine Fisheries Service (2011) NOAA. Anadromous Salmonid Passage Facility Design, <http://www.westcoast.fisheries.noaa.gov> (accessed April 2015)
- White, F. M. (2011) *Fluid Mechanics*, 7th ed., 721, McGraw Hill, New York
- Houghtalen, R. J., Akan, A. O., and Hwang, H. (2010) *Fundamentals of Hydraulic Engineering Systems*, 4th ed., 186-189 Pearson Higher Education, Boston
- Love, M. (2015) Vortex Pool and Chute Fishway Calculations. Arcata, CA.
- Ead, S.A., C. Katopodis, G.J. Sikora, and N. Rajaratnam (2004) Flow Regimes and Structure in Pool and Weir Fishways, *Journal of Environmental Engineering and Science*, Vol. 3(5), 379-390
http://www.researchgate.net/publication/233617887_Flow_regimes_and_structure_in_pool_and_weir_fishways (accessed April 2015)
- Kindsvater, C.E., and Carter R.W.C. (1957) Discharge Characteristics of Rectangular Thin Plate Weirs, *Proceedings of the American Society of Civil Engineers, Journal of the Hydraulics Division*, Vol. 83, No. HY6, December 1957, pp. 1-36
- Villemonthe, J. (1947) Submerged -Weir Discharge Studies, *Engineering News Record*, Vol 866, 54-58, http://www.stream.fs.fed.us/fishxing/fplibrary/Villemonthe_1947_Submerged_weir_discharge_studies.pdf (accessed April 2015)
- Daughtry, R. L., and Franzini, J.B. (1965) *Fluid Mechanics*, 6th ed. 338-349, McGraw-Hill, New York
- Engineering Toolbox (2015) Weirs-Flow Rate Measure, http://www.engineeringtoolbox.com/weirs-flow-rate-d_592.html (accessed April 2015)
- LMNO Engineering Research and Software Ltd. (1999), Focus on Open Channel Flow, Newsletter Vol 1, [http://www.lmnoeng.com/Newsletters/newsletters.htm#v.1\(17\)](http://www.lmnoeng.com/Newsletters/newsletters.htm#v.1(17)) (accessed April 2015)

ABOUT THE STUDENT AUTHORS

Brian Draeger, Mat Nyberg, and Brian Weekly completed this research while pursuing B.S. degrees in Environmental Resources Engineering at Humboldt State University in Arcata, California. The authors share a common interest in the fields of hydrology and hydraulics. The analysis presented in this paper was conducted as a semester project for a River Hydraulics course instructed by Dr. Eileen Cashman. The students' interest and effort toward this project has continued beyond their course work under the direction of Dr. Cashman and Dr. Margaret Lang.

PRESS SUMMARY

Fish passages play an important role in the preservation of migratory fish species in waterways that are impacted by man-made structures. Fish passages (or fishways) cause resistance to flow and create specific hydraulic conditions sufficient for fish migration. This paper offers methods and results for the effectiveness of a vortex pool-and-chute fish passage design on the basis of Chézy resistance coefficients.

Dynamic Visualization of Time-Based Changes to Data Generated by Reddit.com: The Real Time Conversation Project

Daniel Wang, Andy Luse*, and Jim Burkeman

Department of Management Science and Information Systems, Oklahoma State University, Stillwater, OK

Student: danile.wang@okstate.edu

Mentors: andyluse@okstate.edu*, jim.burkeman@okstate.edu

ABSTRACT

With the increased amount of data generated by social networking sites there is also increased difficulty in the analysis of this data, including time-based changes, which can provide unique insights in social network analysis. Information visualization is a vital tool in assisting social scientists with analysis of large quantities of data; however, the gathering, formatting, and visualizing of time-related data from social networking sites still remains an obstacle. This research explores the process of gathering time-based data in real time and using dynamic visualization techniques to visualize and analyze time-based changes in data generated by discussions on the social networking site Reddit. The outcome culminates in our deliverable, the Real Time Conversation Project.

KEYWORDS

Visualization; Network Analysis; Social Network; Reddit; Gephi

INTRODUCTION

Information visualization is the process of creating a visual representation of an abstract set of data through the usage of conventional or computer graphics.¹ The goal of information visualization is to assist end users in perceiving and analyzing the underlying patterns and effects that exist within the data set that may otherwise be difficult to recognize. The visualizations created from the data set assist analysis through the addition of another dimension to the data beyond the alphanumeric dimension, typically by displaying the relationships between different units of data within a set or the relationships between different sets of data.

A social network is, in the simplest terms, the connections that are formed by the social interactions of a group of individuals. More specifically, a social network is a group of people who may form personal relationships, or socially interact with one another on a regular basis.² Traditional social networks are formed geographically and generally remain localized. The advent of the internet has significantly changed the social spaces within which people are active and has redefined the concept of social networks.³ The term social network has begun to be more commonly used to refer to people connected through the use of computer communications.⁴

Social networking sites are the localization, organization, and facilitation of social activities in the online realm, often sporting a formal and organized format that makes social connections visible to the public. These social networking sites grew from the more informal and less organized systems of communication services such as internet relay chat and blogging services. In the modern day, each of these social networking sites prioritizes and orients itself toward a specific context of social interaction, usually delineated by goals, interests, or activities.⁵

Social network analysis is a subcategory within the social sciences that focuses on the study of interactions between individuals within a social circle or social network utilizing mathematical techniques and metrics native to graph and network analysis.⁶ With the widespread adaptation of social networking sites in the last few decades there has been a dramatic increase in the interest in using social network analysis to process the large amount of data that has become available to social scientists.⁷

Visualization in social network analysis

While there has been an increase in the amount of data available to social scientists, the gathering and organizing of this large amount of data into usable forms has been a hurdle to large scale social network analysis. The user interfaces of social networking sites are usually designed to facilitate discussion and interactions between users, and while these interfaces work for their intended purpose, they do not serve well as an information visualization system to assist social network analysis, lacking most of the components that are vital to information visualization systems.⁸

The creation of information visualization systems for social network analysis is also complicated by the differences of the underlying structures of social networking sites. The very thing that social networking sites do in order to separate themselves

from their competition has made it difficult to create a universal tool to gather, analyze, and visualize data generated by social networking sites. Information gathering and visualization systems need to be tailored to individual social networking sites. Many such tools have been created to facilitate data gathering and analysis on a number of different popular social networking sites,⁹ by joint efforts between social scientists and programmers as well as by hobbyist programmers driven by curiosity.⁴

One such social networking site is Reddit, an online community image and message bulletin board with content submitted and curated by users through the usage of a real time scoring system. While a few visualization tools have been created in order to visualize the data generated by Reddit, none have incorporated dynamic, time-based changes as an aspect of social interaction.

Goal of project

With the changes in content consumption habits of individuals through the use of the internet, popularity has become a very important metric for measuring the success of produced content, and dynamic, time-based changes to consumption and sharing of content can impact end-state popularity. Using Reddit as a platform, the goal of our Real Time Conversation project is to create a visualization system that can view the dynamic time-based changes of social interaction that take place on Reddit.

OVERVIEW OF REDDIT

Reddit is an online community image and message bulletin board that hosts content submitted and curated by its users. Reddit's unique structure allows users to cast votes on the submitted content using a democratic system of one up or down vote per user. From the perspective of an individual user, upvoting submitted content is to judge that content to be likeable, good, and interesting, whereas downvoting the submitted content is to evaluate that content as unlikable, low quality, or repulsive. The aggregate effect of this rapid-fire democracy is to determine, via input from the user base, what content submitted to Reddit can be considered constructive and beneficial to the community as a whole and what content is considered to be disruptive and destructive to the community. The same democratic voting process is also extended to the individual comments within the discussion section of Reddit, where users discuss and comment on the submitted content.

The majority of the content sharing and discussion activities on Reddit take place in user created sub-communities known as Subreddits. These Subreddits are often delineated by and focused on specific topics or themes. Some Subreddits focus on sharing content and having discussions over a specific subject such as technology, politics, or the latest popular television show, while other Subreddits may be for sharing content in a specific format, such as discussions, videos, or animated gifs. Because of the ability of users to create Subreddits, one of Reddit's strengths as a social networking site is that there is a Subreddit for just about every topic. Of course, the activities and popularity of these Subreddits do still rely on the level of user activity within.

Reddit selects a choice subset of Subreddits that they consider to be either interesting and relevant to current events or active and well moderated to make up the default landing page for Reddit.com. This landing page, known as the Front Page, displays a list of twenty of the most popular content submissions from the subset of Subreddits, allowing casual users or unregistered readers a low commitment way of engaging with the content available on Reddit. Content with the highest aggregated score within the choice subset of Subreddits is displayed towards the top of the front page, allowing it to gain a great amount of visibility and, in turn, additional popularity and score. Content that makes it to the front page is periodically moved down the list in a process independent from voting in order to allow room for fresh new content. The tiered process for displaying content on Reddit works synergistically with the democratic voting process. Content that is considered "better" gains more upvotes and becomes more visible, where it is capable of gaining even more upvotes and more visibility until it has run its course. The best content within a period of time is capable of picking up the most traction from this positive feedback loop, with the greatest amount of traction being gained when it reaches the front page. On the other hand, low quality or bad content becomes less visible as it is downvoted, making it harder for other users to view, and more importantly, vote on the content, thus causing it to sink into obscurity.

In addition to the ability to aggregate high quality content through the democratic voting process, Reddit is also host to discussions regarding the submitted content, with certain Subreddits such as AskReddit, askscience, explainlikeimfive and IAmA being solely dedicated to discussions. In these discussion-focused Subreddits, the original submitted content, usually in the form of a question, is treated as a conversation starter and the discussion makes up the real substance of interest. The discussion section is also subjected to the same democratic voting process that governs the submitted content, allowing comments to be upvoted or downvoted by the community, creating a similar aggregate effect within the discussion. Likewise, the individual comments of the discussion are also sorted and displayed in a manner similar to the submitted content, with a vertically nested list of comments that contain a list of its child comments and so forth. After a certain number of nested comments, Reddit puts a link to the additional discussions in place of another round of child comments, keeping the threads from becoming too long or losing focus (see **Figure 1** for an example Reddit page).

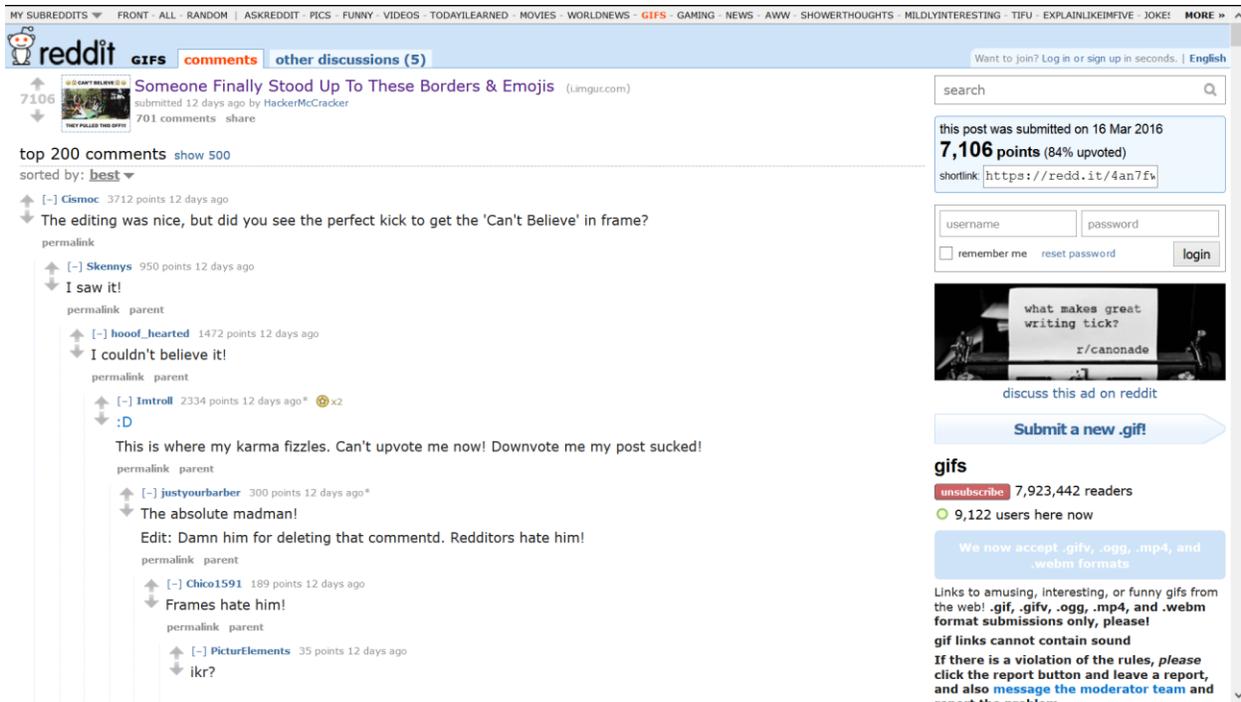


Figure 1. Screen capture of Reddit.

While there are a few methods to sort the list of comments in the discussion, Reddit, by default, sorts the comments in a manner that promotes discussion, with a mixture of score and number of child comments determining the position of that comment within the nested list. This same process affects each group of nested child comments. Eventually, a reader of the discussion section is left with a distilled version of discussions that have taken place in relation to the original submission. These distilled discussions are usually informative, entertaining, or humorous, and they should be, since hundreds of other people have already evaluated these comments as being such and sorted them through the voting process. Discussions do not remain solely in their respective discussion sections; however, as users move across different Subreddits and different discussions bringing humorous anecdotes and stories with them to share within another discussion. Similarly, content that is submitted to a Subreddit not within the choice subset of Subreddits that make up the front page, may oftentimes be resubmitted on a different Subreddit where, hopefully, more people would be capable of seeing it with the ultimate goal of reaching the front page.

Lastly, individual users on Reddit have a cumulative score known as Karma, a sort of reputation. There are two types of Karma, one for submitting content and one for engaging in discussions, but they function in the same manner, being the cumulative score of all the content or comments that a user has submitted. The Karma system is a sort of encouragement for users on Reddit to remain respectful and follow community guidelines with the threat of losing credibility and reputation.

Current visualizations

Andrei Kashcha (of Yasiv.com) created a visualization of Reddit in 2012 that shows the relationship between individual Subreddits. Using hyperlinks that point to other Subreddit's communities that are contained in any individual Subreddit's description, the visualization created a network graph of the communities of Reddit. This visualization allowed viewers to see the connections between different Reddit communities that may or may not share some common theme. Andrei Kashcha's visualization contains roughly 3800 Subreddit communities and 14000 connections between them, but does not include any of the large number of isolated communities that are unconnected to others.¹⁰

Randal S. Olson (of the University of Pennsylvania's Institute for Biomedical Informatics) also created a visualization of the Reddit sub-communities in order to explore how a network map of primary topics of interest may assist users in finding and organizing into specific interest groups.¹¹ In the same vein, in 2012 user Laurel Quade created a fictional map of the major communities of Reddit, grouping the major areas of interest with the cities scaled to the number of subscribed users. While not the most accurate or technical of visualization, it does demonstrate the vast variety present in the Reddit community and different entertaining methods of visualization.¹²

Kawandee Virdee (of Embed.ly) created a visualization of discussions on Reddit in a networks graph in 2014. The visualization created a color coded network graph of the comments of users that participated in any particular discussion, showing both how and where in a discussion users post and how often users engage in back and forth discussion with others by posting more than once. The visualization is publically accessible on Github and is an excellent tool to analyze the different structures of conversations that occur in different Subreddits. The visualization highlighted the differences in discussion structures between regular and specialized Subreddits. Specific Subreddits, such as IAMA where users share their profession and answer questions posed by the community, were naturally more prone to reciprocative discussions than other Subreddits.¹³

While all the above visualizations offer interesting pictures of Reddit activity, all these visualizations offer only a static picture of what occurs on the site. These static pictures are not able to depict the longitudinal aspects of Reddit data to enable understanding of the rise and/or fall of popularity trends of Reddit data and the connections between these changes across Subreddit communities.

REAL TIME CONVERSATION PROJECT

Reddit has recently taken steps to reduce outside influences to the voting process by making the score of submitted content and comments unknown for a short period of time after submission. This new policy, coupled with the fact that a post and content have their scores displayed as a cumulative number instead of separate upvotes and downvotes mean that regular browsing and static visualizations do not, or are not capable of, detailing time-based changes in the score of discussions. While this does not interfere with the common usage of Reddit, it does make analysis of the effects of popularity in a discussion inaccurate. A simplistic solution to the lack of time-based changes to data may be to assume a linear increase in comment score over the period of existence of a particular post or comment; however, assuming a linear increase in comment score fails to take into account any variances in score that may be caused by opposing reactions in conversations over time or any effects of popularity that may boost the content's score. The solution to the lack of readily available data regarding time based changes to the score on Reddit would be to gather data over a period of real time and integrate that additional dimension of data into current visualizations.

To that end, our Real Time Conversation project aims to create a solution to the lack of data regarding time-based changes to the score on Reddit by recording data in real-time. The end product of the Real Time Conversation project application gathers Reddit submissions from the rising page and recursively parses the information related to the submission and comments, with the major variable of interest being score. After twelve hours, the submissions that made it to the front page are exported as an undirected relational network in a GEXF¹⁴ file for visualization within Gephi.¹⁵ Each node of the network graph represents a comment, and the relation between the nodes represents the nesting of comments within the discussion. Gephi is capable of organizing the relational network into informative formats through built in algorithms and is also capable of dynamic visualization, adapting and changing aspects of the visualization over time as defined dynamic attributes are changed.

Within the scope of this visualization system there are two dynamic attributes: the period of existence and the score of a node. The period of existence of a node corresponds to the time when that node's comment is posted in the discussion, the original content submission will always be the first node, and parent comments will always exist before their child comments. The score of a node is used to dynamically affect the size of the node over time. The increase and decrease in size correspond to the popularity of the comment. The goal of this dual dynamicity is to present to the end user, in a natural and intuitive manner, the growth of the size and popularity of a discussion in relation to the number and popularity of the comments within that discussion.

Implementation

The solution to the lack of data regarding time-based changes to the score of content and comment submissions on Reddit is to gather data from Reddit in real time. There are several methods to gather data from Reddit (such as parsing Reddit pages as JSON files) but the most formalized one is to use the Reddit application program interface (API).¹⁶ The Reddit API was designed to allow the automation of a variety of complex functions related to using Reddit and moderating Subreddits. Subreddit moderators can use the Reddit API to write scripts that automate certain aspects of the moderation process, such as complaint handling or automatic keyword detection. Hobbyist users have used the Reddit API to create simple automated applications dedicated to specific functions, such as the user account MetricConversionBot, which automatically detects imperial units within any comment and converts them to metric units in a child comment. Similarly, the user account JiffyBot converts Youtube videos to animated Gif files and rehosts them on an easily accessible website and posts a link to that image in a child comment.

As a central function, the Reddit API allows automated applications to gather and sort through a large amount of data from Reddit. However, there are certain rules and restrictions to using the Reddit API such as a limit of one connection per two seconds from any unique IP address, these limitations were created in order to regulate automated applications and prevent intentional or unintentional overloading of the Reddit servers. To simplify usage of the Reddit API, some independent programmers created the Python Reddit API Wrapper.¹⁷ PRAW is a package that is available to the programming language Python that allows programmers to streamline their development process by handling the data retrieval and Reddit API interface

process, automatically bringing the application in line with Reddit API regulations. PRAW also assists programmers by providing better organization to both the process, and the data retrieved through the Reddit API. Using the Reddit API and PRAW, submissions and comments are retrieved from Reddit as objects with associated variables containing all the relevant information for that submission or comment. Each submission or comment is uniquely identified by a 6-digit alphanumeric string, and the string can be used to directly retrieve the referenced submission or comment page on the Reddit website.

PostgreSQL¹⁸ is an enterprise scale open source relational database system, being both feature-rich and scalable. PostgreSQL was chosen for this project in part due to its ability to handle a large amount of data and in part due to its popularity and open source nature encouraging developers to create additional systems and modules that can interface with it. PSYCOPG2¹⁹ is used in this project to connect PostgreSQL and Python. The Real Time Conversation application is written in Python 3.4.3. Using the PRAW and PSYCOPG2 modules, the application moves data from the Reddit API into PostgreSQL by using the available variables of the pre-defined submission and comment objects in PRAW.

To start, a couple of sets are made in order to keep track of the submission parsed over the period that the program runs, one set (A) is of all of the unique IDs of submissions that appear on the front page over the 12-hour period. The other set (B) contains the unique IDs of all of the submissions that are being parsed during the same time. To start parsing data, a list of the submissions from the rising page is retrieved from Reddit using PRAW. For each submission in the list, if its unique ID is not already in set (B), the unique ID is retrieved and used to create two unique tables in PostgreSQL: one for nodes, storing the comments, and one for edges, storing the relationship between comments. After the appropriate tables have been created, the application retrieves each of the submissions in set (B) from PRAW and all of the comments for that submission are placed into a set and recursively parsed to extract the unique ID and current time variable. These variables are used to create a new record in the database, containing additional information parsed from variables such as the body text of the comment. Once the process of parsing all comments is completed, the application returns to the beginning and retrieves the new list of submissions on the rising page, starting the process over.

While this process seems to retrieve redundant data many times, such as submission and comment ID, these are used in conjunction with the current time to uniquely identify each record. Variables that very rarely change, such as the body text of the comments, are only re-recorded if there was actually a change. This is the simplest of several methods that could be used to record information of comment scores at different times; it is also the one that records the most complete information. Alternative methods have been implemented for testing, but the basic concept of checking each individual comment for change in score is still necessary. This method simply records all the data and leaves any organizational process to another portion of the code that does not have to interface with the Reddit API and retrieve information from Reddit.

After 12 hours have elapsed, any discussion that did not reach the front page is dropped in order to reduce the amount of unnecessary information stored in the database. This is achieved by using any unique ID that does not exist in both set (A) and set (B) to retrieve and drop its respective tables. After the data reduction process, the remaining tables for this 12-hour cycle is run through a conversation script that retrieves the necessary data from the tables and forms it into a GEXF file that can be visualized using Gephi. The conversation script uses PSYCOPG2 and the unique ID of each table to query the available data from each database table, organizes the data by nodes, and sorts each node's worth of data by time. The conversion script then wraps all of the data with the appropriate GEXF format, adds the header and footer information, and writes the entire file out to a user defined directory. **Figure 2** shows the process diagram and related procedural list summarizing the workflow of the solution.

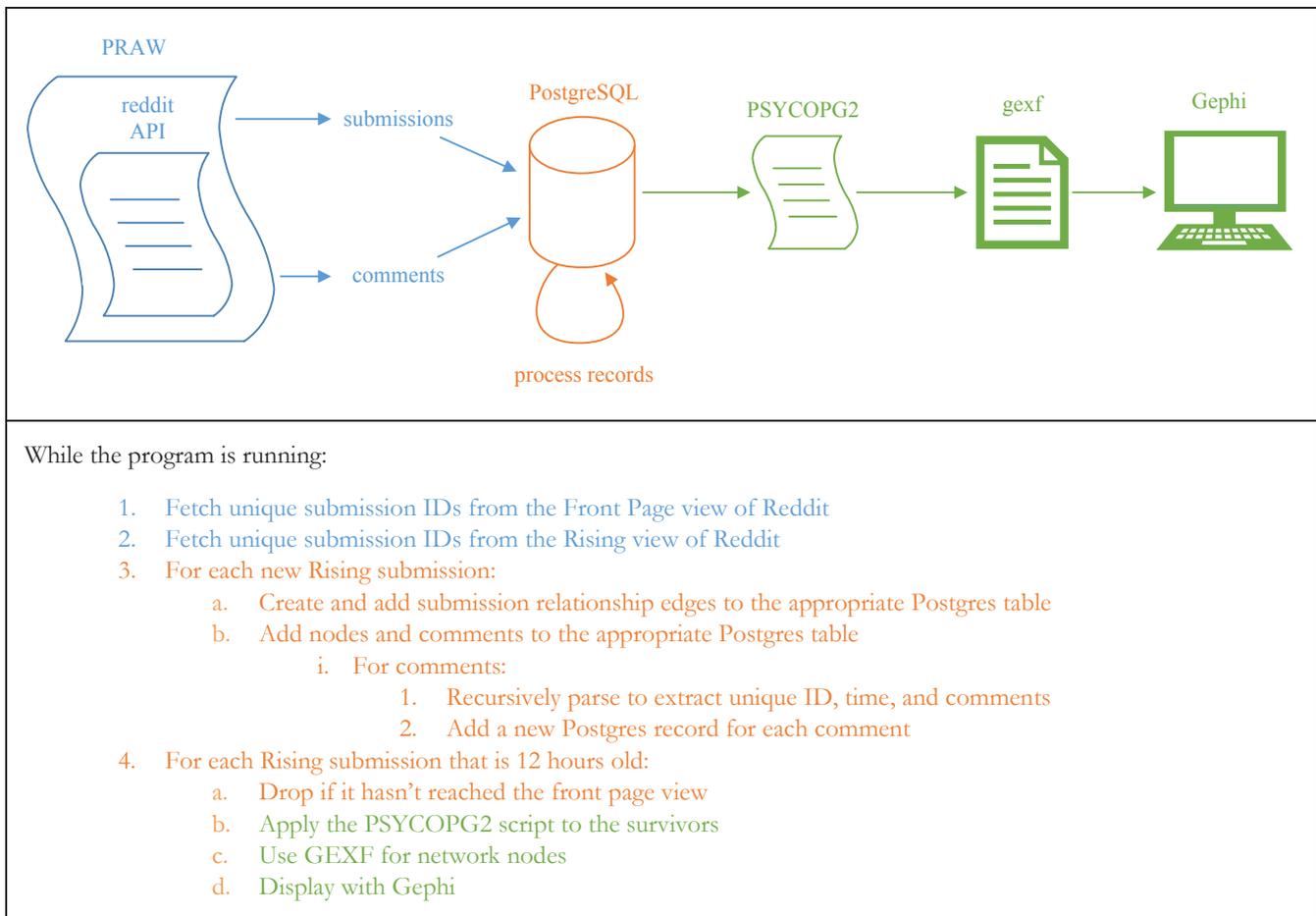


Figure 2. Process diagram and procedural list for solution.

GEXF, or Graph Exchange XML Format, is a language that details the data and dynamics of complex network structures. Created by members of the Gephi project solely with the goal of representing network graphs, the GEXF graph format was created to be able to represent nodes, edges and data associated with a network graph. With the additional ability to handle any hierarchical structures or dynamic functions that may be present within a network graph, GEXF is a powerful, extensible, and open format that is mature enough for specific applications.

Gephi is an open source network graph visualization platform. Built as a tool to assist analysis of graph data, Gephi enables users to manipulate the structure, shape, and color of the visualized graph data in order to assist with data analytics. Gephi is an excellent information visualization system, containing all of the components necessary to an information visualization system, with additional functionality for exploration of other aspects of data analytics and statistics. Due to its open source nature, Gephi allows users to extend its functionality through the usage of plugins, making it a very versatile platform for visualization; however, in the same sense that Photoshop is not the best tool to use to view pictures, Gephi is not suited to casually viewing data visualizations. Gephi is more specialized towards an analytical role with its complex functions.

Because of the community friendly nature of the open source format, Gephi streamlines the installation of plugins. The role of plugins within Gephi is to extend Gephi's information visualization or data analytic functionality. Being open source assists Gephi in this regard, and there is an extensive library of plugins available for users seeking specialized functionalities. Plugins used in the Real Time Conversation project were specialized to extend the extraction functionality of Gephi as an information visualization system. Gephi, by default, is able to export visualizations in a couple of different formats including SVG, PDF and PNG files. The plugins Seadragon Web Export and Google Maps Exporter extended the extraction functionality by providing a different method of extracting and displaying visualization data.

Seadragon Web Export²⁰ allows the exported graphic to be viewed using Microsoft's Seadragon Software through a web browser. Seadragon is the powerhouse behind Microsoft's Silverlight, Pivot and Photosynth applications and focuses on smooth browsing

of graphics of any size. Google Maps Exporter²¹, on the other hand, extracts the visualization created in Gephi for usage with the Google Maps API, allowing end users to view the extracted visualization in a browser in the same intuitive method that they would use Google Maps. Both of these plugins give the extracted visualization a couple of additional information visualization functionalities such as smooth zooming and detail views instead of a simple static overview. However, the functionalities of Seadragon Web Export and Google Maps Exporter do not meet to the goal of the Real Time Conversation project, which is to visualize the effects of dynamic time-based changes to data that take place on Reddit. While the Gephi plugin library is large, there was no plugin that could specifically address the requirement for dynamic visualization exporting; however, Gephi is able to export screenshots of the visualization during each stage of the dynamic visualization. Utilizing this function, it was possible to use extracted screenshots at each interval of the dynamic visualization to create an animated Gif file that shows the progressive growth of the dynamically visualized network over time, which achieved the goal of exporting a dynamic, time-based visualization. Given the static nature of printed media, we are only able to display static stills of our product. Figure 3 shows a static, black and white version of a specific Subreddit conversation; while **Figure 4** shows a static colored view of another Subreddit conversation in Gephi.

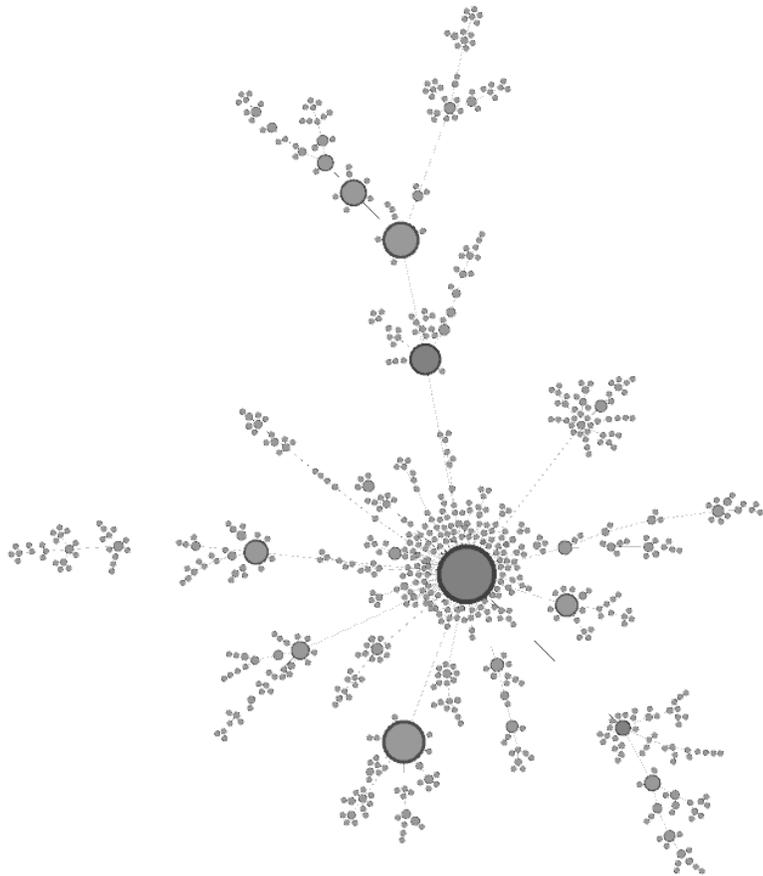


Figure 3. “The Rock” posts first still from “Baywatch” from Subreddit R/movies.

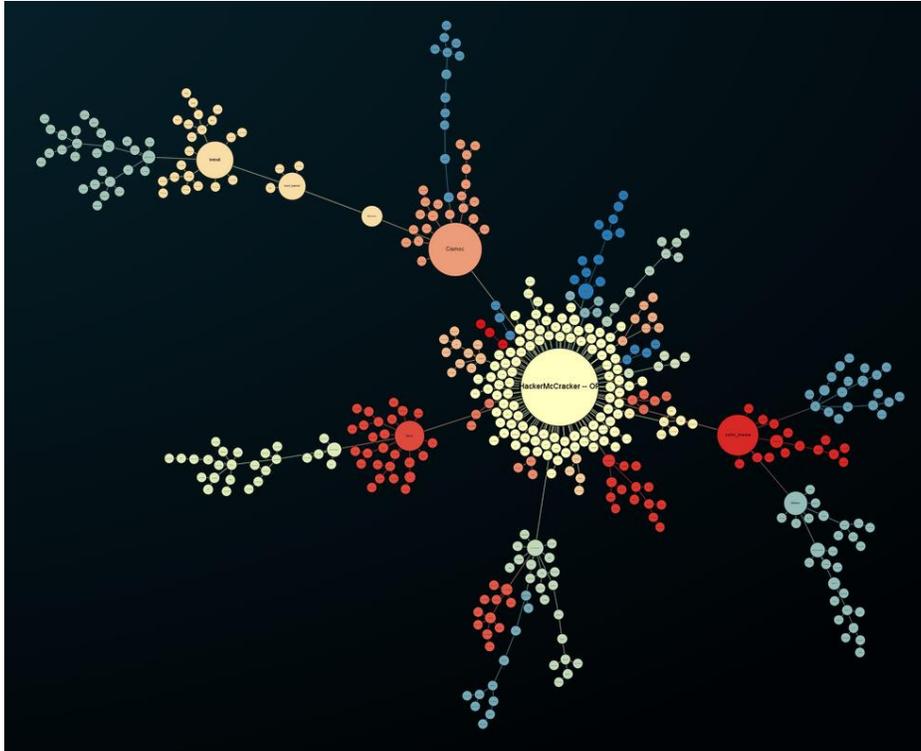


Figure 4. Visualization from Subreddit pictured in screencap from Figure 1.

DISCUSSION

The usual focus of information visualization with regards to social networking sites has generally been the visualization of relationships and connections between individuals within a social group at one moment in time. The Real Time Conversation project developed for this research demonstrates the possibility of *dynamically* visualizing discussions that occur *over time* within a community on social networking sites. The visualizations created from the Real Time Conversation project application met the baseline goals of detailing the growth of both the size and the popularity of discussions on Reddit through a dynamic time-based visualization of comments and their score. When compared to existing static visualizations, dynamic visualization provides additional information regarding the growth of the discussion while retaining a similar sense of the discussion structure.

While Reddit is uniquely formatted to encourage community discussions, the same process of gathering conversational data for visualization using a relational network graph is applicable to other social networking sites, as there are usually systematic manners in which users of social networking sites engage in conversations. At the same time, there are several limitations of gathering conversational data that are specific to Reddit. Since Reddit hides certain comments behind hyperlinks to reduce the need to transfer unnecessary data, a certain amount of data is passed over by the data gathering process; attempts to fill in the gaps required excessive connections to Reddit in order to retrieve the additional data and negatively impacted the efficiency of the data gathering process.

Using Gephi as a visualization system allowed end users to engage with the visualization through the usage of the components of a visualization system; however, a certain amount of knowledge regarding the use of Gephi was necessary to use this visualization system to its full extent. Exporting visualizations from Gephi caused the visualization to lose several interactive components in addition to requiring a significant amount of manual labor. The good news is that the use of Gephi itself for analyzing the time-based images provides a fully interactive environment for users, and the necessary training to use Gephi is quite minimal.

While the focus of this project did not include social network analysis or data analytics, visual analysis of a small sample of ten visualizations did show similarities with regards to the growth of the relational network graph and the spread of popularity among the different nodes. From the static visualizations, it can be seen that the discussions are structured to have between one and five significantly sized branches of conversation, of which, one is significantly larger than the others and may contain a significant branch of its own. Of the nodes in the discussion, the submission itself has the most aggregate score, and the first node of the largest or second largest branch will have the second highest aggregate score. When the dynamic visualization is viewed, it is revealed that the submission itself gains most of its score early on in the discussion's lifetime, but the longer the discussion lasts,

the more likely that the submissions actually begin losing score towards the end of the discussion's lifetime. The comments start gaining in score after the submission has started, however the slowing of the rate at which the submission gains in score does not necessarily affect the comment's aggregate score.

This basic analysis seems to suggest that there may be a systematic trend or underlying method to the progression of discussions that take place within communities on Reddit. Due to medium limitations, only the end-state static visualizations of sample sets have been included in Appendixes. The result of using this visualization to intuitively recognize patterns in the data regarding discussions on Reddit is encouraging, and indicates that there may be an application for the Real Time Conversation project as a tool to assist social scientists in conducting social network analysis or finding specific patterns of interest.

CONCLUSIONS

In this project we defined a need for data visualization and analysis regarding time-based changes to the popularity of online content within the scope of the online social networking site Reddit. With no data readily available for analysis, an application was developed that was capable of logging and exporting time-based changes to content popularity on Reddit for visualization through the information visualization and analysis program Gephi. The visualizations generated by Gephi could be used to analyze the growth of the size and popularity of the discussions over time. It may potentially be applicable to the fields of social psychology and other related social sciences.

On the basis of this project, some additional projects could be pursued. A plugin for Gephi that specializes in extracting and exporting dynamic visualization in animated image and video formats could be beneficial to the graph visualization community in the long run. Additionally, future work should be done to extend the gathering and visualizing of conversational data generated by Reddit in order to allow the visualization to be viewed in a real time, interactive web format that may serve as a companion or alternative to the standard Reddit web user interface. An extension of the data gathering application that does not restrain the data logging to specific cycles, and can continuously gather data, may also assist in providing more complete and additional information for future analysis.

REFERENCES

1. Luse A, Scheibe KP, Townsend AM. A component-based framework for visualization of intrusion detection events. *Information Security Journal: A Global Perspective*. 2008;17(2):95-107.
2. Wasserman S, Faust K. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press; 1994.
3. Christensen M, Jansson A, Christensen C, Christensen M, Jansson A, Christensen C. *Online Territories: Globalization, Mediated Practice, and Social Space*. New York: Peter Lang Publishing; 2011.
4. Garton L, Haythornthwaite C, Wellman B. Studying Online Social Networks. *Journal of Computer-Mediated Communication*. 1997;3(1).
5. Boyd DM, Ellison NB. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*. 2007;13(1):210-30.
6. Scott J. *Social network analysis*. 3rd ed. London: Sage; 2012.
7. Serrat O. *Social Network Analysis*. Asian Development Bank. 2009;28:1-3.
8. Ellison NB, Steinfield C, Lampe C. The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*. 2007;12(4):1143-68.
9. Brandes U, Wagner D. Analysis and Visualization of Social Networks. In: Jünger M, Mutzel P, editors. *Graph Drawing Software*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004. p. 321-40.
10. Kashcha A. Visualizing Communities of Reddit.com 2012 [cited 2016 April 14]. Available from: <http://blog.yasiv.com/2012/07/visualizing-communities-of-redditcom.html>.
11. Olson RS, Neal ZP. Navigating the massive world of reddit: Using backbone networks to map user interests in social media. *PeerJ Computer Science*. 2015;1(4).
12. Quade L. I made a map of Reddit 2013 [cited 2016 April 14]. Available from: https://www.reddit.com/r/MapPorn/comments/o9j6a/i_made_a_map_of_reddit_xpost_from_rpics_1787x2157/.
13. Virdee K. Visualizing discussions on Reddit with a D3 network and Embedly 2013 [cited 2016 April 14]. Available from: <https://blog.embed.ly/visualizing-discussions-on-reddit-with-a-d3-network-and-embedly-e3ac9297bebd#.nblg4hmx4>.
14. project GC. GEXF File Format 2009 [cited 2016 November 22]. Available from: <https://gephi.org/gexf/format/>.
15. Gephi.org. Gephi: The Open Graph Viz Platform 2016 [cited 2016 November 22]. Available from: <https://gephi.org/>.
16. reddit. reddit API documentation 2016 [cited 2016 November 22]. Available from: <https://www.reddit.com/dev/api/>.
17. Boe B. PRAW: The Python Reddit Api Wrapper 2014 [cited 2016 November 22]. Available from: <https://praw.readthedocs.io/en/stable/>.
18. Group PGD. PostgreSQL 2016 [cited 2016 November 22]. Available from: <https://www.postgresql.org/>.
19. Varrazzo D. psycopg 2014 [cited 2016 November 22]. Available from: <http://initd.org/psycopg/>.
20. Microsoft. OpenSeadragon 2016 [cited 2016 November 22]. Available from: <https://openseadragon.codeplex.com/>.
21. PM. Google Maps Exporter 2015 [cited 2016 November 22]. Available from: <https://marketplace.gephi.org/plugin/google-maps-exporter/>.

ABOUT THE STUDENT AUTHOR

Daniel Wang graduated this past May 2016 with an undergraduate degree in Management Information Systems.

PRESS SUMMARY

This research, the Real Time Conversation Project, explores the use of dynamic visualization techniques to visualize and analyze time-based changes in data generated by discussions on the social networking site Reddit.

Effects similar to anxiolysis in an organic extract of *Styopodium zonale* on an anxiety related behavior in *Drosophila melanogaster*

Keysh Mejías^{*a}, Grisel Robles^b, Zulmari Martínez^b, Anamaris Torres^b, Lee Algarín^b, Genesis López^c, Ricardo Chiesa^a

^aDepartment of Biology, University of Puerto Rico, Cayey

^bGeneral Natural Sciences Program, University of Puerto Rico, Cayey

^cDepartment of Chemistry, University of Puerto Rico, Cayey

Students: *keysh.mejias@upr.edu, grisel.robles@upr.edu, zulmari.martinez@upr.edu, anamaris.torres@upr.edu, lee.algarin@upr.edu, genesis.lopez1@upr.edu

Mentor: ricardo.chiesa@upr.edu

ABSTRACT

Nature has a great diversity of organisms whose bioactive compounds may potentially be studied. When it comes to aquatic life we find that algae are organisms that are well suited for screening and identification of bioactive compounds due to their widespread distribution in both salt and freshwater. Our hypothesis is that a crude organic extract of the brown algae *Styopodium zonale* can decrease anxiety-related behaviors in *Drosophila melanogaster*. *Styopodium zonale* was collected in the south coast of Puerto Rico and the potential anxiolytic-like effects of the extract were studied in an anxiogenic-like behavioral paradigm in *Drosophila melanogaster*. This behavior is called centrophobia and is measured using an Open Field Arena (OFA). Validation of the paradigm gave the expected results as reported in the literature, in which *Drosophila* exhibits a phobia (avoidance) of remaining in the center of the OFA, which corresponds to a behavior with anxiety components. The organic extract was dissolved with dimethyl sulfoxide (DMSO). Toxicity tests were performed both for DMSO and the crude organic extract, and neither showed positive results. To perform the behavioral trials, 1 mL of the crude extract and 4 mL of water were mixed with 1.8 g of *Drosophila* food. The final concentration of the crude extract in the food was 5.4 mg/mL. The adult flies were grown in a tube with the extract until a considerable quantity of larvae was observed, and then the adults were removed. These new larvae, once turned into adult flies, were used for the behavioral trials. The behavior of control flies (food without extract) and experimental flies (extract-containing food) was recorded with a video camera and the results of the centrophobic behavior were analyzed and compared using quantitative criteria. Both the control and experimental trials were performed in triplicate. The results show that flies grown in food containing the crude extract present a significant reduction in centrophobia compared with control flies. In conclusion, our results suggest that the organic crude extract from *Styopodium zonale* has anxiolytic-like effects in a *Drosophila melanogaster* model with anxiety components. We are currently performing Nuclear Magnetic Resonance (NMR) studies on the crude extracts to identify the most abundant secondary metabolites. Future experiments should include the administration of the crude extracts (or fractions of the most abundant secondary metabolites) to a vertebrate model in to test the effect in a behavior with anxiety components. We are also in the process of developing a preliminary model of possible mechanisms of action of the crude organic extract in the reduction of centrophobia.

KEYWORDS

Anxiety; Algae; *Drosophila melanogaster*; Open Field Arena; Organic extracts; Centrophobia; *Styopodium zonale*

INTRODUCTION

Algae are aquatic organisms that can be found both in fresh and salt waters. They are classified into classes according to their color, in which three principal groups can be defined: the brown algae (class *Phaeophyceae*) the green algae (class *Chlorophyceae*), and the red algae (class *Rhodophyceae*). In this project, we worked with the *Phaeophyceae*, specifically the alga *Styopodium zonale*, found in the coasts of Puerto Rico. For classification purposes, we followed descriptive and structural criteria described in the book *Caribbean Reef Plants*.¹

In the world, there are a total of 1500 species of brown algae;² of these, 65 are found in Puerto Rico. In this project, we used the brown algae *Styopodium zonale*, which is easily found in the south coast of Puerto Rico, specifically at Pozuelo Beach in the town of Guayama and Punta Guilarte Beach in the town of Arroyo.

Algae contains a great variety of bioactive compounds, which is why it is being studied and used for the development of alternative pharmacotherapies to treat different health conditions.³ A great diversity of organic compounds with an ample spectrum of biological activity have been isolated from marine algae in the last decades⁴ including compounds with anti-inflammatory capacities, antioxidant activity, and antimicrobial activity, among others.⁵ More recently, algal compounds with neuroprotective abilities have also been reported.⁶ Since neuroprotection is an inherent requisite for proper neuronal connectivity in any circuit, including neuronal circuits related to emotions and current drugs for the treatment of anxiety disorders (such as benzodiazepines) have the potential to generate tolerance, dependence, and addiction, it is important to search for alternative chemical agents for the pharmacotherapy of anxiety disorders. Moreover, extracts of the brown seaweed *Ecklonia cava* evoke anxiolytic-like effect in rats.⁷ These findings set the rationale for this project.

The aim of this project was to assess the potential anxiolytic effects of crude organic extracts obtained from brown algae (class Phaeophyceae) as an alternative pharmacotherapy for the treatment of anxiety disorders. Specifically, we aimed to study the effects of a crude organic extract of the brown algae *Styopodium zonale*, the only species of the genus *Styopodium* in Puerto Rico. We used the fruit fly *Drosophila melanogaster*, as the model system for the behavioral studies. *Drosophila* has been used for decades as a model system for genetics, and more recently it has emerged as a model for the study of neurobiological disorders.⁸ *D. melanogaster* is neuroanatomically different from humans, but it is also very similar regarding neuronal physiology and neurotransmitters systems. These similarities make *Drosophila* a versatile and ideal model for the study of behavior and neuropsychiatric disorders. Some examples of the similarities are the neurotransmitter systems, learning behaviors, and memory.⁹ In addition to these similarities in cerebral physiology with the human brain, *Drosophila* are easily maintained and grown under controlled laboratory conditions.

The paradigm used to generate an anxiety behavior in *Drosophila* in this project was centrophobia, an innate aversive behavior.^{10,11} Centrophobia is observed with the use of an Open Field Arena (OFA).¹² The centrophobic behavior can be observed when *Drosophila* is placed in the center of the OFA and it rapidly moves to the borders, avoiding staying in the middle of the arena.^{13,14} The organic crude extract of *Styopodium zonale* was administered in the food from the embryo stage until adulthood, and these adults were used for the behavioral assays.

In summary, the experimental hypothesis is that an organic crude extract of the brown algae *Styopodium zonale* will reduce centrophobia in *Drosophila* (measured in an OFA). Another important goal of this project is to chemically characterize the secondary metabolites that could be responsible for the anxiolytic effects of the crude extract. A long-term goal of this project is to administer the crude *Styopodium zonale* extract and/or purified metabolites to vertebrates, such as rats, in specific brain regions related to anxiety and fear, to assess for any anxiolytic effects.

MATERIALS AND METHODOLOGY

Collection, Storage, and Classification of Brown Algae

Two collections of *S. zonale* were made for this project. The first collection of algae was in Pozuelo Beach, in the town of Guayama, Puerto Rico. The second collection was made in Punta Guilarte Beach at Arroyo, Puerto Rico. For both collections, the algae were stored in plastic bags with seawater and were transported to the laboratory surrounded with ice. They were rinsed with filtered sea-water, blotted and air-dried, classified, and stored at -20°C.

Extraction of the Organic Compounds

Two extractions of the algae *Styopodium zonale* were made, one in October 2014 and the other in September 2015. To carry out the two extractions, an initial quantity of 150 grams and 35 grams respectively were used to obtain a crude extract yield of 2.20 grams (2,200mg) and 0.27 grams (270 mg) of organic solids. The organic solids obtained from each extraction were dissolved in a total volume of 10 mL dimethyl sulfoxide (DMSO) 23% and 10%, respectively. The step-by-step process for the organic compound extraction was as follows: algae were crushed in a mortar, triturated in a blender with a mix of dichloromethane and methanol in a 1:1 proportion. After filtering in a vacuum using a flask, a funnel, two filter papers, and celite sand, the filtrate was placed in a distillation flask to evaporate the liquid in a rotary evaporator. The organic solids obtained from the second extraction, dissolved in 10 mL of DMSO 10%, were used for the experimental group of the behavioral assays. The concentration of crude extract in the total volume of 10 mL of DMSO 10% was 27 mg/mL.

Toxicity Assays with Dimethyl Sulfoxide (DMSO) at 23%

In order to perform a toxicity assay for DMSO at its highest concentration (23% DMSO used in the first organic extraction), 16 flies were placed in a tube. This tube contained about 2 mL of bactoagar 2% and sucrose 5%, with DMSO at 23%, which solidified at the bottom. The flies were placed under observation for 24 hours. In the first extraction, we were able to dissolve the solids of the crude extracts with a 23% concentration of DMSO. Although for the second extraction we obtained complete

dissolution of the solids in the crude extracts using 10% of DMSO, we tested the toxicity of DMSO at its highest concentration of 23%, for a better assessment of its potential toxic effects.

Preparation of OFA for Assessment of Centrophobia

The Open Field Arena (OFA) was built using two acrylic concentric circles. To cover the circles, a square acrylic top was used. Air flowed into the circles through four small spaces that remained in each border between the top acrylic cover and the circles. To hold the camera used for the behavioral recordings, a polyvinyl chloride (PVC) tripod was constructed. It consisted of two PVC tubes, one horizontally oriented and the other one vertically oriented. The one placed vertically contained a screw in the superior part to secure the camera, making it stable. Lastly, two 15 Watt flood lights were placed near the OFA, to maintain uniform photo-taxis in *Drosophila melanogaster* throughout the assay.

Control Assays for the Validation of Centrophobia

To perform the validation of this paradigm, a triplicate was made, for a total sample size (n) of 35 flies. These flies grew eating regular food, which contains 1.8 g of Formula 4-24 food (Carolina Biologicals; proprietary mixture) and 5 mL of water. The flies were immobilized by exposing them to a temperature of -20°C for five minutes. After this, they were moved to the Open Field Arena. As soon as the first fly awoke, their behavior was recorded for 10 minutes without interruption. The criteria used to determine the behavior of centrophobia was the time (in seconds) that the flies remained in the center of the OFA.¹⁵ As expected, the video recordings showed that once the flies awakened, they immediately began moving towards the borders of the OFA, avoiding staying in the center.

*Toxicity Assay with the Organic Extract of *Styopodium zonale**

To assess the potential long-term toxicity of the crude organic extract of *Styopodium zonale*, 1.8 g of Formula 4-24 food were mixed with 4 mL of water and 1 mL of the crude extract (at a concentration of 27 mg/mL). The final concentration of the crude extracts in the 5 mL of food was 5.4 mg/mL. The general behavior and sexual behavior of the flies were observed during nine days.

*Experimental Centrophobia Assays with the Organic Extract of the Algae *Styopodium zonale**

For the experimental behavioral assays, and to assess the potential anxiolytic effects of the crude extract, a triplicate was made, for a total sample size (n) of 41 flies. The adult flies used had been exposed to 5 mL of food containing the crude extract at a concentration of 5.4 mg/mL since embryo stage. The assays were performed as described in *Control Assays for the Validation of Centrophobia*. The video recordings showed that flies exposed to the crude extract remained in the center of the OFA for a longer period of time than the control flies. With this, it was concluded that the crude organic extract of the algae *Styopodium zonale* has anxiolytic effects on this particular aversive behavior in *Drosophila melanogaster*.

*Chemical Analysis of the Crude Extract from *Styopodium zonale**

After evaporation of the solvent mixture, the flask containing the crude extract from *S. zonale* was placed in a vacuum for one hour to remove any remaining traces of solvents. A spatula was used to obtain a small sample of the extract, place and dissolve it in an NMR vial containing 0.5 mL of CDCl₃ (deuterated chloroform). Once this was done, the vial was transferred to a Bruker Avance NMR spectrophotometer to obtain the ¹H-NMR (400 MHz) and ¹³C-NMR (100 MHz) data.

*Statistical analysis of the effects of the crude extract from *Styopodium zonale**

Results obtained from both the validation of centrophobic behavior assay and the experimental centrophobic assay with the extract were analyzed using the program Logger Pro 3. The mean and median were obtained for both assays (control and experimental) and their means were compared for statistical significance using a one-tailed student t-test.

RESULTS

Organic Extractions

Two extractions were made with the *Styopodium zonale* algae with a yield of 220 mg/mL and 27mg/mL, respectively.

Toxicity Assays with Dimethyl Sulfoxide (DMSO) at 23%

In the DMSO toxicity assay, no convulsions or deaths were observed. This implies that the DMSO 23% is not toxic for *Drosophila melanogaster*.

*Toxicity Assay with the Organic Extract of *Styopodium zonale**

For the crude organic of *S. zonale* toxicity assay no changes in mating, copulation, oviposition, and development into adulthood were observed. This indicates no long-term toxicity of the crude extract to the *Drosophila melanogaster*.

Validation of Anxiety Behavior (Centrophobia)

The anxiety behavior of centrophobia in *Drosophila melanogaster* when placed in an OFA was successfully validated. The behavior in the OFA of flies exposed to regular food (**Table 1A**) was registered and compared with the behavior of flies exposed to the crude extract (**Table 1B**). The mean latency of the flies (time spent in the center of the OFA) was 265 seconds with a standard error of 212 seconds (**Figure 2A**). The median latency was 236 seconds (**Figure 2B**).

Behavioral Assays to Assess the Possible Anxiolytic Effects of the Organic Crude Extract of *Styopodium zonale*

The flies grown in food containing the crude organic extract of *S. zonale* showed a considerable and statistically significant decrease in centrophobia, remaining for longer times in the center of the OFA. The mean latency of the flies was 467 seconds with a standard error of 180 seconds (**Figure 2A**). The median latency was 600 seconds (**Figure 2B**).

Chemical Analysis of the Crude Extract from *Styopodium zonale*

The crude extract of *S. zonale* was analyzed with Nuclear Magnetic Resonance (NMR), which is a powerful technique for the identification of plant metabolites. Analysis of metabolites using NMR is a fast and successful tool to determine the most important regions of the spectrum for further analysis. The NMR analyzed in this study was ^1H -NMR and ^{13}C -NMR. In the ^1H -NMR (**Figure 1A**), signals were present of vinylic protons ($\delta=5.2$ - 5.4 ppm), α -heteroatoms ($\delta=3.0$ - 3.5 ppm), allylic protons ($\delta=1.5$ - 2.0 ppm), and aliphatic protons ($\delta=0.5$ - 1.5 ppm). Meanwhile, in the ^{13}C -NMR (**Figure 1B**), signals of vinylic carbons ($\delta=130$ ppm) and aliphatic carbons ($\delta=14$ - 32 ppm) were present. These two signals are shown only in the ^{13}C -NMR spectrum because of the time the extract was analyzed. The ^1H -NMR and the ^{13}C -NMR spectrums showed signals that are present in cyclic organic compounds. This is common to find in natural compounds, such as terpenoids, according to the scientific literature.¹⁶ We are currently working in the purification and characterization of the crude extract.

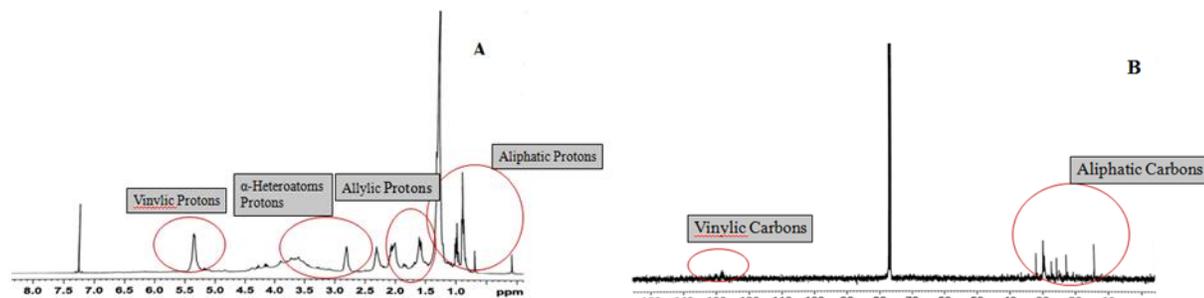


Figure 1. NMR spectra of the crude extract using CDCl_3 as a solvent. A. ^1H -NMR (400 MHz) spectrum. B. ^{13}C -NMR (100 MHz) spectrum.

DISCUSSION

The objective of toxicity assays was to assess the possibility that the solvent (DMSO) or the *S. zonale* organic extract could affect adult or developing flies. First, a toxicity trial of DMSO 23% was done, since it was the highest percentage used to dissolve the organic solids. Flies exposed to 23% DMSO behaved normally and did not convulse or die. Then we assessed the potential toxicity of the crude organic extract of *S. zonale* on a long-term perspective. Adult flies courted and copulated normally, and embryos developed into normal and sexually active adults.

Centrophobia in *Drosophila* was validated using flies exposed to regular food ($n = 35$) and then placed in an OFA. The expected avoidance behavior in which the flies move quickly to the walls of the OFA was observed, recorded, and quantified. These flies used for validation of the paradigm also served as the control group.

Table 1A	
Fly Number	Time (s)
1	30
2	34
3	68
4	77
5	77
6	84
7	88
8	88
9	137
10	143
11	160
12	172
13	177
14	188
15	210
16	217
17	217
18	256
19	600
20	600
21	100
22	111
23	128
24	131
25	135
26	160
27	279
28	434
29	589
30	600
31	600
32	600
33	600
34	600
35	600

Table 1B	
Fly Number	Time (s)
1	107
2	116
3	143
4	152
5	172
6	246
7	247
8	249
9	251
10	264
11	271
12	271
13	334
14	450
15	475
16	494
17	558
18	585
19	600
20	600
21	600
22	600
23	600
24	600
25	600
26	600
27	600
28	600
29	600
30	600
31	600
32	600
33	600
34	600
35	600
36	600
37	600
38	600
39	600
40	600
41	600

Table 1. Time spent by the flies in the center of the OFA. **A.** Flies exposed to regular food and the time each spent (in seconds) in the center of the OFA. Time was measured from the moment the fly moved from the center towards the walls of the OFA. The mean obtained was 265 seconds with a standard deviation of 212 seconds. **B.** Flies exposed to food with algae extract (5.4 mg/mL) and the time each spent (in seconds) in the center of the OFA. Time was measured from the moment the fly moved from the center towards the walls of the OFA. The mean obtained was 467 seconds with a standard deviation of 180 seconds.

Once the centrophobic behavior of control flies was validated, the assays for assessing the potential anxiolytic effects of the crude organic extract of *S. zonale* were performed. The flies used for these assays ($n = 41$; **Table 1B**) were exposed to food containing 5.4 mg/mL of the crude organic extract since the embryo stage until adulthood. When compared to control flies ($n = 35$; **Table 1A**), these treated flies showed a dramatic and statistically significant increase in time spent in the center of the OFA, which translates into a decrease in the centrophobic behavior and indicates an anxiolytic effect of the *S. zonale* organic extract. A quantitative and statistical analysis of the results obtained for both groups (control and treated flies) was performed and is shown in **Figure 2A** (mean, t-test, p-value) and **Figure 2B** (median). Based on our toxicity assays, DMSO at the concentrations used in this project is not toxic for the flies. Moreover, there are no reports in the literature of DMSO affecting the behavior of flies on an OFA. In addition, DMSO does not have any effect on the behavior of flies in the dark/light box test or on flies acutely exposed to ethanol. These are two other behavioral and motor activity paradigms we are currently working on in our laboratory.

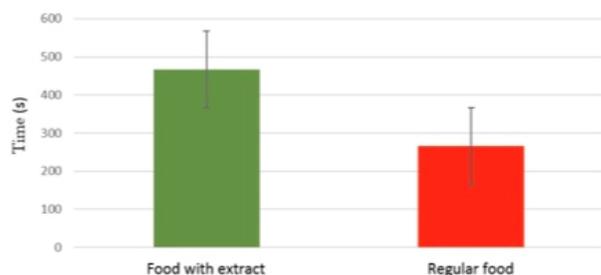


Figure 2A.

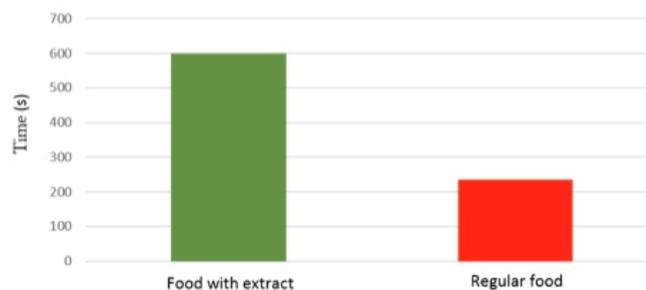


Figure 2B.

Figure 2A. Bar graph of the mean time spent in the center of the OFA by flies exposed to food containing 5.4 mg/mL of the organic extract of *S. zonale* (green bar; mean = 467 seconds, S. E. = 180 seconds) and flies exposed to regular food (red bar; mean = 265 seconds, S. E. = 212 seconds), p-value = 0.000013, which is significant at $p < 0.05$, one-tailed student t-test.

Figure 2B. Bar graphs of the median of the time spent in the center of the OFA by flies exposed to food containing 5.4 mg/mL of the organic extract of *S. zonale* (green bar; median = 600 seconds) and flies exposed to regular food (red bar; median = 236).

CONCLUSIONS

As expected, *Drosophila melanogaster* grown on regular food (control flies) placed in an OFA show centrophobic behavior. Our results support our experimental hypothesis that flies exposed since the embryo stage to a crude organic extract of the brown algae *S. zonale* show a dramatic and significant decrease in centrophobic behavior. Regarding future projects, protein extracts of *S. zonale* and organic extracts of *Sargassum polyceratum* (another brown algae) will be tested for their potential anxiolytic effects. In addition, we plan to perform the assays by sex (females only and male only) to assess for differences in behavior and response to the *S. zonale* extract due to variations in the hormonal milieu between sexes. Finally, we aim to perform these experiments, in a vertebrate model, such as rats, to assess the anxiolytic effects of the *S. zonale* crude organic extract or purified metabolites using other anxiogenic and fear-inducing paradigms.

ACKNOWLEDGMENTS

Our gratitude to Dr. Claudia Ospina, of the Chemistry Department of the UPR-Cayey, for the collaboration in the organic extraction procedure and chemical characterization of the *S. zonale* crude extract. This project was funded by seed money from the FIDI (Institutional Funds for the Development of Research) program, the Dean of Academic Affairs, and from seed money from the Institute of Interdisciplinary Investigations and the BRIC Program of UPR-Cayey. Last, but not least, we give our gratitude to the rest of the research team, whose hard work and dedication made this research project possible.

REFERENCES

1. Littler DS and Littler MM. (2000). Caribbean Reef Plants. OffShore Graphics, Inc. Washington, D. C. ISBN 0-9678901-0-1.
2. Raven P, Evert R, Eichhorn S. (1999) *Biology of plants*. W.H. Freeman and Company/ Worth Publishers. 944.
3. Shalaby, E. (2011) Algae as promising organism for environment and health. Landes Bioscience.
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3258063/pdf/psb0609_1338.pdf
4. Matsubara K, Matsuura Y, Hori K, and Miyazawa K. (2000). An anticoagulant proteoglycan from the marine green alga, *Codium pugniformis*. J. Appl. Phycol. 12: 9–14.
5. Wijesekara I, Yoon N, and Kim S. (2010). Phlorotannins from Ecklonia cava (Phaeophyceae): Biological activities and potential health benefits. BioFactors, 30(6): 408–414.
6. Pangestuti R, Kim S. (2011) Neuroprotective effects of marine algae. *Mar. Drugs*, 9: 803-818
7. Suengmok C, Daeseok H, Seon-Bong K, et al. (2012) Depressive Effects on the Central Nervous System and Underlying Mechanism of the Enzymatic Extract and Its Phlorotannin-Rich Fraction from Ecklonia cava Edible Brown Seaweed. *Biosci. Biotechnol. Biochem.*, 76(1): 163-168
8. Iliadi K. (2009) The genetic basis of emotional behavior: has the time come for a Drosophila model? *Journal of Neurogenetics*, 23: 136-146.
9. O'Kane C. (2011) Drosophila as a model organism for the study of neuropsychiatric disorders. *Pub Med*, 7: 37-60.
10. Götz K, Biesinger R. (1985) Centrophobism in Drosophila melanogaster. *Journal of Comparative Physiology A*, 156(3): 319-327.
11. Crawley J. (1985). Exploratory behavior models of anxiety in mice. *Spring*, 9(1):37-44.
12. Roman G, Soibam B, Goldfeder R, Manson-Bishop C, et al. (2012) Modeling Drosophila positional preferences in open field arenas with directional persistence and wall attraction. *Plos One*, 7(10): e46570
13. Besson M, Martin J. (2005) Centrophobism/thigmotaxis, a new role for the mushroom bodies in Drosophila. *Journal of Neurobiology* 62: 386–396.
14. Gotz K (1994) Exploratory strategies in Drosophila. *Basic Life Sci* 16: 391–407.
15. Valente D, Golani I and Mitra P. (2007) Analysis of the Trajectory of Drosophila melanogaster in a Circular Open Field Arena. *Plos One* 2(10): e1083. DOI: 10.1371/journal.pone.0001083
16. Dorta, E.; Cueto, M.; Brito, I.; Darias, J. (2002). New Terpenoids from the Brown Alga *Styopodium zonale*. *Nat. Prod.* 65, 1727-1730.

ABOUT THE STUDENT AUTHORS

Keysh Mejías Martínez is a senior in the Biology Bachelors Program at the University of Puerto Rico at Cayey. She has been working on the article since its beginning and it has been a great experience for the development of her scientific abilities. One of the most satisfying parts of the project was presenting preliminary results at the 56th Annual Drosophila Conference in Chicago in March 2015. Her next goal is to earn an MS Degree in Neurobiology and eventually study medicine.

Grisel Robles Del Valle studies at the University of Puerto Rico at Cayey, and is currently earning a bachelors in the General Natural Sciences program. She has been working on Dr. Ricardo Chiesa's research project for two years. This has helped her to decide on pursuing a Ph.D. degree in a field related to neurobiology. One of her most gratifying experiences was to travel to Chicago for a poster presentation at the 56th Annual Drosophila Conference, in March 2015. She is also on the board of directors to the student organization called "Clinical Laboratory Professionals", also known as T-MED and is a member of the Microbiology Society of Puerto Rico.

Zulmari Martinez Resto is currently in her last semester of her bachelor's at the University of Puerto Rico at Cayey. She is part of the General Natural Science Program and five other student organizations. She has been part of Dr. Ricardo Chiesa's research team since the beginning of the project and thanks to this great experience, she was able to travel to Chicago for the 56th Annual Drosophila Conference in March 2015. During summer 2014, she participated in Dr. Garcia Rinaldi's Medicine Internship. All of these experiences made her major goal to become a neurosurgeon.

Anamaris Torres Sánchez is currently on her fourth year in the General Natural Science Bachelors Program at the University of Puerto Rico at Cayey. She has been on three different research experiences as an undergraduate. She is also a historian for the $\beta\beta\beta$ National Biological Honor Society Z-E Chapter and is part of three other student organizations such as MEDLIFE. She aspires to pursue graduate studies in nursing and medicine.

Lee Juan Algarín Sepúlveda is a senior in the General Natural Sciences Bachelor's Program at the University of Puerto Rico, in Cayey. In the interest of pursuing higher goals in the fields of research and medicine, he is currently preparing himself by experiencing the life of an undergraduate research student. His biggest desire is to obtain an MD-PhD degree in a neurobiology-related field. He is a member of the American Medical Student Association (AMSA) and $\beta\beta\beta$ National Biological Honor Society, Z-E Chapter.

Genesis López Feliciano is a student at the University of Puerto Rico at Cayey who is doing her bachelor's degree in the Natural Sciences Program with a major in Chemistry. López is working in the Organic Laboratory with her mentor, Claudia Ospina, Ph.D. Currently, López is working in the chemical analysis of the brown algae *Styopodium zonale* to isolate new and potentially versatile bioactive compounds. During her bachelor's degree, López has participated in two different research experiences. She is a member of student organizations such as "Círculo de Química"-ACS Chapter, American Society of Biochemistry and Molecular Biology (ASBMB) and Legacy Professional Association.

PRESS SUMMARY

Nature has a great diversity of organisms with the potential to study their bioactive compounds. When it comes to aquatic life we find that algae are organisms that are understudied, but this does not recognize the capacity for them to be used as potential sources of bioactive compounds with therapeutic capabilities. The hypothesis for this project is that a crude organic extract of the brown algae *Styopodium zonale* can decrease anxiety behaviors in *Drosophila melanogaster*. *Styopodium zonale* was collected in the south coast of Puerto Rico. The anxiolytic effects of the extract were studied in an anxiogenic behavioral paradigm in *Drosophila melanogaster*. To generate an anxiety behavior in *Drosophila*, we used centrophobia in an Open Field Arena (OFA) as the behavioral paradigm. Validation of the paradigm gave the expected results as reported in the literature, in which *Drosophila* exhibits a phobia (avoidance) of remaining in the center of the OFA, which corresponds to an anxiety behavior. The organic extract was dissolved with Dimethyl Sulfoxide (DMSO) 10%. Toxicity trials were performed, and the results obtained eliminated the possibility of any toxicity by the DMSO and the crude extracts. To perform the behavioral trials, the crude extract was placed in 5 ml of water at a concentration of 5.4 mg/ml as a solution to be mixed afterward with 1.8 g of *Drosophila* food. The adult flies were placed in a tube with the extract until a considerable quantity of larvae was observed, and then the adults were removed. These newborn larvae, once turned into adult flies, were used for the behavioral trials. These trials were filmed with a video camera in order to analyze the results of quantitative measurements. After performing the behavioral trials in a triplicate (flies exposed and not exposed to the extract), a comparison of the results was made. Flies grown in food containing the crude extract showed a significant reduction in centrophobia when compared with flies who grew exposed to regular food (control flies). In conclusion, our results suggest that the organic crude extract from *Styopodium zonale* has anxiolytic effects in an anxiety model of *Drosophila melanogaster*.

