

(ur) DIMENSIONS

The Journal of Undergraduate Research in Natural Science and Mathematics
California State University, Fullerton

VOLUME 14 / SPRING 2012

MARKS OF A CSUF GRADUATE FROM THE COLLEGE OF NATURAL SCIENCES AND MATHEMATICS

Graduates from the College of Natural Sciences and Mathematics:

Understand the basic concepts and principles of science and mathematics.

Are experienced in working collectively and collaboratively to solve problems.

Communicate both orally and in writing with clarity, precision and confidence.

Are adept at using computers to do word processing, prepare spreadsheets and graphs, and use presentation software.

Possess skills in information retrieval using library resources and the Internet.

Have extensive laboratory/workshop/field experience where they utilize the scientific method to ask questions, formulate hypotheses, design experiments, conduct experiments, and analyze data.

Appreciate diverse cultures as a result of working side by side with many people in collaborative efforts in the classroom, laboratory and on research projects.

In many instances have had the opportunity to work individually with faculty in conducting research and independent projects. In addition to the attributes of all NSM students, these students generate original data and contribute to the research knowledge base.

Have had the opportunity to work with very modern, sophisticated equipment including advanced computer hardware and software.

DIMENSIONS

Dimensions: The Journal of Undergraduate Research in Natural Sciences and Mathematics is an official publication of California State University, Fullerton. Dimensions is published annually by CSUF, 800 N. St. College Blvd., Fullerton, CA 92834. Copyright ©2012 CSUF. Except as otherwise provided, Dimensions: The Journal of Undergraduate Research in Natural Sciences and Mathematics grants permission for material in this publication to be copied for use by non-profit educational institutions for scholarly or instructional purposes only, provided that 1) copies are distributed at or below cost, 2) the author and Dimensions are identified, and 3) proper notice of the copyright appears on each copy. If the author retains the copyright, permission to copy must be attained directly from the author.

ABOUT THE COVER

Just like in science, our curiosity offers new perspectives which reveal intricacies filled with beauty and intricacy. Our world is a cycle of creation and destruction. Nothing escapes the mighty power of science. Even mundane objects which, at first appear simple, reveal mechanisms which have evolved from a primary conceived human idea. Anything under the proper light can appear beautifully complex.

ACKNOWLEDGEMENTS

EDITOR-IN-CHIEF

Amber Shah

EDITORS

Christopher Baker, Geology
Manar Adem, Biology
Peter Ho, Physics/Mathematics

ADVISORS

Dr. Rochelle Woods, Assistant Dean for Student Affairs

GRAPHIC DESIGN

Brian Huezo, Cover Designer
Courtney Kim, Layout Editor

COLLEGE OF THE ARTS

Dr. Joseph H. Arnold, Jr., Dean
Andi Sims, Assistant Dean for Student Affairs
Chen Wang, Graphic Design Professor
John Drew, Graphic Design Professor

COLLEGE OF NATURAL SCIENCES & MATHEMATICS

Dr. Robert A. Koch, Acting Dean
Dr. Mark Filowitz, Associate Dean
Dr. Kathryn Dickson, Acting Chair Department of Biological Science
Dr. Christopher Meyer, Chair Department of Chemistry
Dr. David Bowman, Chair Department of Geological Sciences
Dr. Paul De Land, Chair Department of Mathematics
Dr. James M. Feagin, Chair Department of Physics

SPECIAL THANKS TO

President Milton A. Gordon and Dean Robert A. Koch
for their support and dedication to *Dimensions*.

ARTICLES

06

Hydrogeologic Investigation of the Sheep Creek Fans Regional Aquifer and Chromium's Origins and Impacts on Water Quality, San Bernardino County, CA
Erik M. Cadaret, *Geological Sciences*

31

Measuring Xylem Vulnerability in Three Citrus Species
Emily Wieber, *Biological Sciences*

07

Evidence for Glaciovolcanism at Long Canyon Dome, Sierra Nevada, California
John F. J. Hennessy, *Geological Sciences*

37

Maximum Likelihood Estimation of the Fat Fraction Using Magnetic Resonance Imaging at High Signal-to-Noise Ratio
Jorly Chatouphonexay, *Mathematics*

14

Measuring debris-flow timing relative to rainfall in burned watersheds: Implications for residential evacuations and early warning
Robert J. Leeper, *Geological Sciences*

45

Assessing Uncertainty of Clustered Neuronal Intensity Curves
Mikhail Y. Popov, *Mathematics*

19

A Magma Mixing Origin for Andesites Erupted in 2006 from Augustine Volcano, Alaska
Ashley Melendez, *Geological Sciences*

53

An Inverse-Based Analogue of the Probability That Two Elements in a Finite Group Commute
Nick Blackford, Daniel Lenders, Danny Orton, *Mathematics*

24

Fault Scarp Morphology Along The Northern Eureka Valley Fault Zone, Eureka Valley, Eastern California, U.S.A.
Ernest Nunez, *Geological Sciences*

59

To Err is Human, to Trisect Divine...
Jonathan Sayre, *Mathematics*

25

Profile of energy delivered to chicks by the Elegant Tern (*Thalasseus elegans*) over a recent 16-year period (1994-2009) at the Bolsa Chica Ecological Reserve
Laura Tolentino, *Biological Sciences*

64

Biographies

Hydrogeologic Investigation of the Sheep Creek Fans Regional Aquifer and Chromium's Origins and Impacts on Water Quality, San Bernardino County, CA

Department of Geological Sciences, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

Erik M. Cadaret

Advisors: Dr. W. Richard Laton and Dr. John H. Foster

Abstract

The Sheep Creek Fan is located in the Mojave Desert between the towns of Wrightwood and El Mirage dry lake. The hydrogeology of the fan was modeled in 3D to determine what areas dictated by the Mojave Water Agency 2.4 km (1.5 mi) buffer zone from the California aqueduct is favorable for artificial recharge. The sand vs. clay ratios created by the subsurface lithologies being categorized as sand, silt, and clay indicated that the southern middle portion of the fan is suitable for recharge. Determining a suitable recharge location prompted further investigation on the recharge water impacts on water quality. Chromium impacts the water quality and the concentrations were graphed to determine the concentration versus distance traveled from the source. The chromium source is likely from the mineral actinolite from the Pelona Schist and it was found to be in high concentrations in sediments at the base of the Wrightwood watershed in Sheep Creek and decreased with distance from the source. The correlation between our sediment samples and the concentrations of chromium in water is an inverse relationship.

Evidence for Glaciovolcanism at Long Canyon Dome, Sierra Nevada, California

Department of Geological Sciences, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

John F. J. Hennessy

Advisor: Dr. Brandon Browne

Abstract

Glaciovolcanic deposits are deposits that form during the interaction of lava with glacial ice and/or melt water. Pyroclastic deposits that form under these conditions can be recognized and distinguished from those subaerial deposits that form under surface conditions by their unique textures including, but not limited to: jig-saw brecciated textures, fine-grained pumice and ash clast sizes, poorly vesicularity morphology, and fractured crystal textures. It is also well-documented that subglacial rhyolites observed in Iceland and other glaciovolcanic settings contain aggregates of fine material that accumulate on the surface of glassy pumice fragments, especially within vesicle hollows. The presence of these features and more within volcanoclastic deposits have often been used to argue that the deposits occurred from subglacial eruptions. Therefore, some macroscopic structures have also been used to determine the extent of localized glaciation during eruptions. The stratigraphy of the Long Canyon Dome (LCD) rhyolites, situated in the southern Sierra Nevada mountain range, California, suggests that these deposits are of a glaciovolcanic origin based on their textural characteristics. The presented textures and structural features of these units are similar to those described in Icelandic rhyolites and other subglacial rhyolite deposits. Few publications on the LCD rhyolite are known, thus very little has been documented about this location. This study pioneers the textural analysis of the LCD deposits and describes microscopic features that have been compared with other known glaciovolcanic deposits to argue in favor of a subglacial origin.

Introduction

Although there are many publications on lithofacies analysis of glaciovolcanic and phreatomagmatic deposits in Iceland (Furnes et al. 1980; Heiken and Wohletz 1985; Gudmundsson et al. 1997; Tuffen et al. 2001; Stevenson et al.

2006; McGarvie 2009) and elsewhere (Heiken and Wohletz 1985; Lescinsky and Fink 2000; Lodge and Lescinsky 2009), very little analysis regarding possible lava-ice interactions has been done in the southern Sierra Nevada of California. Long Canyon Dome (LCD) is situated atop the Kern Plateau of the Sierra Nevada mountain range in eastern California, approximately 10 km west of Olancho Peak surrounded by 3 proximal rhyolitic domes: Kern Peak, Templeton Mtn, and Monache Mtn. (Figure 1).

The rhyolites of LCD display unique textures that are comparative to textures described in glaciovolcanic deposits and may have been ejected beneath glacial confines. Field and laboratory analysis of the LCD rhyolites suggests the erupted material came into contact with glacial ice and/or melt water, as their microtextures indicate flash-cooling and quenching along lava-ice contact margins. In addition, the deposits have bimodal distributions of grain sizes: large, smooth, blocky glasses and nano- to micron-sized platy glass particles that typically accumulate to form aggregates on the large

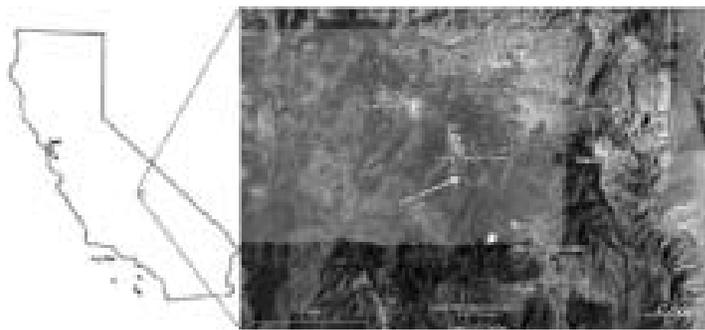


Figure 1 – Google Earth image showing location of Long Canyon Dome (target marker with white arrow pointer) in the southern Sierra Nevada mountain range approximately 10km west of Olancho Peak and within proximity of 3 other rhyolitic domes: Kern Peak, Templeton Mtn, and Monache Mtn.

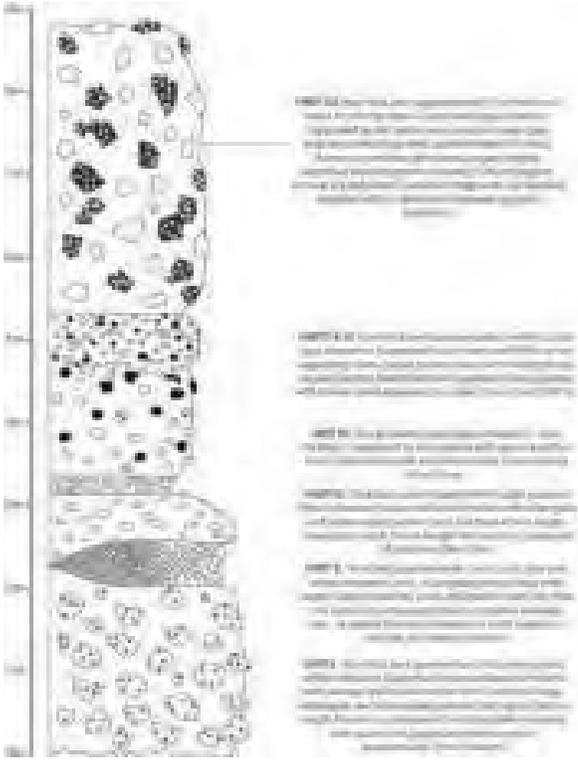


Figure 2 – Stratigraphic column of the exposed outcrop flanking the northwestern edge of Long Canyon Dome.

clasts, especially in vesicle voids – indicative of subglacial or phreatic tephra (Heiken and Wohletz 1985; Stevenson et al. 2006). Therefore, it is likely that perennial alpine glaciation had extended to this region of California during the Pleistocene and further geologic mapping of the deposits may aid in assessing the paleoclimate of this region at the time of the LCD eruption. If alpine glaciation was prevalent during eruptions of the other southern Sierra domes, many of the deposits in this region may have been erroneously described as subaerial and further investigation of the extent of glacial influence on these deposits may need to be considered.

Geological setting

LCD is located at the northern head of Long Canyon near 3 other rhyolitic domes, which yield eruption ages of ~2.4 Ma based on K-Ar dating (Bacon and Duffield 1981). The LCD crest (elev. 2880) has a diameter of 700 m

and lies approximately 100 m above the base (elev. 2780 m), which has a diameter of about 900 m (Bacon and Duffield 1981). The rhyolite deposits described in this study are exposed in outcrop at the northwestern edge of the dome along a small stream channel approximately 55 m above the base. The subunits are distinguished on the basis of textural differences (e.g. vesicularity, pyroclast size, framework, etc.) and/or variations in crystallinity (e.g. crystal size, mineral abundance, color, etc.). K-Ar dating of sanidine crystals provide an eruption age of ~185 ka (Bacon and Duffield 1981), which coincides with a relatively high Pleistocene glacial period where sea level was about 80-100 m below modern sea level. Pre-eruptive temperatures of the LCD rhyolites are 676-723°C and 664-789°C based on magnetite-ilmenite and Ti-in-Zircon geothermometry calculations, respectively (Lewis et al. 2010). LCD is encompassed by a tuff ring containing ash identical to ash found in a tuff layer greater than 12 cm at Lake Manix near Barstow, CA. If the tuff is indeed derived from the LCD event then a violent eruption allowed ash to escape from glacial confines and be transported in suspension as far as 300 km away (Wells, 2009).

Analytical methods

The exposed outcrop at the northwestern flank of LCD was described in detail in the field and stratigraphic unit boundaries were determined on the basis of textural variations between adjacent layers. Multiple samples were also collected from each of the units for laboratory analysis. A stratigraphic column of the exposed units was constructed (Figure 2) and a scanning electron microscope (SEM) was used to image samples removed from respective units for microscopic examination (Figure 3). In addition, SEM and thin section images of the LCD outcrop samples were compared with samples taken from the outer tuff ring as well (Figure 4).

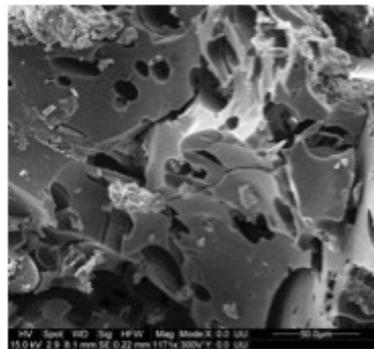


Figure 3 – Example scanning electron micrograph of a pumice sample from the base of the rhyolite tephra (Unit I). Vesicle voids are tubular and appear elliptical in cross section along the long axis and circular in cross section along the short axis. Note aggregates of fines attached to smoothed glassy walls of pumice.

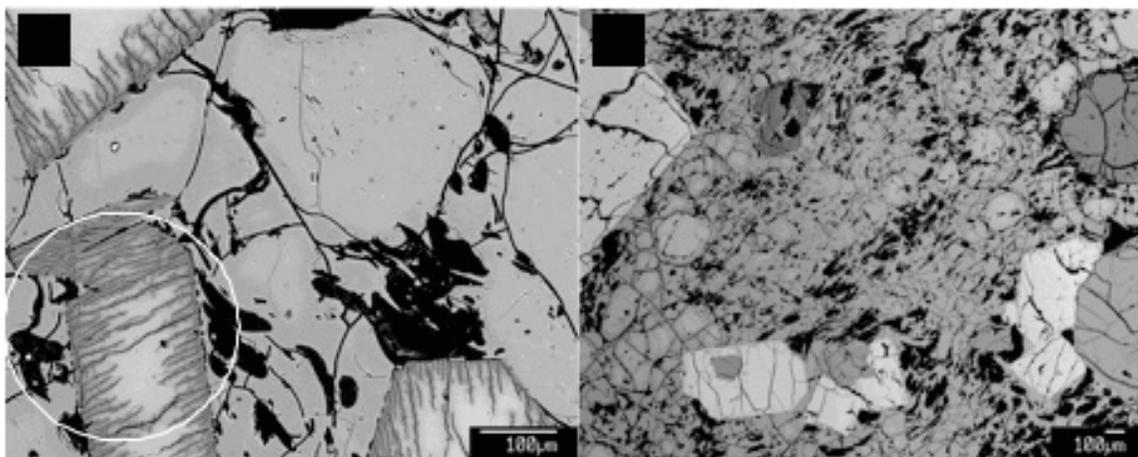


Figure 4 – Thin section images of A) a pumice fragment from the rhyolitic outcrop deposits and B) a pumice fragment from the surrounding tuff ring. Note that images are not shown at the same scale, however the tuff ring pumice is highly vesicular compared with the sample taken from the described outcrop. Sanidine crystals in photo A (circled) display rare dendritic cleavage fractures that are interpreted as a result of violent expansion of steam during contact of lava with glacial margins.

Macroscopic observations

The basal unit (Unit I) is a breccia apparently >2 meters thick as it is exposed at the surface along the incised drainage with no discontinuity for at least 2 meters (Figure 5). The weathered surface is gray, and a fresh surface is light gray and white. The unit is matrix-supported. Pumice resistance in this unit is fairly high, although the ash matrix weathers easily and is brittle within a few centimeters of surficial exposure.

Unit II is a lens of light gray angular, polygonal pumice clasts that has a poorly consolidated clast-supported framework (Figure 6). There is a horizontal grading from left to right with smallest pumice clasts (<1 cm) grading into largest clasts (>13 cm). Maximum unit thickness is about 55cm and pumice clasts contain microcrystalline minerals of sanidine and plagioclase.

Unit III is a layer that is approximately 60 cm thick and is a breccia composed of light gray and white subangular and subrounded pumice clasts supported by an ash matrix. Average pumice clast size is ~4 cm with largest clasts reaching ~7 cm. Pumices have well defined edges (Figure 7) and are easily distinguishable from ash matrix, which is composed of small pumice fragments <2 mm (Figure 8).

Unit IV is the fourth distinguishable layer and is light gray with white and pinches and swells from thin ~5 cm to medium ~15 cm thickness. It is a recessive unit and has very low resistance to weathering. Largest pumice



Figure 5 – Side-view Photograph of Unit I in outcrop facing down the stream channel toward the west. Field notebook and rock hammer handle for scale.



Figure 6 – Side-view photograph of Unit II (outlined in white) facing up the stream channel toward the east. Horizontal grading decreases toward the top of the photo. Rock hammer for scale.



Figure 7 – Photograph of a large pumice sample from Unit III collected for further examination. Note well-defined edges in contact with ash matrix. Pencil for scale.

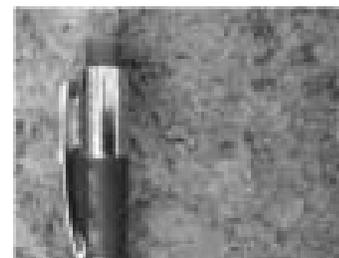


Figure 8 – Close up photograph of ash matrix from Unit III. Ash fragments are <2 mm. Pencil for scale.

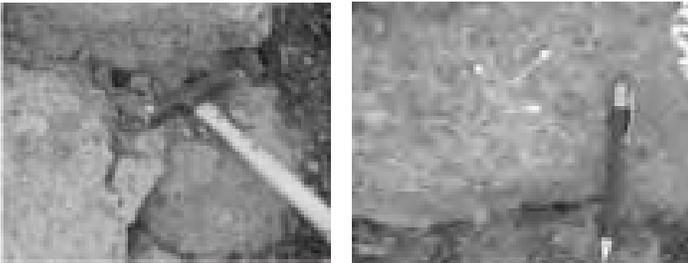


Figure 9 (top left)– Photograph of recessive Unit IV (outlined in white). Pumice clasts sizes range from ~3-6 cm. Rock hammer for scale.

Figure 10 (top right)– Photograph of Unit VI showing abundant pumice clasts and partially devitrified black perlitised pumice clasts ~1-3 cm (examples outlined in white). Unit V looks similar but with larger, less abundant clasts. Pencil for scale.

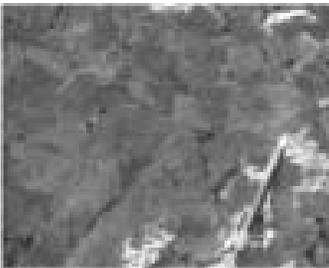


Figure 11 (left)– Photograph of Unit VII breccia showing black “salt and pepper” pumice clasts (some welded together) embedded with white pumice clasts that are ~6 cm. Pencil for scale.

clasts range from ~3-6 cm and are subangular and subrounded. Framework is an ash matrix composed of poorly consolidated pumice fragments with average grain size <5 mm (Figure 9).

Units V and VI are identical lithologically but and have been divided by the apparent differences in clasts sizes. Unit V is approximately 1.5 m thick and is a breccia containing light gray angular to subangular pumice clasts and partially devitrified black perlitised pumice clasts that typically range from ~4-6 cm. This unit is dominantly an ash matrix and large clasts are evenly dispersed. Unit VI is about 0.5 m thick and appears to be compositionally identical to Unit V, however Unit VI grades into smaller clasts that range from ~1-3 cm and becomes more abundant in pumices towards the top of the unit (Figure 10).

Unit VII is the top unit. It is a breccia exposed for at least 3.5 m thick before being concealed toward the dome crest by vegetation detritus. It has a gray to dark gray color and consists of large (~6 cm) gray angular pumices and partially devitrified black pumices in which some are welded together with no boundary between conjoined clasts. The black glassy pumices have a “salt and pepper” appearance with platy perlitised obsidian flakes combined with white quartz or plagioclase crystals ~1 mm (Figure 11). Pumices are supported by a fine gray ash matrix <2 mm.

Microscopic observations

Thin section images were taken from a sample collected from the LCD outcrop and a sample collected from the surrounding tuff ring. Figure 4A shows smooth, vesicle-poor, microlite-abundant glass. The vesicles appear to be sub-parallel and are restricted to accumulating along fracture planes and weaknesses within the glass. The sanidine crystals display unusual sub-parallel microfractures along cleavage planes (examples circled in Figure 4A). The sample in Figure 4B is from the tuff ring and shows vesicle-rich, microlite-free glass. The vesicles appear to occur independently of fractures in the glass and do not appear to be sub-parallel. The vesicles have a “wispy” sense in which localized groupings are stretched and distorted in similar orientations. Interestingly, sanidine crystals in lavas and pumice samples found elsewhere do not display the microfractures seen in Figure 4A.

Samples collected from the basal layer of the LCD outcrop (Unit I) were imaged using a scanning electron microscope. Figure 12A shows a bimodal distribution of grain sizes with blocky, smooth, vesicle-poor glass comprising the bulk of the sample and submicron-sized platy glass particles that accumulate on the face of the larger grains, especially within vesicle hollows. There are curved fractures that arch with no apparent structural purpose and terminate within the glass (examples circled in Figure 12A).

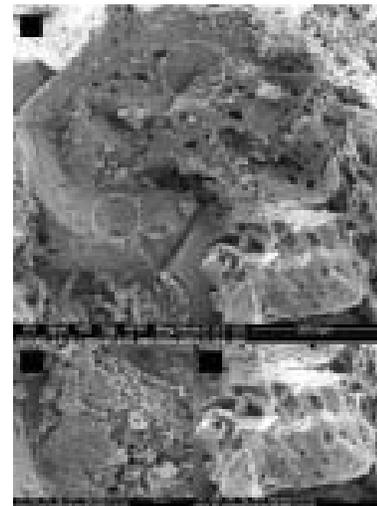


Figure 12 – SEM images of a pumice sample collected from Unit I. A) Arched fractures terminating within glass are circled. B) Closer image of features seen in middle left of A. C) closer image of features seen in lower right of A. See text for explanation.

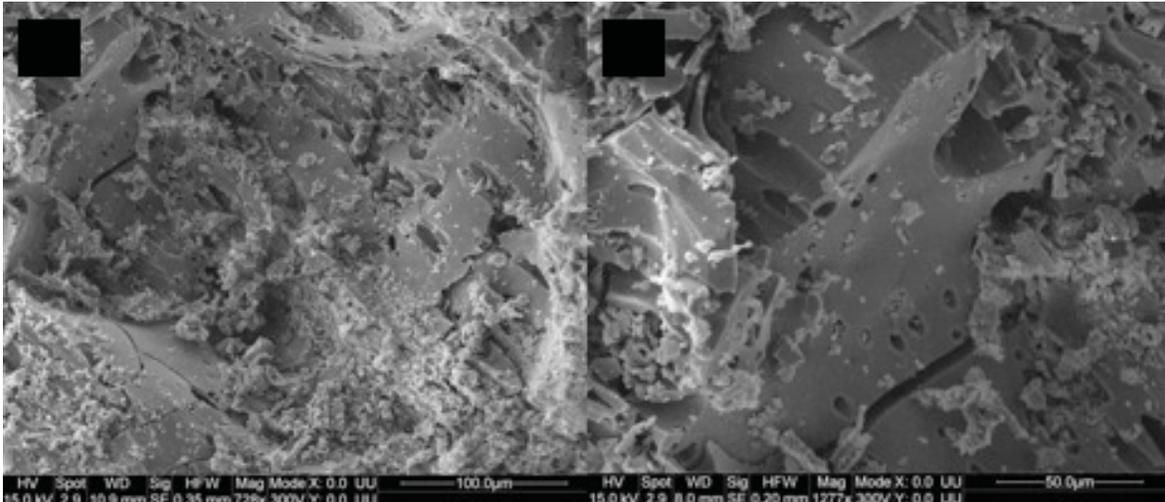


Figure 12 – SEM images of a pumice sample collected from Unit I. A) Arched fractures terminating within glass are circled. B) Closer image of features seen in middle left of A. C) closer image of features seen in lower right of A. See text for explanation.

Vesicles appear to be spatial and do not typically form in groups. Figure 12B displays parallel joints that form a stack of tabular glass. Figure 12C displays sub-parallel curvilinear extensional fractures near the top of the image and aggregates of nano-sized glass particles accumulate in vesicles (circled). Figure 13A is another image from the same sample and shows aggregates of fines accumulating on the surface of the smooth pumice faces. The nano- to micron-sized platy glass shards form a cauliflower-like texture along pumice faces when condensed in groups, which do not appear to be disrupted where adhered over fractured surfaces. The vesicles all appear to be tubular and elongate in the same direction, as shown more clearly in Figure 13B. Smooth and sinusoidal fracturing caused by degassing gives the face a jig-saw puzzle piece appearance.

Discussion

The following five lines of evidence suggest that LCD formed as a result of a subglacial volcanic eruption based on similarities observed here with those described from subglacial rhyolitic deposits elsewhere:

1. Most of the units in the LCD outcrop sequence are breccias and display a number of features that indicate interaction of the ejecta with water. First, the “jigsaw fit” of the pumices in Units V-VII suggests the

flow made contact with a wet substrate and clast rotation allowed the fragments to fit into place (McGarvie 2009). Also, the consolidated muddy matrix that supports these clasts is indicative of subglacial environments as caused by violent fractionation of the flow in direct contact with glacial margins and within a confined environment that resulted in poorly-sorted deposits, whereas a subaerial eruption would typically allow the majority of the fines to be removed in proximity to the dome (Tuffen et al. 2001). Lastly, the perlitized obsidian pumice clasts observed in Units V-VII are interpreted to have been formed by eruption in a hydrous environment as excess water could not be removed from the hot flow as vapor (Tuffen et al. 2001).

2. The lens that forms Unit II is unique to rhyolite deposits and likely formed from flow through a constrained channel. The poorly consolidated framework of this unit may be due to saturation of the flow by water that filled the channel and restricted welding of the clasts as water filled the voids and removed most of the fine material that would otherwise occupy the voids between the polygonal clasts.

3. In all likelihood there had to be a unique restraining element during eruption since volatiles were not free to escape in any direction as is usual for rhyolites in subaerial environments. The vesicles viewed in

SEM images are spatial, well defined, and typically elongate parallel or sub-parallel and it appears that volatiles were constrained possibly due to strengthening of the glass as it cooled. Vesicles orientations are likely perpendicular to flow direction and parallel to glacial contact margins.

4. The occurrence of extremely fine-grained ash shards adhered to pumice walls and hollows. This is considered to be indicative of phreatomagmatic eruptions (Heiken and Wohletz 1985; Buttner et al. 1999). These fines are suspected to have occurred from the violent interaction of rhyolitic lava with melt water (Stevenson et al. 2006). The bimodal distribution of grain sizes in these deposits suggests that only some of the flow came into contact with water and was phreatomagmatically fragmented (Stevenson et al. 2006).

5. Since the above features of the LCD rhyolite deposits indicate phreatomagmatic activity, it is possible for this eruption to have occurred beneath a non-glacial lake. However, the age of this eruption coincides with a high glaciated period and low sea level stands and it is more probable that the paleoclimate of this alpine region was glacial. Although the Sierras have been uplifting for millions of years, there is no evidence to indicate that Long Canyon had a paleotopography 185 ka that could sustain a non-glacial body of water and thus these features could only be explained by a subglacial eruption environment (Tuffen et al. 2001).

Conclusions

The deposits at LCD contain features that are consistent with features described elsewhere in subglacial rhyolitic deposits, particularly in Iceland where lava-ice interactions are more common. Considering the age of the LCD eruption and a coincidentally low sea level stand, it is probable that alpine glaciers were prevalent in this region during the Pleistocene. Although there is no definitive evidence that ice was the medium in which the rhyolite came into contact, there is evidence that a considerable amount of surface water was present during eruption periods. However, based on the possible paleotopography of the region during the Pleistocene it is not likely that a non-glacial lake could be present in this region, especially during a high glacial period; therefore, the medium was probably glacial ice and/or melt water. The most compelling feature of the LCD rhyolites is the bimodal grain size of the deposits with the platy micro-particles of glass

adhering to the larger bulk material. This feature has been documented in phreatomagmatic and glaciovolcanic deposits and is widely considered to be a diagnostic component for these types of eruption environments. Also, the unique microfractures in the sanidine crystals observed in thin section analysis are not well documented. The uniqueness of this particular feature suggests that these rhyolite deposits were at least emplaced under abnormal conditions, whether glacial or not. More analysis of LCD could provide additional information to support a glaciovolcanic origin including mapping the surrounding flow deposits and documenting any abnormalities that could not be explained by present topographic constrictions. Also, the perlitised obsidian flakes described in the upper layers of the sequence should be chemically analyzed for water content to determine the extent of excess water interaction with the flow. Until additional studies reveal supportive or contradictory evidence for a subglacial eruption at LCD it is apparent that these deposits were emplaced beneath glacial confines prior to break through of the glacier cauldron to release ash into the upper atmosphere. This study was intended to provide substantial evidence to establish glaciovolcanism at LCD and further detailed studies to reconstructing the eruption history would be beneficial to developing an order of events and the extent of lava-ice interactions.

Acknowledgements

I am grateful to Dr. Browne for offering this project to me. Working in the Sierra Nevada was a challenging and very rewarding experience, and I am grateful to have learned how to analyze samples for morphological data via scanning electron microscopy at both CSUF and CSUN for this project.

References

- Bacon CR, Duffield WA (1981) Late Cenozoic rhyolites from the Kern Plateau, southern Sierra Nevada, California. *American Journal of Science* 281: 1-34
- Buttner R, Dellino P, Zimanowski B (1999) Identifying magma-water interaction from the surface features of ash particles. *Nature* 401(6754): 688-690
- Furnes H, Fridleifsson IB, Atkins FB (1980) Subglacial volcanic – on the formation of acid hyaloclastites. *Journal of Volcanology and Geothermal Research* 8: 95-110
- Gudmundsson MT, Sigmundsson F, Björnsson H (1997) Ice-volcano interaction of the 1996 Gjalp subglacial eruption, Vatnajökull, Iceland. *Nature* 389: 954-957
- Heiken G, Wohletz K (1985) *Volcanic ash*. University of California Press, Berkeley
- Lescinsky DT, Fink JH (2000) Lava and ice interaction at stratovolcanoes: use of characteristic features to determine past glacial extents and future volcanic hazards. *Journal of Geophysical Research* 105: B10 23,711-23,726
- Lewis B, Browne B, Van Sickle D, Vazquez J (2010) Geological and petrological observations of the Long Canyon Dome rhyolite, Sierra Nevada. *Geological Society of America Cordilleran Section (Abstract 18-6 BTH 6)*
- Lodge RWD, Lescinsky, DT (2009) Fracture patterns at lava-ice contacts on Kokostick Butte, OR, and Mazama Ridge, Mount Rainier, WA: implications for flow emplacement and cooling histories. *Journal of Volcanology and Geothermal Research* 185: 298-310
- McGarvie D (2009) Rhyolitic volcano-ice interactions in Iceland. *Journal of Volcanology and Geothermal Research* 185: 367-389
- Stevenson JA, McGarvie DW, Smellie JL, Gilbert JS (2006) Subglacial and ice-contact volcanism at the Öræfajökull stratovolcano, Iceland. *Bull Volcanol* 68: 737-752
- Tuffen H, Gilbert J, McGarvie D (2001) Products of an effusive subglacial rhyolite eruption: Bláhnúkur, Torfajökull, Iceland. *Bull Volcanol* 63: 179-190
- Wells, Chrysta (2009) Major Element Fingerprinting of Lake Manix Tephra – Origins in the Southern Sierra Nevada? California State University, Fullerton Department of Geological Sciences Thesis

Measuring debris-flow timing relative to rainfall in burned watersheds: Implications for residential evacuations and early warning

Department of Geological Sciences, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

Robert J. Leeper
Advisor: Dr. Brady Rhodes

Abstract

Data on the timing of post-fire debris flows are rare and mostly limited to eyewitness accounts. I contributed to the development of a method to measure the timing of post-fire debris flows relative to rainfall that utilizes inexpensive pressure transducers that are commonly used to measure stream and groundwater levels. I measured the timing of 11 debris flows relative to rainfall intensity in seven watersheds burned during the 2009 Station Fire. The watersheds upstream from the pressure transducers ranged from 0.02 km² to 0.63 km². All of the recorded debris flows took place within 10-minutes or less of local peaks in short duration rainfall intensity, which demonstrates that post-fire debris flows respond quickly to intense rainfall.

I also reconstructed the debris flow events of February 6, 2010. On that date two early morning pulses of rain mobilized over 13,000 y³ of debris and filled the Mullally debris basin to capacity, while later rain pulses triggered debris flows within the same watershed that damaged or destroyed 41 homes downstream from the debris basin (Kim et al., 2010). Reconstruction of the relationship between rainfall intensity and the timing of debris flows emphasizes the importance of early evacuation by residents and will inform deliberations on how best to communicate post-fire debris-flow hazards to the public.

Introduction

Post-fire debris flows are likely to continue in the western United States and occur even more frequently in other regions of the world (Cannon

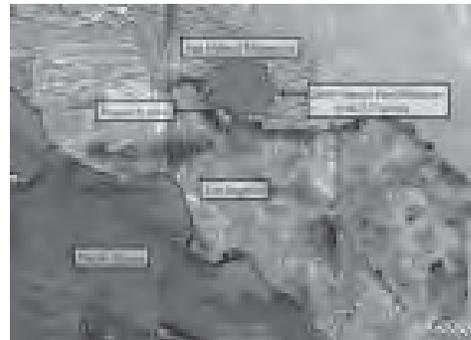


Figure 1 – Research area is located in the San Gabriel Mountains, 25km north of Los Angeles



Figure 2 – Oblique view of research area and 11 instrumented watersheds along the front of the San Gabriel Mountains

and DeGraff, 2008). As human development continues to encroach upon mountainous terrain and as the influence of climate change on the magnitude, frequency, and duration of wildfire becomes better understood (Westerling et al., 2006), residents that live in or near mountain drainages must be aware of potential geologic hazards that can put them and their families at risk.

Since 1933, the San Gabriel Mountains above the communities of La Crescenta, Montrose, and Glendale have experienced two wildfires. Winter storms hit the burn area soon after both fires were extinguished and caused debris flows. In late November 1933, 7,000 acres burned along the front of the San Gabriel Mountains. On January 1, 1934, intense rainfall above the burn area triggered debris flows that killed 42 people, destroyed 371 homes, and damaged hundreds more below the burn area (Chawner, 1934). On August 26, 2009, the Station Fire ignited and burned 160,577 acres in 51 days and resulted in the evacuation of thousands of residents who lived in the mountain and surrounding foothill communities (Fig. 1).

Methods

I installed Solinst Levellogger Model 3001 pressure transducers in 11 burned watersheds along the front of the San Gabriel Mountains (Fig. 2). Pressure transducers measure pressure head in meters of water and have a self-contained datalogger and power supply. The datalogger can be set to sample at different rates but memory capacities limit measurements to 40,000 pressure readings. Recording can occur as a continuous loop, with the oldest data overwritten, or as a single loop that shuts down when the memory chip is full. For this study, I chose a 1-minute sampling rate and a single loop. At this sampling rate the memory chip fills in 27.7 days. This method ensures that some data are captured and preserved, even when access to a site is delayed by days or weeks. The peak pressure reading recorded was used to identify the timing of flows.

Channel cross-sections were measured at pressure transducer sites using a Topcon® GPT-9000 Total Station (Fig. 3a). Marking cross-section control

points approximately 2 m up the wall of the channel ensures consistent survey lines before and after flow events. Drilling a hole (2.5 cm diameter by 20 cm deep) into strong bedrock along the axis of the channel and inserting the pressure transducers to a depth of 1 cm below the bedrock surface provides a relatively secure installation site (Figs. 3b and 3c). At four installation sites, pressure transducers installed in weaker rock at the downstream end of an outcrop were lost during debris flows that caused bedrock erosion. A tipping-bucket rain gauge installed as close as possible to the location of the pressure transducer provides rainfall data that are correlated with debris-flow timing (Fig. 3d).

Following a storm and after locating and recovering the pressure transducers, the data are downloaded to a field laptop (Figs. 3e and 3f). Surveying the high-flow mark and examining sedimentary deposits determine whether a debris flow or flooding occurred (Figs. 3g and 3h). It is necessary to perform these last two steps because the pressure transducer may miss the peak of a fast-moving flow because the pressure transducer cannot sample fast enough, and the transducer is unable to distinguish a debris flow from flood because the density of the flow is unknown. Nevertheless, we are using the pressure signal to identify timing, not to identify the type of event. In the field, the presence of matrix supported debris flow levees and mud-veneers classified an event as a debris flow.

Results

Two storms, each with a 2-year recurrence probability, affected the research area on January 17-23, 2010 and February 6, 2010. The data recorded during the storms show that multiple basins had similar responses. Pressure



Figure 3a –
Survey cross section

Figure 3b –
Drill a hole

Figure 3c –
Insert pressure
transducer

Figure 3d –
Install rain gauge

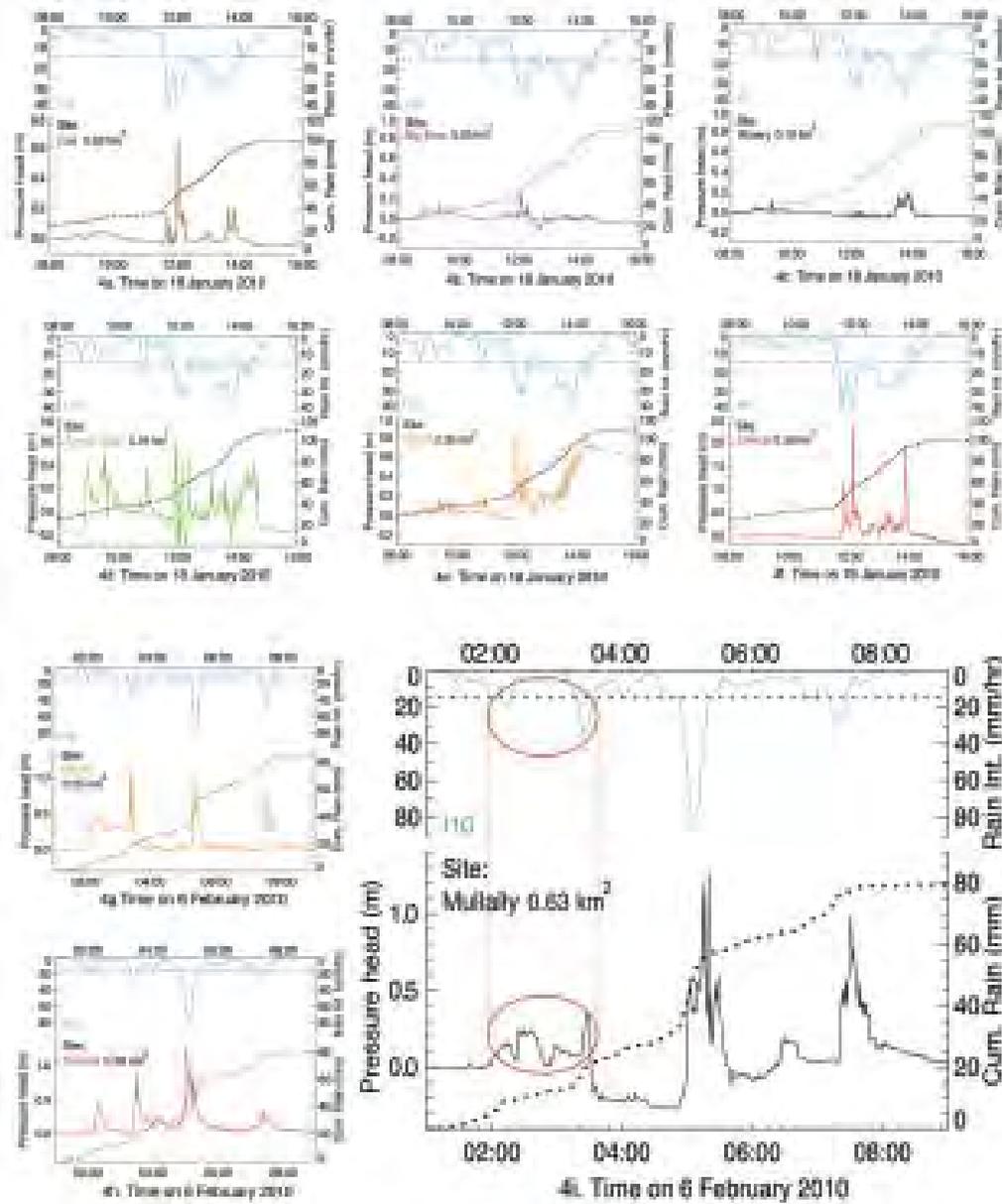
Figure 3e –
Record an event

Figure 3f –
Locate pressure
transducer

Figure 3g –
Check sedimentation

Figure 3h –
Survey high-flow marks

Figures 3a-3h – Images show pressure transducer installation, maintenance, and sedimentation identified during post-event surveys. Photo credit: Robert Leeper and Jason Kean



Figures 4a-4i – Plot of pressure data showing pressure head in meters of water across the lower x-axis panel and the 10-minute rain intensity measured near each basin across the top x-axis panel. Cumulative rainfall is shown with the dotted black line. The dashed blue line is the empirical debris-flow threshold (15mm/hr) of Cannon *et al.*, 2008.



Figure 5a –
185 meters from the Mullally debris basin



Figure 5b –
190 meters from the Mullally debris basin



Figure 5c –
220 meters from the Mullally debris basin



Figure 5d –
250 meters from Mullally debris basin

Figures 5a-5d – Images showing damage and marooned residents on the morning of Feb 6th, 2010 after debris flows breached the Mullally containment basin and inundated the neighborhood below it. Photo Credit: Robert Leeper

data recorded at each site indicate multiple flow periods, flow timing that corresponds very closely with peak 10-minute rainfall intensity (I10), and that peak rainfall intensity was at or above the 15-mm/hr, 10-minute duration threshold (T10) of Cannon et al. (2008) when debris flows occurred (Figs. 4a-4h).

The January 17-23, 2010 storm triggered debris flows in the Arroyo Seco, Big Briar, Gould, Oak Creek, Shields, and Winery watersheds. Bedrock erosion destroyed the pressure transducers installed in the Goss, Hay, and Mullally watersheds during the storm. However, post-storm sedimentation surveys at these three sites confirmed that debris flows did occur.

The Mullally debris basin was completely empty on January 14, 2010, and at the time had a maximum capacity of 13,000 y3. The January 17-23, 2010 storm filled the Mullally debris basin to capacity (Leeper and Kean, 2010).

On February 5, 2010, excavation crews had almost completely emptied the Mullally debris basin as another storm was forecast to hit southern California one day later.

Eyewitness accounts from Los Angeles Department of Public Works (LADPW) personnel monitoring debris levels in the basins below the burned

area provide additional information on the timing and flow characteristics during the February 6, 2010 storm. Between 2:00 am and 3:25 am, three peaks in rainfall were recorded near the Mullally watershed. The peaks exceeded T10 and had values of 16-mm/hr (2:00 am), 20-mm/hr (2:25 am), and 35-mm/hr (3:25 am) and pressure data show flow periods that coincide with T10 being exceeded (Fig 4i). At 4:25 am, a LADPW patrol crew arrived at the Mullally debris basin overlook and reported the basin was full and clear water was flowing over the basin wall and into the spillway. At 4:32 am, the crew attempted to leave the Mullally overlook. The crew's path was blocked by an active debris flow that breached the Mullally debris basin, blocked the spillway, and inundated the homes just downstream of the debris basin.

Combining the pressure transducer and rainfall data along with the information provided by the first-hand accounts of LADPW personnel, confirms the debris flows associated with the two early morning pulses of rain on Feb 6, 2010 mobilized approximately 13,000 y3 of material and filled the Mullally debris basin to capacity. The flows associated with the later pulses of rain that exceeded T10, flowed over the Mullally debris basin and inundated the community downstream, marooning residents, and causing major damage and destruction (Figs 5a-5d).

Conclusion

The relationship between rainfall intensity and the timing of post-fire debris flows should be a topic of discussion in residential areas that abut recently burned slopes. Our timing data show that post-fire debris-flows are closely tied to local peaks in I10. This means that post-fire debris flows begin in less than 10 minutes after intense rainfall. Effectively communicating the geologic hazard that post-fire debris flows present will help the public better understand the hazard and its risk. If members of the public choose not to follow emergency management evacuation orders and are in need of rescue, it is unlikely that emergency responders will be able to reach them during a storm. Effective warning and early evacuation needs to come well before a storm is over a burn area: There is very little lead-time for post-fire debris-flow warning once heavy rainfall starts.

References

Cannon, S.H., and DeGraff, J., 2008, The Increasing Wildfire and Post Fire Debris Flow Threat in Western USA, and Implications for Consequences of Climate Change: Springer-Verlag Berlin Heidelberg 2009, p. 14.

Cannon, S.H., Gartner, J.E., Wilson, R.C., Bowers, J.C., and Laber, J.L., 2008, Storm rainfall conditions for floods and debris flows from recently burned areas in southwestern Colorado and southern California: *Geomorphology*, v. 96, p. 250-269.

Chawner, W.D., 1934, The Montrose-La Crescenta (California) flood of January 1, 1934, and its sedimentary aspects. [Master's thesis].

Kim, V., Vives, R., Rong-Gong, L., and Gottfried, A.J., 2010, Mandatory evacuations ordered for 500 homes in La Canada Flintridge, La Crescenta and Acton, Volume February 6, 2010, L.A. Times. (<http://articles.latimes.com/2010/feb/06/news/la-mem-storm7-2009-feb07>)

Leeper, R., and Kean, J., 2010, Some debris basins were nearing capacity, USGS: MHDP (<http://urbanearth.gps.caltech.edu/2010/01/29/some-l-a-county-debris-basins-were-either-near-or-at-full-capacity/>)

Westerling, A.L., Hidalgo, H.G., Cayan, D.R., and Swetnam, T.W., 2006, Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity: *Science*, v. 313, p. 940-943

A Magma Mixing Origin for Andesites Erupted in 2006 from Augustine Volcano, Alaska

Department of Geological Sciences, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

Ashley Melendez

Advisor: Dr. Brandon Browne

Abstract

The 2006 eruption of Augustine Volcano in Alaska produced andesite magma containing quenched basaltic inclusions, suggesting a “magma mixing” origin for the petrogenesis of the andesite. To test the magma mixing origin hypothesis, amphibole crystals contained in quenched basaltic inclusions and the andesite were analyzed using an Electron Microprobe as a way of compositionally fingerprinting them. Results show the existence of (at least) two distinct populations of amphibole crystals; a high-Al and low Si population that nucleated and grew within the basaltic inclusion-forming magma and a low-Al and high-Si population that nucleated and grew within the andesite magma. The presence of two distinct populations of amphibole not only can be used to verify the magma mixing hypothesis for 2006 Augustine andesite formation, but also suggests a very short time period between mixing and eruption, as intermediate compositions of amphibole are absent.

Introduction

Prior studies have argued that Augustine Volcano erupts magma formed via complicated open-system magmatic processes, including magma mixing (Steiner, 2009; Larson et al., 2010). Because the process of magma mixing has long been considered an important mechanism for the triggering of volcanic eruptions, (e.g., Sparks et al., 1977), studies aimed at understanding volcanoes characterized by open system processes are important because results may aid in the prediction and mitigation of volcanic eruptions in general.

Magma mixing is defined as the hybridization of at least two compositionally unique magmas in the liquid state. A very similar process, known as magma mingling, is defined as the incomplete hybridization of compositionally unique magmas, which results in the presence of blobs or drops of incompletely mixed magma compositions. The unmixed blobs are often called “quenched basaltic inclusions” because they are usually basaltic in composition and exhibit textures indicative of rapid quenching.

An intriguing characteristic about magmas that form through the magma mingling process is that they themselves are actually very well mixed, with less than 1% of the material being preserved as quenched inclusions, suggesting that inclusion formation may be the exception and not the rule. After all, extensive degrees of hybridization between compositionally and thermally unique magmas with contrasting viscosities requires that pre-mingling viscosities of the compositional end-member magmas must converge in such a way where liquid-liquid mixing (as opposed to mingling) can occur. To achieve this, previous studies indicate that the high viscosity (SiO₂-rich andesite magma) end member is heated by a lower viscosity (SiO₂-poor basalt magma) end member as the basalt intrudes into the andesite from below (e.g., Browne et al., 2006). As a result, the basalt magma cools as the andesite magma warms, which causes their viscosities to converge so that liquid-liquid mixing can occur.

A consequence of this mixing process is that minerals inherent to each magma end member must be exchanged from one magma composition to the other. This paper describes results of a study that use major element concentrations of amphibole crystals to show magma mixing between a basaltic magma and an andesite magma prior to (and during?) the



Figure 1 – Location map of Augustine Volcano (from Alaska Volcano Observatory)

2006 eruption of Augustine Volcano in Alaska. Amphibole is a perfect mineral family to examine for this study because (1) Amphibole occurs in both the basaltic inclusions and the andesite lava, and (2) amphibole is a hydrous mineral phase that is sensitive to changes in the composition, temperature, and pressure of the melt from which it crystallizes. Comparing the compositions of amphibole in the basaltic inclusions with those in the andesite lava will allow me to determine if they have the same or different magma sources, which will allow me to say whether or not magma mixing between two different magmas formed the 2006 Augustine Volcano eruptive products.

Geologic Setting

Augustine Volcano is a small, unpopulated (~100km²) island located in the Cook Inlet of Alaska ~120km southwest of Homer, Alaska (Figure 1). It is situated along the eastern portion of the Aleutian Island Arc and is similar in eruption style to the slightly larger, but less frequently active, volcanoes there: Spurr, Redoubt, Iliamna. Augustine Volcano erupts as an andesite dome complex and has erupted 6 times in the past 130 years, with the last eruption in 2006. The oldest rocks on Augustine are not volcanic at all, but Jurassic sedimentary rocks of the Naknek formation. The oldest volcanics on the island are poorly sorted with clasts of olivine basalts of Pleistocene age (Waitt and Beget, 2009).

The 2006 eruption was marked by three main eruptive phases following a precursory phase: explosive, continuous and effusive. The precursory phase (May 2005 to January 11, 2006) was marked by an increase of daily number of earthquakes on the island as well as an inflation of the volcano (Coombs et al., 2010). The explosive phase began on January 11, 2006 and continued with 13 explosive events until January 28, 2006. The beginning of this phase was marked by a strong swarm of earthquakes and a powerful explosive event reaching heights of 9km above sea level on January 12. The explosive eruptions continued for days reaching a maximum height of 14km. During this time, pyroclastic and volcanoclastic materials were deposited, notably the largest single deposit of the eruption, the Rocky Point pyroclastic flow (Figure 2). Inflation of the volcano continued during this time as evidenced by drumbeat earthquakes. These earthquakes are often associated with flux of new lava (Coombs et al., 2010). The continuous phase (January 28-February 10) immediately followed the Rocky point pyroclastic flow. The eruption style shifted abruptly and the volcano began to deflate. The eruption emitted a nearly continuous plume at this point of the eruption, rich in ash and producing many pyroclastic flows. During this time, ash fall was carried as far as Homer and Seldovia, some 115km away (Coombs et al., 2010). The continuous phase is named so because the volcano continuously built and destroyed domes during this time and emplaced block-and-ash flow deposits

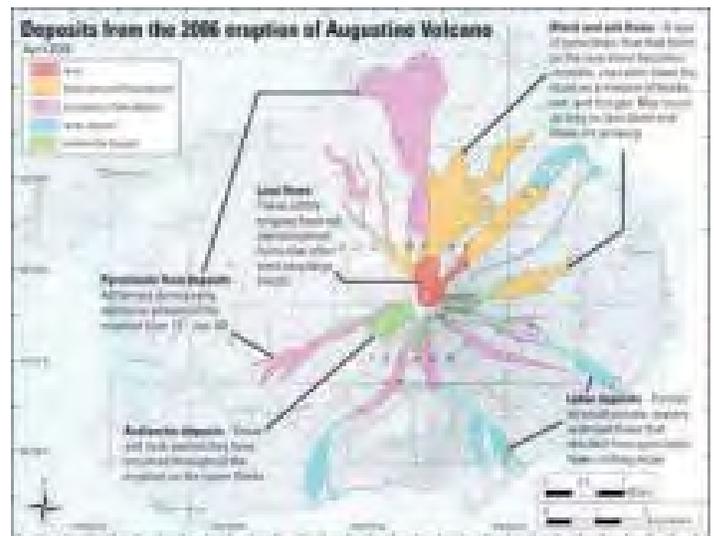


Figure 2 – Geologic map showing the 2006 eruption deposits (from Alaska Volcano Observatory)

(Figure 2) (Coombs et al., 2010). The final phase of the eruption, the effusive phase, (March 3–March 16) occurred after a 3-week hiatus in the eruption. This phase produced two lava flows and many gravity-driven block-and-ash flows and rock falls (Figure 2) (Coombs et al., 2010).

Methodology

We collected samples of both the basaltic inclusions and andesite from Augustine Volcano in the summer of 2010. We looked for inclusions in the rock that were dark gray to black, vesicular and that contained prominent plagioclase phenocrysts in a fine-grained groundmass. We also looked for rounded, cusped boundaries. With a chisel and rock hammer, we removed the inclusions from the surrounding rock along with a piece of the immediately surrounding the inclusion. From these rock samples, we made petrographic thin sections in order to analyze the mineralogy, composition and textures of both the “host” andesite lava and the inclusions. A portion of the thin sections was made professionally and the rest of the thin sections were made using the equipment at CSUF. Four thin section samples containing the most abundant amphibole were compositionally and texturally analyzed using electron microprobe (EMP) analysis with backscatter electron (BSE) imaging. EMP analysis operates by directing an accelerated stream of electrons at a carbon-coated thin section sample in a vacuum chamber. The electron stream

bombards the sample and ejects electrons from an inner electron orbit. Outer electrons fall back to fill vacant spaces, which result in the generation of X-rays. Spectrometers count and measure the wavelength of the X-rays released by the sample, which are related to specific elements. Because EMP analysis is non-destructive and can be done on a polished thin section, very small areas (1 micron) can be analyzed in situ without damaging the sample. For this study, a 15 KeV accelerating voltage was used for the electron beam on my samples.

Results

The EMP compositional data from the four andesite thin sections and the basaltic inclusions were divided into data taken from the core of an amphibole crystal and data taken from the rim of a crystal. The samples were plotted for amphibole composition with a resulting majority of a range from hornblende to pargasite in composition with the most of the samples falling into the hornblende and edenite composition range (figure 3). Notice that there are two main clusters with regards to the composition of the amphibole samples analyzed. The two clusters include both core and rim points. One cluster contains relatively higher amounts of Na and K, but lower amounts of Si. This grouping falls into the edenite range of compositions on the cusp of the pargasite region. The second cluster contains relatively less Na and K, but

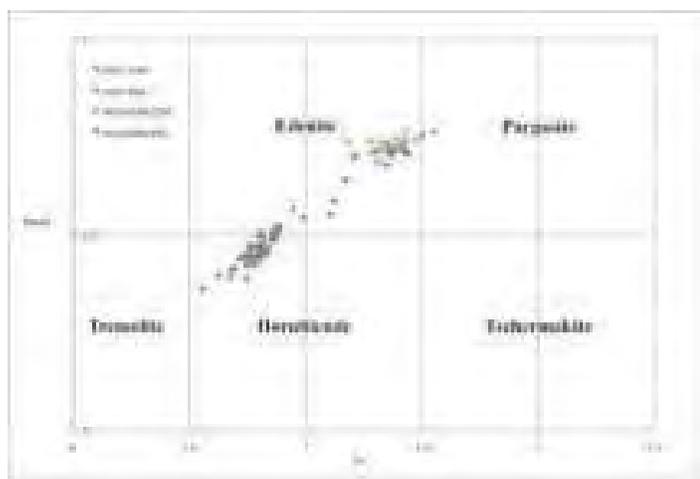


Figure 3 – Classification diagram (adapted from Deer et al., 1992) for amphibole crystals. Amphibole crystals from basaltic inclusions range from edenite to pargasite, whereas andesite “host” lava amphiboles are predominantly hornblende with some edenite.

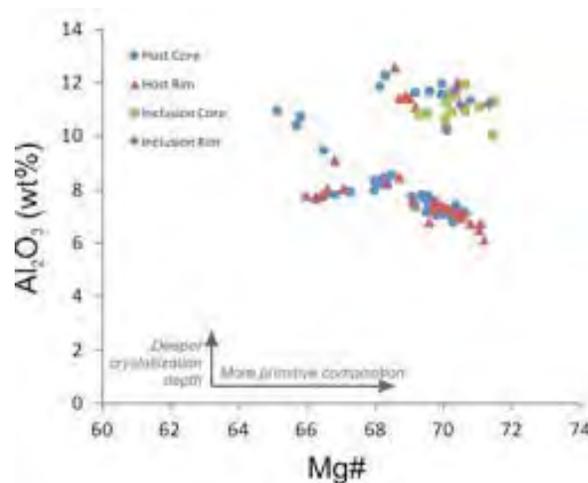


Figure 4 – Concentrations of Al_2O_3 plotted against $Mg\#$ ($Mg/(Mg+Fe)$) of amphibole crystals, which shows that basaltic inclusion amphiboles are slightly more primitive and formed at greater depths compared to amphiboles from the “host” andesite lava.

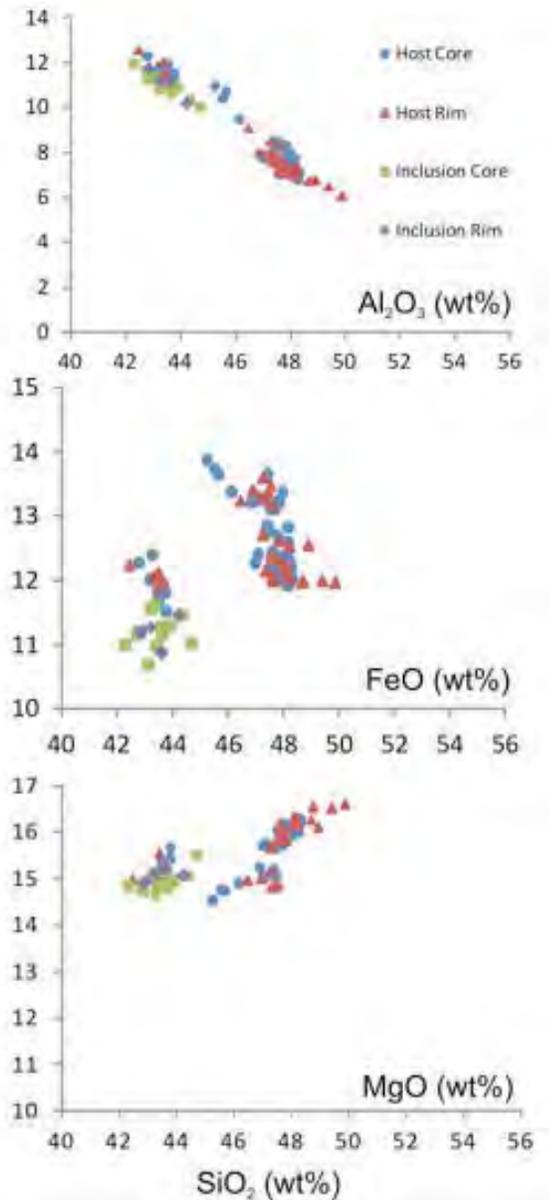


Figure 5 – Major element concentrations of Al_2O_3 , FeO , and MgO plotted against SiO_2 by weight percent for rim and core portions of amphibole crystals from the basaltic inclusions and “host” andesite lava. Note that inclusion amphibole generally plots away from most andesite amphibole, although some andesite amphibole exists in inclusions, suggesting that hybridization of the two magmas occurred prior to inclusion formation.

greater Si. This cluster represents the largest number of samples from the andesite and falls into the hornblende/edenite boundary. The two clusters are linked together in a nearly linear pattern. Of particular interest here is that all the points taken from the basaltic inclusions fall into the lower Si, higher Na+K cluster, while points from the andesite are distributed in both populations. Comparing the concentration of Al_2O_3 in amphiboles to their magnesium number ($\text{Mg\#} = \text{Mg}/(\text{Mg}+\text{Fe})$) is useful because the aluminum concentration in amphiboles is controlled by pressure (i.e., crystallization depth) and the Mg# is a record of how primitive (i.e., Mg-rich) the melt from which the amphiboles crystallized. In this plot (figure 4), a range of compositions is expressed with clear clusters of concentrated compositions. Amphibole crystals from basaltic inclusions form a cluster composed of higher Al_2O_3 and relatively higher Mg# compared to the amphibole from the “host” andesite lava. There does not appear to be a compositional linkage between the two clusters, but the two clusters exhibit almost parallel linear relationships. Again, it is interesting to note that all amphibole from the basaltic inclusion are only located in the higher Al_2O_3 cluster, while the points from the andesite are found in both populations.

Amphibole crystals can also be compared in terms of Al_2O_3 , FeO , and MgO concentration versus SiO_2 concentration (figure 5). Here, two distinct clusters of compositions that are linked together by a few points of andesite samples that form an intermediate composition between the two clusters. The basaltic inclusion amphibole is higher in Al_2O_3 , lower in FeO , equivalent in MgO , and lower in SiO_2 compared to amphiboles that nucleated and grew in the andesite magma.

Discussion

Aluminum concentration in amphiboles increases with depth of nucleation and growth (Deer et al., 1992). The compositions as plotted in figure 3 exhibit at least two distinct concentrations of compositions that represent two end-member amphibole compositions. The plot in the edenite to pargasite range represents a much higher stability temperature than the other plot in the high hornblende range (Deer et al., 1992). The compositional data has all indicated strong evidence for at least two distinct amphibole populations within the andesite. Through all of the compositional plots, there have been two distinct clusters of core and rim compositions that trend nearly parallel to each other. In the case of Augustine volcano, the source of these high-Al, high-temperature amphiboles would be the intruding basalt, one compositional end-member. The lower Al content and temperature composition would

represent a shallower source. The majority of the material erupted on Augustine Volcano is the andesite and represents the mixing between these two compositional end-members.

The data from the basaltic inclusions all plot within the same cluster for each individual graph and are not distributed into the two clusters like the andesite data. The amphiboles from the andesite, however, are found to exist in the same compositional cluster as some of the basaltic inclusion samples. This requires that hybridization of the andesite and basalt magma – and therefore crystal exchange between the two end-members – preceded the formation of inclusions. Moreover, because intermediate compositions of amphibole crystals exist between the inclusion-type and the andesite-type, even in terms of rim compositions, the time span for hybridization, inclusion-formation, and eruption must have been very short, possibly even days.

Acknowledgements

Thank you to Dr. Browne for including me in such an exciting project that included traveling to an active volcanic island in Alaska. This research was supported by the US Geological Survey and Alaska Volcano Observatory through a grant awarded through the American Recovery and Reinvestment Act (Grant # G10AC00033). I would also like to thank the adorable foxes on Augustine Island for keeping me happy during field work despite the wind and rain!

References

Browne, B.L., 2009, Magma mixing in 'real time': Evaluating the role of basalt in eruption triggering and the generation of lithological diversity in deposits from the 2006 eruption of Augustine Volcano, Alaska. USGS Proposal

Browne, B.L., Gardner, J.E., 2006, The influence of magma ascent path on the texture, mineralogy, and formation of hornblende reaction rims. *Earth and Planetary Science Letters*, v. 246, p. 161-176.

Coombs, M., Bull, K., Vallance, J., Schneider, D., Thoms, E., Wessels, R., and McGimsey, R., 2010, Timing, distribution, and volume of proximal products of the 2006 eruption of Augustine Volcano, in Power, J.A., Coombs, M.L., and Freymueller, J.T., eds., *The 2006 Eruption of Augustine Volcano, Alaska*: U.S. Geological Survey Professional Paper 1769.

Deer, W.A., R.A. Howie and J. Zussman, 1992, *An introduction to the rock forming minerals*. Pearson Education Limited, Essex, England.

Sparks, R.S.J., H. Sigurdsson, and L. Wilson, 1977. Magma mixing: a mechanism for triggering acid explosive eruptions. *Nature* v. 267 p. 315–318.

Steiner, A.R., Browne, B., and Nye, C., In Press, Quenched basaltic inclusions in <2,200 yrs B.P. deposits at Augustine Volcano, Alaska. *International Geology Review*.

Waitt, R.B., Begét, J.E., 2009, *Volcanic processes and geology of Augustine Volcano, Alaska*: U.S. Geological Survey, Professional Paper 1762, 78 p., 2 plates, scale 1:25,000.

Fault Scarp Morphology Along The Northern Eureka Valley Fault Zone, Eureka Valley, Eastern California, U.S.A.

Department of Geological Sciences, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

Ernest Nunez
Advisor: Dr. Jeffery Knott
LSAMP Grant #HRD-0802628

Abstract

Eureka Valley, CA, in northwestern Death Valley National Park, is one of a series of valleys formed by oblique extension within the Eastern California shear zone and Walker Lane belt. In Eureka Valley, extension is accommodated by the normal-oblique, north-south trending Eureka Valley fault zone (EVFZ) that bounds the valley's east side at the foot of the Last Chance Range. A Mw 6.1 earthquake on May 17, 1993 is attributed to a northeast-southwest-trending trace of EVFZ located in the west-central portion of the valley. The earthquake produced ground deformation best detected by remote sensing with only minor ground cracks observed after the earthquake. The minor ground rupture is in stark contrast to the prominent fault scarps found along valley-bounding fault on the east side. Here we present measured fault scarp profiles offsetting Quaternary alluvial-fan deposits along the northern traces of the EVFZ at the foot of the Last Chance Range. In this study, I will measure fault scarp height and slope angle to determine the earthquake magnitude and earthquake age. I will compare these data with previous studies of fault scarps along the southern EVFZ. The parallel trends of the northern scarps and the 1993 event suggest that a cross-valley fault is developing in Eureka Valley. The development of a cross-valley fault is consistent with clay models and field observations of a basin in the late extension phase.

Profile of energy delivered to chicks by the Elegant Tern (*Thalasseus elegans*) over a recent 16-year period (1994-2009) at the Bolsa Chica Ecological Reserve

Fish, Seabird and Conservation Biology Lab
 Department of Biological Science, College of Natural Science and Mathematics
 California State University, Fullerton, CA, USA

Laura Tolentino, Tyler Flisik
 Advisor: Dr. Mike Horn

Abstract

Seabirds have high metabolic rates and feed primarily on energy-rich fish to sustain their high energy demands. Over a recent 16-year period (1994-2009), pipefish (*Syngnathus* spp.) have increased in the diet of the chicks of the Elegant Tern (*Thalasseus elegans*) from about 2% to 66% at the Bolsa Chica Ecological Reserve (BCER). We predicted that pipefish have a lower energy density (kJ g⁻¹) than other major prey species of this tern, making pipefish an insufficient energy source for proper growth and development of the tern's chicks. After assessing the energy density and lipid content of the top six prey species (~75% of total diet) of Elegant Tern chicks over 10 of the last 16 years at BCER, we found a positive correlation between (1) standard fish length (mm SL) and dry mass (g), (2) standard fish length and fish energy content, and (3) an inverse correlation between yearly energy content delivered to chicks and pipefish proportions. Our results lead us to conclude (1) that pipefish do not contain a lower energy, protein or lipid content, but a higher ash content than most other major prey (2) the increase in pipefish did not affect the energy delivered to the Elegant Tern colony at BCER, and (3) that the increase of pipefish did not negatively affect Elegant Tern nest numbers. The energy content and prey composition data for all six major prey species provided a profile of the energy delivered to the Elegant Tern chicks each year. Such a profile has value for predicting energy content for chick diets and perhaps Elegant Tern breeding success in future years.

Introduction

Birds have some of the highest metabolic rates among vertebrates, with seabirds averaging rates 60% higher than those of terrestrial bird species

(Nagy 1999). To compensate for their high metabolic requirements, most seabirds feed heavily on high energy-fish species, such as northern anchovy (*Engraulis mordax*) and Pacific sardine (*Sardinops sagax*). The Elegant Tern (*Thalasseus elegans*) is the most abundant coastal seabird that breeds in southern California with nest numbers ranging from 1,200 to 20,000 nests per year for 1993-2010 (Patton et al. 2010). During the nesting season, Elegant Tern parents deliver a single fish to their young about 4 times a day (Cole 1997), making the quality (energy content) of each fish delivered critically important. Shifts in diet away from energy-rich prey species can slow the growth and reduce the lipid stores of seabird chicks at fledging (Wanless et al. 2005, Romano et al. 2006), including those of the Elegant Tern (Dahul and Horn 2003). In a laboratory feeding experiment with Elegant Terns, chicks that were fed a high-energy diet had a higher body mass and lipid reserves at fledging than those fed a low-energy diet (Dahdul and Horn 2003). Substantial lipid reserves are important because they provide energy to the chicks during periods of food deprivation (Taylor and Konarzewski 1989) and when chicks are learning to fly and forage for themselves (Reid et al. 2000).

The diet of Elegant Tern chicks nesting at the colony at the Bolsa Chica Ecological Reserve (BCER) in southern California has been recorded for most of the past 16 years (1994-2009) by Mike Horn and his students. During this period, the proportion of pipefish (*Syngnathus* spp.), a reputedly low-quality prey, has increased in the bird's diet while the proportions of anchovy, sardine and other fish species have declined. In 1996, pipefish represented only 2.2% of the chick diet (Horn unpublished data), whereas in 2008 the proportion had increased to 66% (Hendricks 2011). The increase in pipefish raises concern for the future of the Elegant Tern colony at BCER because such an increase may result in lower reproductive output (i.e., lower numbers of nests or

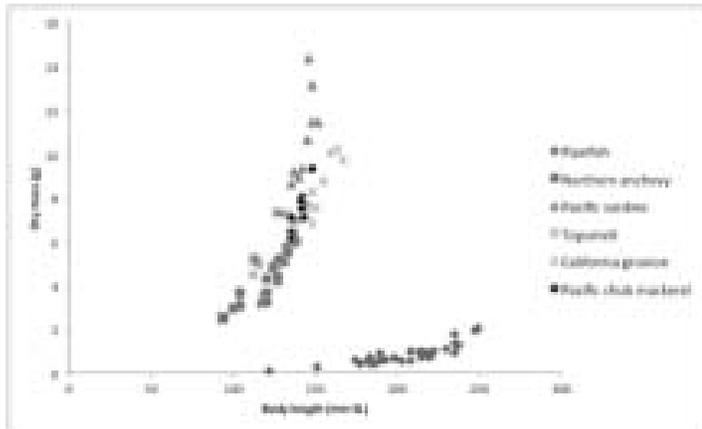


Figure 1 – Scatter plot of body mass and length of the six major prey species from the Elegant Tern chick diet for 10 years of a 16-year interval (1994-2009).

chicks fledged), or chicks fledging below normal weight and with too few lipid reserves to sustain them while they learn to feed for themselves.

To begin the process to determine any effects of an increase of pipefish consumption on the number of nesting pairs of Elegant Tern chicks at BCER, we tested three hypotheses in this study: (1) that pipefish are lower in energy density, protein and lipid content than other major prey species; (2) that over the 16-year study period the increase of pipefish resulted in decreased energy delivered to Elegant Tern colony at BCER; and a (3) the decrease in energy delivered to the chicks results in a decrease in Elegant Tern nest numbers at BCER.

Methods

The major prey species in the Elegant Tern diet were determined by calculating species proportions in dropped fish collected over a 16-year

period (1994-2009) at BCER. The energy, lipid, protein, and ash contents of these six main prey species delivered by Elegant Terns to their chicks for the 10 years of the 16 years with the largest data sets were determined to obtain the annual nutritional value of the fish delivered to the chicks. The 10 years for which we had the most dietary data at BCER were 1994, 1995, 1996, 1997, 1999, 2001, 2004, 2007, 2008 and 2009. Pipefish, northern anchovy, Pacific sardine, topsmelt (*Atherinops affinis*), California grunion (*Leuresthes tenuis*), and Pacific chub mackerel (*Scomber japonicus*), which together made up $\geq 75\%$ of the Elegant Tern total diet for the 10 years, were collected using a either a beach seine or beam trawl at Cabrillo Beach in Los Angeles Harbor or obtained from a bait barge in Long Beach Harbor.

The energy content (kJ g^{-1}) of each fish species was determined using a two-step method. First, total lipids were extracted using a Soxhlet apparatus with petroleum ether as a solvent, and, second, protein and ash contents were determined through incineration using a muffle furnace. The fish were dried to a constant mass in a convection oven at 60°C . The fish were homogenized using a mortar and pestle and weighed before the lipids were extracted. Once the lipids were removed, the remaining matter was dried and weighed before placing it in the muffle furnace for ten hours at 600°C for the incineration of the proteins. The remaining inorganic material was weighed and represented the ash content of the prey.

The lipid, protein, and ash contents were determined using a subtraction according to the equation from Harris et al. (2008), and energy content was determined using energy equivalents of 39.3 kJ g^{-1} for lipids and 17.8 kJ g^{-1} for proteins. The total lipid, protein and ash values obtained from this equation were used along with the available provisioning rate (4 fish per day per chick) for the Elegant Tern (Cole 1997) to determine the amount of energy delivered to the chicks, i.e., an energy-delivery profile for each of the 10 nesting seasons.

Prey species (n)	Regression equation	R ² value
pipefish (12)	$y = 0.0108e^{0.0001x}$	0.848
northern anchovy (12)	$y = 0.3703e^{0.0001x}$	0.898
Pacific sardine (8)	$y = 0.3067e^{0.0001x}$	0.891
topsmelt (12)	$y = 0.1842e^{0.0001x}$	0.947
California grunion (12)	$y = 0.1842e^{0.0001x}$	0.947
Pacific chub (mackerel) (11)	$y = 0.2487e^{0.0001x}$	0.742

Table 1 – Equations and R² values obtained from the body length (mm SL) and dry mass (g) relation of the six major prey fish species in the diet of Elegant Tern chicks

Prey species	Regression equation	R2 value
Pipefish (12)	$y = -1.5996x + 24.6$	0.046
Northern anchovy (12)	$y = -11.705x + 73.914$	0.840
Pacific sardine (8)	$y = -0.5526x + 19.813$	0.006
Topsmelt (12)	$y = -4.3726x + 36.634$	0.605
Grunion (12)	$y = -4.2936x + 37.922$	0.299
Pacific chub mackerel (11)	$y = 11.909x - 42.348$	0.216

Table 2 – Equations and R2 values obtained from the body length (mm SL) and energy content relation of the six major prey fish species in the diet of Elegant Tern chicks

The equations used to determine lipid, protein and ash contents are as follows:

$$\text{lipid content (g)} = \text{dry mass minus lipid-free mass}$$

$$\text{protein content (g) diet} = \text{lipid-free mass minus inorganic (ash) mass.}$$

The average dry mass of each major prey species in the chick diet for the 10 years was calculated and then multiplied by its energy density. The equation for energy delivered provides an estimate for the amount of energy received by a single chick each season, $Ed = (v)(R)(P)$, where Ed represents the estimate of the amount of energy delivered to each chick, v represents the energy content of each prey species, $R=140$ (4 provisions/day x 35 days for the fledging period) represents the average number of provisions per chick each season and P represents the proportion of each prey species present in the diet of the chicks. Total prey amount was adjusted to equal 100% for each year to be able to compare the different proportions of the major prey over the 16-year period. The total prey value represents the major prey species in the Elegant Tern chick diet.

Results

Body length (mm SL) and dry mass (g) of each prey fish species were positively correlated (Table 1, Figure 1). For the prey species, except pipefish, the body mass increased steeply with increased body length, but the body lengths of the specimens obtained for these five species were within a narrow range. Conversely, pipefish body mass increased much more gradually and the specimens represented a wider range of body lengths. The differences in the rate of change in body growth compared to length in pipefish and the other Elegant Tern prey species (Figure 1) reflect the fact that pipefish are long and slender while the other species are thicker and deeper bodied. This relationship can best be understood by looking at the slope of the

corresponding regression equations (Table 1). All prey species analyzed, except pipefish, had a very large initial value for body mass, which translated to a steep increase in the dry mass as body length increases. The high R^2 values suggest a strong relationship between body length and dry mass for all six major prey species (Table 1).

Standard length and energy content varied among the six major prey species of the Elegant Tern. Five of the prey species, pipefish, northern anchovy, Pacific sardine, topsmelt and California grunion, exhibited a negative correlation of standard length and energy content. Pacific chub mackerel, however, showed a positive correlation between standard length and energy content (Table 2).

Mean lipid, protein, ash and total fish energy contents varied among the six major prey species (Table 3). Northern anchovy showed the highest lipid content while Pacific chub mackerel exhibited the lowest lipid content. The four other prey species had intermediate lipid contents. Similar mean protein values were seen among the six species with a range of 72% to 80%. Ash values for northern anchovy, Pacific sardine and Pacific chub mackerel were around 14% while that for California grunion was 16.5%. Topsmelt and pipefish had the highest ash contents at >19%. Similar mean energy densities were observed for all six species with a range of 15.9 kJ g⁻¹ to 17.9 kJ g⁻¹.

The whole-fish energy content “per bite” varied greatly among the prey species. Whole-fish energy content refers to the energy content of one whole fish of a particular species. Pipefish had a whole-fish energy content (4.2 kJ), nearly an order of magnitude lower than the five other prey species. The total energy provided by an average-sized Pacific sardine (56.6 kJ) is 14 times greater than that of an average-sized pipefish. The content for the remaining four species showed intermediate increase from topsmelt (23.8 kJ), to Pacific chub mackerel (32.3 kJ), to northern anchovy (47.8 kJ) and California grunion (51.3 kJ) (Table 3).

Species	Length (mm)	Dry mass			Energy density (kJ g ⁻¹ dry mass)	Total fish energy content ("per bite") (kJ)
		(g)	% Lipid	% Protein		
pipefish (12)	212.8 ± 6.5	0.8 ± 0.1	7.5 ± 0.8	73.1 ± 0.3	18.5 ± 0.6	4.2 ± 0.4
northern anchovy (12)	124.4 ± 2.4	4.7 ± 0.3	12.2 ± 1.8	73.8 ± 1.6	14.2 ± 0.5	47.8 ± 5.8
Pacific sardine (8)	140.4 ± 2.3	8.5 ± 0.4	7.8 ± 0.8	78.4 ± 1.0	13.73 ± 0.8	65.0 ± 18.5
topsmelt (12)	114.5 ± 3.8	5.2 ± 0.5	8.2 ± 0.7	72.4 ± 0.8	18.1 ± 0.4	38.8 ± 0.4
California gnatfish (12)	163.4 ± 3.7	8.8 ± 0.4	7.4 ± 1.0	75.7 ± 0.6	16.8 ± 0.6	61.3 ± 8.6
Pacific chub mackerel (11)	149.7 ± 4.1	7.5 ± 0.5	5.1 ± 0.5	80.5 ± 0.8	13.8 ± 0.4	32.3 ± 9.2

Table 3 – Composition and energy content (mean ± SE) of major prey species of Elegant Tern chicks. Total fish kilojoules are based on average size of each species delivered to chicks across all seasons of record.

The proportions of the six prey species provisioned by Elegant Terns varied over the 10 years. Pipefish proportions increased in the chick diet in recent years, while proportions of other major prey species, with higher energy content, decreased overall (Figures 2 and 3), especially since 2001. Pipefish deliveries by Elegant Terns have almost doubled over the 16-year period, whereas northern anchovy deliveries have decreased by about 50% (Table 4). We found, however, that the difference in the average amount of energy delivered between 1994-1999 and 2001-2009, representing intervals before and during the pipefish increase, was not significantly different (t-test, $p > 0.43$).

Discussion

The important result of our study was that the increase of pipefish in the Elegant Tern chick diet appeared to not have an effect on the total energy delivered to the chicks or Elegant Tern nest numbers at BCER. Pipefish and the other major prey species in the Elegant Tern chick diet had similar mean energy densities (kJ g⁻¹ dry mass). Pipefish did not exhibit a lower lipid or protein content than most of the other major prey species, yet had a higher ash content, with the exception of topsmelt. However, we still consider pipefish a low-quality prey fish species because of its low whole-fish energy content. The high ash content and slender body of pipefish could account for their low "per bite" value. These results rejected our first and second hypotheses, which predicted that pipefish are lower in energy, protein and

lipid content as compared to other major prey species and that the increase of pipefish resulted in a decrease of energy delivered to the chicks.

The adjusted energy delivered to the Elegant Tern chicks for the 10 years of the study shows no clear relationship between pipefish contribution in the Elegant Tern chick diet and total energy contribution from the top prey species (Figure 3). Since 2001, pipefish proportion has increased dramatically in the Elegant Tern chick diet; however, the total amount of energy delivered

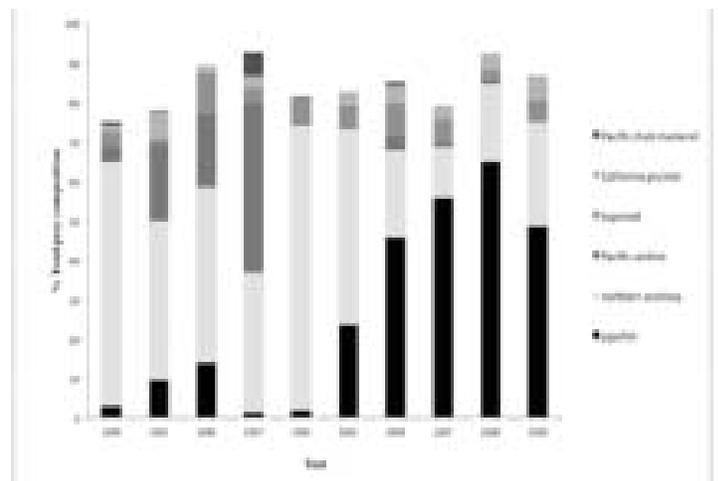


Figure 2 – Composition of the major prey species recorded in the Elegant Tern chick diet from dropped fish for the 10 years of 16 recent years

Species	1994	1995	1996	1997	1998	2001	2004	2007	2008	2009
pipefish	3.89	12.03	18.03	1.37	3.11	38.38	53.70	70.36	70.28	55.84
northern anchovy	49.64	52.34	50.00	38.33	48.73	40.59	26.21	16.20	21.38	30.38
Pacific sanders	3.29	24.78	70.58	46.38	0.35	0.42	4.08	1.02	0.78	0.20
topsmelt	4.98	1.09	11.08	4.15	0.80	0.38	0.70	7.14	3.80	8.18
California grunion	2.34	0.00	1.58	0.48	0.00	4.24	5.38	4.07	4.84	7.44
Pacific chub mackerel	0.85	0.18	0.00	0.32	0.00	0.00	1.07	0.00	0.00	0.00
Total (100%)										

Table 4 – Adjusted proportions (based on 100% for total) of the six major prey species in the deliveries to Elegant Tern chicks for the 10 years of the 16-year interval

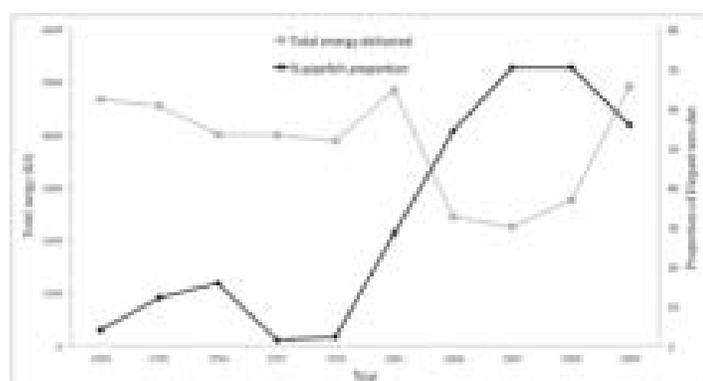


Figure 3 – Estimates for total energy delivered to Elegant tern chicks each year with the adjusted proportion of pipefish in the diet each year

to the chicks over the period has fluctuated. Total energy delivered was lowest when pipefish make up the greatest proportion of the chick diet, with exception of 2009 when the dropped samples of northern anchovy, topsmelt and California grunion represented the largest body sizes in our data set for these species and thus provided more energy “per bite” than previous years. The average amount of energy delivered between 1994-1999 and 2001-2009, however, was not significantly different for these two intervals.

The number of Elegant Tern nest pairs at BCER fluctuated over the 16-year period. Nest pairs were lowest in 2001 when the pipefish proportion in the Elegant Tern chick diet began increasing and were highest during 2007-2009 when pipefish made up a greater proportion of the chick diet and total

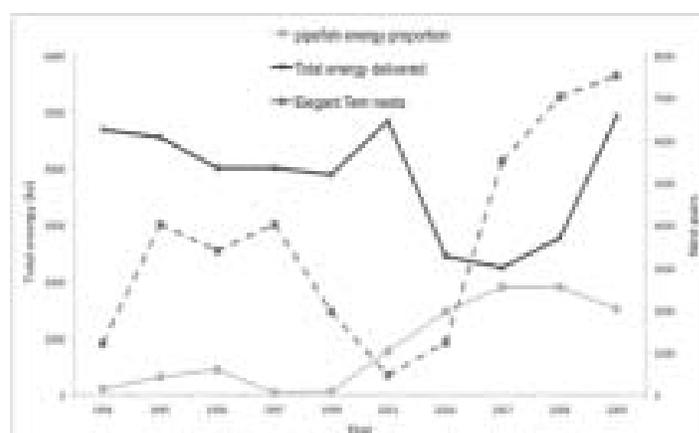


Figure 4 – The total energy delivered by Elegant Terns, contribution of pipefish to that total energy and the estimated numbers of nesting pairs of Elegant Terns at BCER for the 10 years during the 1994-2009 interval

energy delivered was lowest (Figure 4). This result rejects our third hypothesis.

The increase of pipefish proportions in the Elegant Tern chick diet along with dramatic increase of nest numbers at BCER from 2007 to 2009 suggests that the Elegant Tern nest numbers are not affected negatively by the increase of pipefish in the diet. However, the diet of the adult Elegant Tern prey species is not known and may possibly be different from that of their chicks. The adult diet of a closely related species, the Greater Crested Tern (*Sterna bergii*), is somewhat different from that of its chicks (McLeay et al. 2009). If this is true for the Elegant Tern, then the nest numbers at BCER might not have been

affected by the pipefish increase in the diet because the pipefish proportion in the adult diet may not have increased.

The observed increase in pipefish over the 16-year period, nevertheless, could affect the future of the Elegant Tern colony at BCER. Pipefish provide a smaller amount of lipids and energy per fish to the Elegant Tern chicks than any of the five other major prey species. The increase of pipefish in the Elegant Tern chick diet might result in chicks fledging below normal weight and with few lipid reserves to sustain them while they learn to feed for themselves; however, we could not draw such conclusions based on our results. In fact, Elegant Tern nest numbers have remained high during the rise of pipefish in the chicks' diet. Additional studies are needed to examine food quality and its effect on the productivity including chick body composition and fledging rate of the Elegant Tern colony nesting at BCER.

Acknowledgements

Financial support for this research was provided by California State University Fullerton as a part of a Faculty-Student Research and Creative Activities Grant. Thank you.

References

- Burness, G.P., K. Lefevre, and C. T. Collins. 1999. Elegant Tern (*Sterna elegans*). In *The Birds of North America*, no. 404 (A. Poole and F. Gill, Eds.). Academy of Natural Sciences, Philadelphia, and American Ornithologists' Union, Washington, D.C
- Cole, P. 1997. Comparison of chick provisioning in Caspian Terns and Elegant Terns at a mixed-species colony in Southern California. M.A. thesis, California State University, Fullerton.
- Dahdul, W.M., and M.H. Horn. 2003. Energy allocation and postnatal growth in captive Elegant Tern (*Sterna elegans*) chicks: Responses to high- versus low-energy diets. *The Auk* 120: 1069-1081
- Harris, M.P., Newell, F.D., Speakman, J.R., and S. Wanless. 2008. Snake Pipefish *Entelurus aequoreus* are poor food for seabirds. *Ibis* 150: 413-419
- Hendricks, J. 2011. Restoring potential seabird foraging habitat: effects on the feeding ecology of the Elegant tern (*Thalasseus elegans*) in southern California. M.S. thesis, California State University Fullerton.
- McLeay, L.J., Page, B., Goldsworthy, E.S.D., Ward, T.M., Paton, D.C. 2009. Size matters: variation in the diet of chick and adult crested terns. *Marine Biology* 156:1765–1780
- Nagy, K.A., I.A. Girard, and T.K. Brown. 1999. Energetics of free-ranging mammals, reptiles, and birds. *Annual Reviews Nutrition* 19: 247-277
- Patton, R.E. Velarde, M.Horn, and K. Keane. 2010. Elegant Tern colonies in Gulf of California and California current sites: influence of oceanographic events on size and composition. Symposium: Marine birds: dynamic responses to interannual and interdecadal climate variability and change in the Gulf of California and southern California current: models and application. Joint meeting, Cooper Ornithological Society, American Ornithologists Union and Society of Canadian Ornithologists. Sand Diego, California, 7-11 February 2010.
- Reid, P.A., P.A. Prince, and J.P. Croxall. 2000. Fly or die: the role of fat stores in the growth and development of Grey-headed Albatross *Diomedea chrysostoma* chicks. *Ibis* 142: 188-198
- Romano, M.D., Piatt, J.F., and D.D. Roby. 2006. Testing the junk-food hypothesis on marine birds: effects of prey type on growth and development. *Waterbirds* 29: 407-524
- Taylor, J.R.E., and M. Konarzewski. 1989. On the importance of fat reserves for the little auk (*Alle alle*) chicks. *Oecologia* 81: 551-558
- Wanless, S., M.P. Harris, P. Redman, and J.R. Speakman. 2005. Low energy values of fish as a probable cause of major seabird breeding failure in North Sea. *Marine Ecology Progress Series* 294: 1-4

Measuring Xylem Vulnerability in Three Citrus Species

Department of Biological Science, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

Emily Nguyen Wieber

Advisor: Dr. H. Jochen Schenk

Abstract

This study documented xylem vulnerability of *Citrus x sinensis* (Valencia orange), *Citrus x paradisi* (star ruby red grapefruit), and *Citrus x reticulata* (tangerine) under drought conditions. During water shortages, a better understanding of drought tolerance mechanisms in citrus would enable to improve water conservation. Physiological mechanisms of drought resistance were investigated by measuring the formation of xylem embolisms, also known as air bubbles that prevent water flow through vascular tissue. Four methods were compared to determine the best technique for hydrating the stems in order to achieve maximum hydraulic conductivity through the xylem: 1) high pressure flushing for one hour at room temperature, 2) sixteen-hour vacuum treatment of submerged stems at room temperature, 3) low pressure flushing at 12°C for 16 hours, and 4) one-hour vacuum treatment of submerged stems followed by one-hour high pressure flushing at room temperature. Based on previous results and studies, we hypothesized that submergence under a vacuum would be the best method of stem hydration because it would effectively remove trapped air molecules from pit membranes. Based on differences in fruit sizes, we hypothesized that grapefruit would be the most vulnerable to embolism formation. High-pressure flushing was a widely used standard method that produced the lowest hydraulic conductivity, whereas, long-term vacuum and long-term low pressure flushing both caused the highest hydraulic conductivities. The three Citrus species were not significantly different in their vulnerability to embolism formation. In the future, a long-term study would allow for a better understanding of plant hydraulic conductivity in drought-prone regions.

Introduction

Citrus plants are widely grown in tropical and Mediterranean climates. The genus originates from South East Asia (Monrovia, 2010). Citrus farming was

once the main fruit crop in California (Cooley et al. 2009). Due to California's Mediterranean climate, there is a high demand for irrigation dedicated to citrus production, which could result in adverse economic and possibly ecological effects. Orange County was once responsible for producing large amounts of high quality Citrus. With the advancements being made in agricultural farming and urban sprawl, many of these farming communities have moved to rural areas. Yet, a large-scale citrus production still persists in the adjacent counties.

In order to lessen the economic and environmental impact and to increase water conservation it is necessary to examine drought tolerance of citrus trees. More knowledge regarding drought tolerance mechanisms of Citrus could lead to more efficient Citrus selection of varieties for dry-summer climates and could possibly increase water conservation via improved irrigation scheduling. Therefore, it is important to investigate the physiological mechanism of citrus adaptation to a drought prone environment (Poggi et al. 2007).

One crucial physiological aspect of drought resistance in plants is their vulnerability to form xylem embolisms (Hacke & Sperry 2001; Melcher et al. 2003; Poggi et al. 2007). A tree under normal conditions utilizes negative xylem pressure to draw water from the ground into the leaves, replenishing water lost to transpiration and transportation of nutrients. Under drought conditions, a limitation of soil water availability can cause the formation of xylem embolisms, as air is drawn into the xylem through small pores called pit membranes, a process known as air seeding (Hacke & Sperry 2001). Embolisms impede the flow of water through the xylem and, if not removed, can cause leaves to die off or lead to the death of the plant. The process of forming embolisms, often referred to incorrectly as cavitation, occurs when negative pressure in the xylem drops below a certain threshold. In many studies of plant species, a lower vulnerability towards forming embolisms indicates a high tolerance to water deficiency (Poggi et al. 2007). Yet, there

are not many studies of xylem vulnerability for fruit species and only one study of one citrus variety (Poggi et al. 2007). Therefore, the purpose of this study was to document the xylem vulnerability of three Citrus species: Citrus x. sinensis (Valencia orange), Citrus x paradisi (Star Ruby Red grapefruit), and Citrus x. reticulata (Tangerin).

Xylem vulnerability can be measured by determining how much positive air pressure is required to push air laterally into the xylem of detached stems (Sperry & Saliendra 1994). If embolisms are formed in the xylem conduits, the hydraulic conductivity of the stems decreases. A xylem vulnerability curve documents the percent loss of hydraulic conductivity (PLC) as a function of pressure. Measuring xylem vulnerability curves of three citrus species allows gaining insights into drought tolerance mechanisms in Citrus and may aid in developing improved irrigation schedules and water conservation in California.

Three Citrus species were studied Citrus x sinensis (Valencia Orange), Citrus x paradisi (Star Ruby Red Grapefruit), and Citrus x reticulata (Tangerin). Stems of these three species were collected from the Fullerton Arboretum in California. The purpose of this study was to document xylem vulnerability in three citrus species C. x sinensis, C. x paradisi, and C. x reticulata.

To do this, maximum hydraulic conductance of fully hydrated stems had to be measured as a baseline for the measurements. A recent study has found that the standard method of hydrating stems by flushing them with water under positive pressure may not always result in full hydration (Espino & Schenk, 2010). Therefore, the first step of this research project was to determine a standard method to hydrate Citrus stems. This method was then used for all measurements of xylem vulnerability.

Four stem hydration methods were tested, including short term high pressure flushing (T1), long term vacuum (T2), overnight low temperature, low pressure flushing (T3), and a combination of high pressure flushing and short term vacuum (T4). Based on findings by Espino & Schenk (2010), it was hypothesized that the best method to hydrate citrus stems would be to submerge them in an under-water vacuum for 16 hours. The long term vacuum was predicted to out-compete the other methods due to its ability to generate a long-lasting pressure differential that allows for efficient diffusion of air molecules from embolized vessels through pit membranes into water-filled cells.

Based on published research regarding irrigation demands (FAO 2010) and on differences in fruit sizes, it was hypothesized that C. x paradisi would be the most vulnerable to embolism formation. This was based on

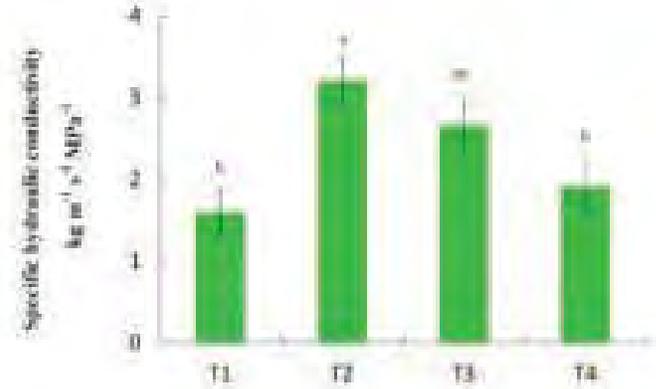


Figure 1 – The effects of hydration treatments (n= 10) on maximum hydraulic conductivity of three Citrus species. T1: 1 hr high pressure flushing, T2: 16 hr vacuum, T3: 16 hr low temperature, low pressure flushing, T4: 1 hr vacuum followed by 1 hr high pressure flushing. Data are least-square means across all three species. Across all species, treatment 2 had significantly higher conductivity ($p < 0.05$) than treatments 1 and 4. Error bars = SE.

the assumption that a higher demand for water corresponded to a higher vulnerability to embolism formation.

Materials and Methods

Collecting stems samples

Stems of three Citrus species: Citrus x sinensis (Valencia orange), Citrus x paradisi (Star Ruby Red Grape Fruit), and Citrus x reticulata (Tangerin) were collected from the Fullerton Arboretum, CA. Fresh stems with diameters between 2.3 mm and 5.0 mm and length of 25 to 40 cm for stem hydration experiments and longer than 80 cm for xylem vulnerability measurements were collected. Stems were cut and placed in a bucket of water which was transported to the plants and H₂O lab at Cal State Fullerton; green stems with minimal or no branching were selected. In the lab, stems were submerged in a container containing de-ionized (DI) water. For stem hydration experiments, stems were successively cut down to 15cm lengths on both ends of the stems, cutting 1 cm at a time. The bark at the proximal end of the stems was removed.

Hydraulic conductivity measurements

Hydraulic conductivity was measured as described by Espino and Schenk (2010). A XYL'EM embolism meter (Bronkhorst, Montigny les Corneilles, France) was used to measure the flow rate of water through the stems. Stems

were connected from their proximal ends to a manifold, which was connected via tubing to the XYL'EM. Water used for measurements and for hydrating stems was degassed using a membrane contactor degassing unit (Liqui-Cel mini-module 1.7 x 5.5, Membrana, Charlotte, NC, USA) and passed through a 0.2 µm filter (model Polycap AS 75, Whatman Inc., Piscataway, NJ). Initial flow rate of water moving into stems was measured in the absence of a pressure differential, (F_0 at 0 kPa). Stems were then measured under a pressure differential of 6 kPa which was generated by raising the water supply relative to the stems. Once the flow rate was stable, typically a duration of one to five minutes, the flow rate was recorded as F_p . Hydraulic conductivity, k was determined as flow rate at 6 kPa, F_p minus flow rate at 0 kPa pressure divided by the pressure differential, Δp : $k = (F_p - F_0) / \Delta p$.

Specific hydraulic conductivity (k_s) of a stem with certain length L and area A was calculated as flow rate at 6 kPa minus flow rate at 0 kPa times length and divided by the differential pressure and by area: $k_s = (F_p - F_0) L / \Delta p A$. All hydrating conductivities reported in this study include a temperature correction to 20 °C to allow for changes in water viscosity with temperature.

Hydration treatments

Citrus stems of all three species were hydrated using four treatments to determine the optimal hydration technique; treatment 1: Ten stems were flushed with DI, degassed water for 1 hour at 150kPa using the high-pressure chamber in the XYL'EM apparatus. This is the standard hydration technique used in almost all xylem vulnerability studies (Espino & Schenk, 2010). Treatment 2: Ten stems were submerged under DI water for 16 hours in a vacuum chamber under 3 kPa absolute pressure, generated using a vacuum diaphragm pump (model DAA-V715A-EB, Gast, Benton Harbor, MI, USA). This treatment was recommended by Espino and Schenk (2010) as to avoid the bubble formation that can be caused by high pressure flushing. Treatment 3: Ten stems were connected to one manifold and were flushed under low pressure with DI, degassed water that was cooled to 12°C before entering the stems. Cooling increases gas solubility in water (Mercury et al. 2003). Flushing was done for 16 hours at a rate of 10g hr⁻¹ for 10 connected stems (i.e., approximately 1 g hr⁻¹ per stem). Treatment 4: Ten stems were submerged under DI water in a vacuum chamber for 1 hour under 3 kPa absolute pressure. This treatment would first remove air from open vessels at both ends of the stems, which could help to avoid bubble formation under the subsequent flushing under high pressure.

Xylem Vulnerability Curves

Xylem vulnerability curves are measurements of the percent loss of hydraulic conductance (PLC, relative to the maximum conductance, k_{max}) as a function of xylem pressure.

The air injection method (Sperry and Saliendra, 1994; Meltcher et al., 2003) was used to measure xylem vulnerability in Citrus. For Valencia Orange stems with 100 cm length, vessel length of the stems were determined by gradually cutting 1 cm until the bubbles being released were observed. The vessel length of an orange stem was 60 cm, as was the vessel length for grapefruit and tangerine. Stems were notched 1 cm apart on two sides of the stem in order to allow for air entry into the xylem. Notches were made by cutting at an angle directly into the xylem, only into the section that was subjected to air pressure; then 65 cm to 100 cm long Citrus stems were hydrated as described in treatment 2. After hydration, they were inserted into cavitation chambers (PMS Instruments, Albany, OR) with both ends protruding. The cavitation chambers were connected to a pressure chamber (model 1000, PMS instruments), which was used to raise the air pressure inside the cavitation chambers successively from 0 to a maximum of 10 MPa. After each successive increase in pressure, the pressure was held for 5 minutes then reduced to 0.1 MPa for about 1 to 2 minutes (Sperry and Saliendra, 1994), which was then reduced to an ambient pressure for measurements of hydraulic conductance using the XYL'EM apparatus. Percent loss of conductance (PLC) was calculated as:

$$PLC = 100 \left(1 - \frac{k}{k_{max}} \right)$$

where a PLC of zero % indicates completely filled xylem and a PLC of 100% indicate a completely embolized xylem. For all xylem vulnerability curves, k_{max} was defined as the k_s measured after application of 0.25 MPa pressure to correct for cavitation fatigue of the xylem (Jacobsen et al. 2007).

Data Analysis

Hydration treatments were compared among the three Citrus species by ANOVA using the statistical software SYSTAT (version 13, SYSTAT Software Inc., San Jose, CA). Xylem vulnerability curves were analyzed using nonlinear regression of PLC vs. air pressure. Logistic dose response curves were fitted to the data using the software TableCurve 2D (version 5.01, SYSTAT Software Inc., San Jose, CA), and the pressures at which PLC reached 50% (Ψ_{50}) was determined from these regressions.

Results

The long-term vacuum treatment (T2), long-term low-pressure and cold flushing treatment (T3) were equally effective in removing embolisms from Citrus vessels (Fig. 1). However, the vacuum treatment was superior to high-pressure flushing, whether combined with vacuum or not. The maximum specific hydraulic conductivity of stems after the long-term vacuum treatment was on average about twice as high as achieved after one hour of high-pressure flushing (Fig. 1), suggesting that the vacuum was much more effective for removing air bubbles from stems. Maximum vessel lengths determined by the air injection method was 60 cm for the Valencia orange, 60 cm for the Star Ruby Red Grapefruit, and 60 cm for the tangerine. Based on these findings on the length of stems, 100 to 120 cm were used for xylem vulnerability measurements were not significantly different in their vulnerability to xylem embolism formation (Fig. 2), as differences between calculated Ψ_{50} values (= pressure at 50% PLC) were not statistically significant.

The Valencia orange, Ruby Red grapefruit and tangerine were not significantly different in their vulnerability to xylem embolism formation (Fig. 2), as differences between calculated Ψ_{50} values (= pressure at 50% PLC) were not statistically significant. This was largely caused by the very high variability of conductivities in response to air injection into the stems, as indicated by the large scatter of points on all graphs of Fig.2.

Discussion

As hypothesized the best method for hydrating Citrus stems was the sixteen hour submerged vacuum treatment (T2). This method proved to be the most effective technique to remove air from the xylem. This method stands out the most among other treatments in providing a long-lasting partial pressure gradient between the gas phase in the vacuum chamber and the gas phase in embolized vessels, which drives the movement of air molecules from vessels through pit membranes into water-filled vessels and ultimately through the water's surface in the vacuum pump. As Espino and Schenk (2010) found previously, submergence under vacuum effectively removed air from vessels, while high-pressure flushing was found to be the least effective in removing air from vessels. High-pressure flushing, the widely used standard method produced the lowest hydraulic conductivity, whereas, long-term vacuum and long-term low pressure flushing both caused the highest hydraulic conductivities. As air removal from wood occurs by way of diffusion, a relatively slow process, the longer the time, the more efficient

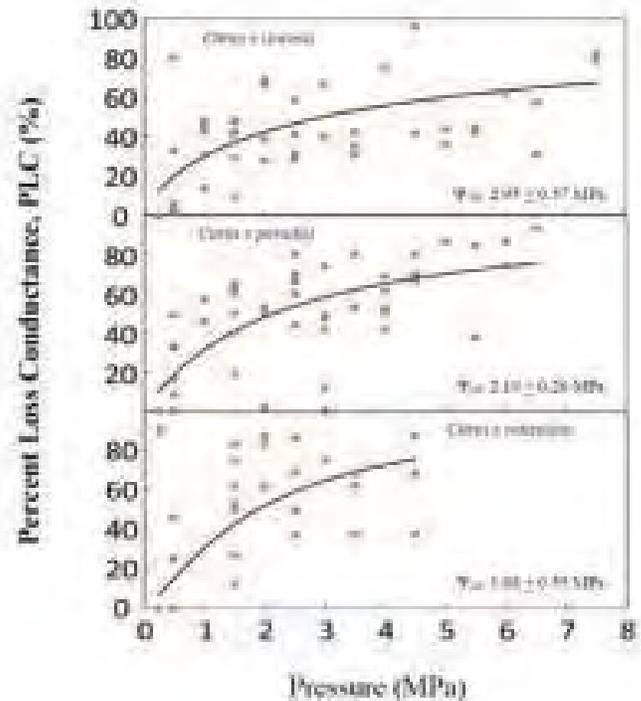


Figure 2 – Xylem vulnerability curves for three Citrus species (n=8). Logistic regressions were used to determine Ψ_{50} , which is the pressure at which 50% of PLC was observed.

the removal of air will be. The second vacuum treatment, being the only one to be significantly different from treatment 1 and, was also more feasible to implement as a standard method than treatment 3, was adopted as the method of choice for hydrating Citrus stems.

The second hypothesis, that grapefruit would be the most vulnerable to embolism formation than orange and tangerine, was not supported. The three Citrus species were not significantly different in their vulnerability to embolism formation, although there appeared to be a trend that showed that Valencia oranges were the most resistant species and the tangerines were the most vulnerable. The inability to detect statistical differences between the vulnerabilities of the three species was entirely due to large variances within species, resulting in large standard errors (Fig. 2). The widely distributed data points suggests that the air injection method, as applied, did not result in highly reliable and reproductive measurements of xylem vulnerability. The xylem vulnerability curves of all three species did not go up to 100%

PLC over the range of pressures measured. Poggi et al. (2007) documented a pattern of xylem cavitation for Clementine that was similar to that of the three Citrus species studied here. These authors suggested that xylem conduits in Citrus may be divided into more vulnerable and less vulnerable “compartments”, but could not find evidence for their species to support this hypothesis.

For future works, it would be constructive to study the xylem’s anatomical structure of the three Citrus species. The anatomical xylem analysis of the three Citrus species would examine any difference in xylem straits that have been associated with vulnerability to embolism formation in previous studies, including vessel diameters and the area of pits connecting vessels (Hacke et al. 2009). Such measurements could be used to test the xylem compartments hypothesis of Poggi et al. (2007), for example if vessels turned out to be present in two distinct diameter classes. Also, the three Citrus species turned out to be anatomically similar, which would explain why the three Citrus species were not statistically different in their vulnerability to embolism formation.

As a follow up study on the findings of the research reported here, I would like to conduct a study of the effects of Citrus root grafting on hydraulic traits of Citrus. Root grafting is currently utilized within the agricultural field to increase the yield and control the longitudinal growth of Citrus trees. Citrus farmers use the practice of grafting one species of Citrus trees to a developed root system of a secondary citrus species to shorten the growth rate which it will take a species to develop from a seedling to an adult tree capable of producing fruit. Another useful and relevant study for the future could be on the effects of grafting on the formation of embolisms within the trees’ roots and stems by comparing grafted and un-grafted trees. This would allow me to test the hypothesis that root grafting leads to the development of more embolism due to a hydraulic mismatch between roots and stems. Steppe et al. (2006) reported evidence for such a hydraulic mismatch in orange trees by observing peculiar stomatal oscillations that may have been due to an insufficient hydraulic conductance of the root system.

In the future, I would like to expand my study within the area of water usage for citrus species and work with Citrus farmers to develop and test root irrigation systems either above or below ground and generate improved irrigation schedules to allow for effective water conservation within drought prone areas, which would ultimately lower costs in yield production for the farmer. I believe that this would facilitate a thriving agricultural economy here in California that could persist in the face of climate change. Moreover,

lower cost production of locally grown produce would be more attractive to local area consumers within a targeted community and would allow for a competitive job growth of small agricultural farmers with the potential of creating job opportunities in a given area.

In conclusion, this study has resulted in a new standard method to hydrate stems for measurements of xylem vulnerability curves, which will now be tested on other plant species. The research on xylem vulnerability in Citrus has remained inconclusive, but will continue to develop better experimental protocols and increase the reliability of the technique.

References

Cooley, H., Christian-Smith, J., Gleick, P. 2009. Sustaining California Agriculture in an Uncertain Future. Pacific Institute, July Report, 1-81.

Espino, S. and Schenk, H. J. 2010. Mind the bubbles: Achieving stable measurements of maximum hydraulic conductivity through woody plant samples. *Journal of Experimental Botany*. Vol. 62, 1119-1132.

FAO 2010. Crop water information; Citrus. Available from http://www.fao.org/nr/water/cropinfo_citrus.html.

Hacke, Uwe., Sperry, John. 2001. Functional and ecological xylem anatomy. Urban and Fischer Verlag. Vol. 4/2, 97-115.

Hacke, U. G., A. L. Jacobsen, and R. B. Pratt. 2009. Xylem function of arid-land shrubs from California, USA: an ecological and evolutionary analysis. *Plant, Cell & Environment*. Vol. 32, 1324-1333.

Mercury, L., M. Azaroual, H. Zeyen, and Y. Tardy. 2003. Thermodynamic properties of solutions in metastable systems under negative or positive pressures. *Geochimica et Cosmochimica Acta*. Vol. 67, 1769-1785.

Monrovia [Internet]. [cited 2010 Oct 10]. Campbell Valencia Orange. Available from <http://www.monrovia.com/plant-catalog/plants/789/campbell-valencia-orange.php>

Poggi, I., Polidori, J., Gandoin, J., Paolacci, V., Battini, M., Albertini, M., Ameglio, T., and Cochard, H. 2007. Stomatal regulation and xylem cavitation in Clementine (*Citrus clementina* Hort) under drought conditions. *Journal of Horticultural Science & Biotechnology*, Vol 82, 845-858.

Sperry, J. S., Donnelly, J. R., and Tyree, M. T. 1988. A Method for measuring hydraulic conductivity and embolism in xylem. *Plant, Cell and Environment*. Vol 11, 35-40.

Sperry, J. S. and N. Z. Saliendra. 1994. Intra- and inter-plant variation in xylem cavitation in *Betula occidentalis*. *Plant, Cell and Environment* 17:1233-1241.

Steppe, K., S. Dzikit, R. Lemeur, and J. R. Milford. 2006. Stomatal oscillations in orange trees under natural climatic conditions. *Annals of Botany*, Vol. 92, 831-83.

Maximum Likelihood Estimation of the Fat Fraction Using Magnetic Resonance Imaging at High Signal-to-Noise Ratio

Department of Mathematics, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

Jorly Chatouphonexay
Advisor: Dr. Angel R. Pineda

Abstract

For our project, we are using mathematics to improve the use of magnetic resonance imaging (MRI) to diagnose nonalcoholic fatty liver disease (NAFLD). NAFLD is diagnosed by using the fat fraction. We defined the true fat fraction as $\rho_{true} = \frac{\mu_{fat}}{\mu_{fat} + \mu_{water}}$, where μ_{fat} is the mean of the signal and μ_{water} is the mean of the water signal. Since true fat fraction is unknown, it is estimated by two methods at high signal-to-noise ratio (SNR): the clinical method and the maximum likelihood estimation (MLE) method. With the clinical method, the fat-fraction estimate is generated by $\hat{\rho}_{clinical} = \frac{|F|}{|F| + |W|}$. The magnitude of the fat and water signals are denoted as $|F|$ and $|W|$, and follow the Rician distribution. However, at high SNR, they approximately follow the normal distribution. With the MLE method, the fat-fraction estimate is generated by $\hat{\rho}_{MLE} = \frac{\hat{F}_{MLE}}{\hat{F}_{MLE} + \hat{W}_{MLE}}$ where \hat{F}_{MLE} is the estimated fat signal and \hat{W}_{MLE} is the estimated water signal obtained by the MLE method. In this work we will show that the MLE method leads to a smaller mean squared error (MSE) in estimating the fat fraction than the standard (clinical) method.

Introduction

Non-alcoholic fatty liver disease (NAFLD) represents one of the most common chronic liver conditions in western society, due to the increased rates of obesity and diabetes [1,5,9]. In the United States alone, NAFLD affects 30% of adults and 10% of children [7]. NAFLD is a spectrum of liver

conditions, ranging from simple steatosis to non-alcoholic steatohepatitis (NASH), which can lead to cirrhosis, hepatocellular carcinoma, and liver failure. NAFLD is detected by the changes of fat in the liver when alcohol consumption is considered to be low (<20 g/day) [1,5,9].

Currently, a liver biopsy is the standard for the diagnosis of NAFLD, but this method is invasive [7]. Additionally, a liver biopsy suggests that the tissue sample represents the organ while the biopsy samples 1/50,000th of the liver. Although a liver biopsy is preferred for diagnosis, it is prone to high sampling error from sampling small amount of liver tissue. Often, a biopsy is an expensive and painful operation to perform [1,5,7,9,10].

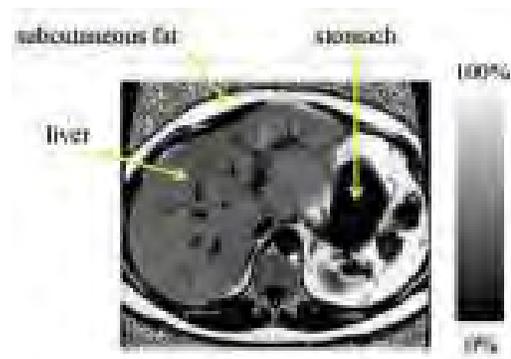


Figure 1 – A clinical image showing the fat fraction from an MRI scan. Shown here, we see how the fat fraction varies in the liver (mostly fat), subcutaneous fat, and stomach (mostly water) tissues. From the amount of fat in the liver, this image shows that this patient has a fat fraction of 50%. Patients with NAFLD exhibit a fat fraction of 10% [8]. Image courtesy of Scott Reeder, used with permission of the authors of [2,3].

As an alternative to a liver biopsy, we find that magnetic resonance imaging (MRI) may be a better option since the procedures of MRI are non-invasive. Also with MRI, this procedure visualizes the internal structures within the human body. An MRI scan is an ideal method over the liver biopsy since a scan provides a clear contrast between the different types of tissues of the human body [7]. In contrast to the other medical imaging techniques, MRI does not expose patients to radiation. For example, computed tomography (CT) uses ionizing radiation in order to produce images of the inside of the human body. Also, MRI is more reliable in differentiating changes in the liver's fat content than ultrasound, which uses high-frequency sound waves to produce images of the inside of the human body [9].

Fat Fraction

To quantify NAFLD, we would need to quantify the fat fraction. The fat fraction is the amount of fat over the amount of fat and water in the liver. A clinical image of a fat fraction is shown in Figure 1. This fraction quantifies how much fat is contained in a tissue. The model for the complex water and fat signals is

$$M = (\mu_r + \epsilon_r) + i * (\mu_i + \epsilon_i)$$

where μ_r is the mean of the real component and μ_i is the mean of the imaginary component of the measured signal, ϵ_r as the sampling error associated with the sampling of the real part of the signal, and ϵ_i is the sampling error associated with the sampling of the imaginary part of the signal. Note, the mean of the measured signal is defined by the equation below

$$\mu_m = \sqrt{\mu_{m,r}^2 + \mu_{m,i}^2}$$

where $\mu_{m,r}$ is the mean for the real part and $\mu_{m,i}$ is the mean for the imaginary part of our model. In this model, M is the complex signal from the MRI, and the probability density function (pdf) of M follows the Rician distribution.

The true fat fraction is given by

$$\rho_{true} = \frac{\mu_{fat}}{\mu_{fat} + \mu_{water}}$$

where μ_{fat} is denoted as the mean of the fat signal and μ_{water} is denoted as the mean of the water signal. Since the true fat fraction is unknown, we have two methods to estimate it: the clinical method and the maximum likelihood estimation (MLE) method.

Methods

Clinical Method

The measurement model for the fat-fraction estimate by the clinical method is

$$M = (M_r + \epsilon_r) + i * (M_i + \epsilon_i),$$

where M_r is the real and M_i is the imaginary components of the measured signal. In our algorithm, we defined M_r to follow a normal distribution with mean μ_r and standard deviation σ_r , M_i to follow a normal distribution with mean μ_i and standard deviation σ_i , ϵ_r as the sampling error associated with the sampling of M_r , and ϵ_i as the sampling error associated with the sampling of M_i .

By taking the magnitude of the measurement model above, we obtain

$$|M| = \sqrt{M_r^2 + M_i^2}.$$

Here, $|M|$ follows a Rician distribution with parameters μ_m and σ_m . For our model, we assumed σ_i and σ_r are both equal to σ_m .

The fat-fraction estimate in clinical settings is generated by

$$\hat{\rho}_{clinical} = \frac{|F|}{|F| + |W|}.$$

The magnitudes of the fat and water signals $|F|$ and $|W|$ follow the Rician distribution. At high signal-to-noise ratio (SNR), the signals approximately follows a normal distribution. SNR is the mean to standard deviation ratio and is considered to be high when $\mu/\sigma \geq 3$ [6], as shown in Figure 2.

The clinical estimates sometimes do not accurately estimate the true fat fraction (Figure 3).

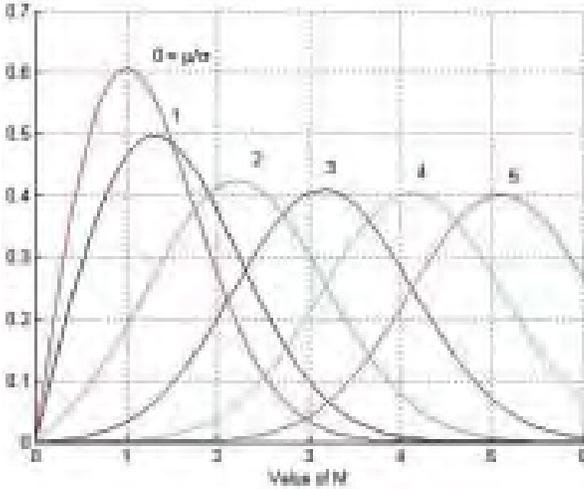


Figure 2 – Rician pdf: $f_M(M) = \left(\frac{M}{\sigma}\right) e^{-(M^2 + \mu^2)/2\sigma^2} I_0\left(\frac{M\mu}{\sigma^2}\right)$ for $M \geq 0$. The Rician distribution is depicted above for several SNRs when $\sigma = 1$. The distribution of M is approximately normal once $\text{SNR} \geq 3$. I_0 is the zeroth order modified Bessel function.

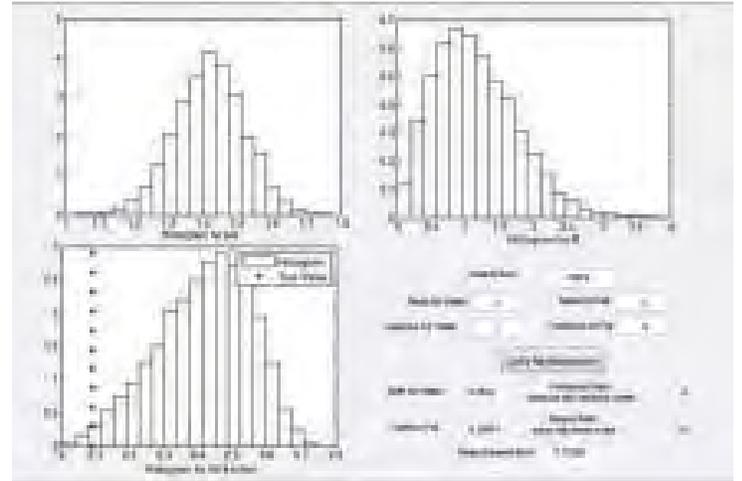


Figure 3 – An image showing one result from the use of the clinical method. This method has provided a poor estimate of the fat fraction when SNR is high for $|W|$ and low for $|F|$. In the histogram of the fat-fraction estimates above we see that the true value is not estimated well. By using the MLE method, we hope to correct this problem. Image courtesy of the authors of [2,3].

Maximum Likelihood Estimation (MLE) Method

Given samples from a statistical model, the method of MLE is used for deriving the estimators of unknown parameters. MLE often yields a reasonable estimator, and if the data set is large, it will be typically become an excellent estimator. The estimated parameter is a value that has the highest probability of generating a given set of data.

Meaning that the estimated parameter is a value at which we would find the maximum of the likelihood function. The likelihood function is a function of the parameters, denoted as $L(\theta|x)$, given observed values from a pdf [3]. Parameters are important in mathematical modeling. In statistics, pdfs are characterized by their parameters. By adjusting the parameters, we can obtain different pdfs. The parameter that we wish to estimate is μ since the fat fraction consists of the mean of magnitudes of the fat and water signals.

Choosing of $\hat{\sigma}_m$

We assumed that the standard deviation of our model will be equivalent to the standard derivation for the real part of the measured signal [2,4] when $\sigma_{m,r} = \sigma_{m,i}$:

$$\sigma_m = \sigma_{m,r}.$$

This means that the standard deviation for the water signal is the standard deviation from the real part of the measured water signal and the standard deviation for the fat signal is the standard deviation from the real part of the measured fat signal,

$$\sigma_w = \sigma_{w,r} \text{ and } \sigma_f = \sigma_{f,r}.$$

These relationships were verified by our simulations for the magnitude of the fat and water signals as shown in Figure 4.

The normal approximation works well at high SNR (as shown in Figure 4) but it does not work as well at low SNR (Figure 5).

Derivation of $\hat{\mu}_m$ by the Maximum Likelihood Estimation (MLE) Method

The MLE of μ is a value or set of values that maximizes the likelihood function. To maximize the likelihood function, we take the derivative of the likelihood function with respect to the parameters that we wish to estimate, and set it to zero. For our research, that parameter is the mean of the normal distribution, μ , and the likelihood function is a normal distribution since it approximates the Rician distribution at high SNR. We solved for the estimate of μ when the derivative is set to zero [4],

$$L(\theta = \mu_m; \sigma_m, m) = \frac{1}{\sqrt{2\pi}\sigma_m} \cdot e^{-\frac{(m - \sqrt{\mu_m^2 + \sigma_m^2})^2}{2\sigma_m^2}} \quad \text{for } m \geq 0.$$

Above is a model of the likelihood function, which is a normal distribution with m restricted to nonnegative values. The derivation of the MLE is provided in Appendix I.

By the method of MLE, we conclude our estimate to be,

$$\hat{\mu}_m = \begin{cases} \sqrt{m^2 - \sigma_m^2} & \text{if } m^2 - \sigma_m^2 > 0. \\ 0 & \text{Otherwise.} \end{cases}$$

The estimate of the fat fraction by MLE is generated by

$$\hat{\rho}_{MLE} = \frac{\hat{F}_{MLE}}{\hat{F}_{MLE} + \hat{W}_{MLE}},$$

where \hat{F}_{MLE} is the estimated fat signal and \hat{W}_{MLE} is the estimated water signal obtained by the MLE method.

Sample Mean Squared Error (MSE)

In both methods, n denotes the size of our measurement sample which is 100,000. This means that we are simulating a patient's liver 100,000 times in an MRI scanner.

The sample MSE will allow us to compare the accuracy of each method's estimation of the fat fraction. The sample MSE of the fat fraction using the clinical method is

$$\widehat{MSE}_{clinical} = \frac{1}{n} \sum_{i=0}^n (\rho_{true} - \hat{\rho}_{clinical_i})^2,$$

The sample MSE of the fat fraction using the MLE method is

$$\widehat{MSE}_{MLE} = \frac{1}{n} \sum_{i=0}^n (\rho_{true} - \hat{\rho}_{MLE_i})^2.$$

Earlier in this paper, we denoted ρ_{true} as the true fat fraction, $\hat{\rho}_{clinical}$ as an estimate of the fat fraction obtained by the clinical method, and $\hat{\rho}_{MLE}$ as an estimate of the fat fraction obtained by the MLE method.

Preliminary Results

Exploring the parameters of μ_w and σ_w and μ_f and σ_f

We tried different parameters to see when \widehat{MSE}_{MLE} will be smaller than $\widehat{MSE}_{clinical}$, under the assumption that the mean of the real and imaginary parts of the fat and water signals were the same.

Case 1) When

$$\begin{aligned} \mu_{w,r} &= \mu_{w,i} = 3, \\ \mu_{f,r} &= \mu_{f,i} = 2, \text{ and} \\ \sigma_{w,r} &= \sigma_{f,r} = 0.01, \end{aligned}$$

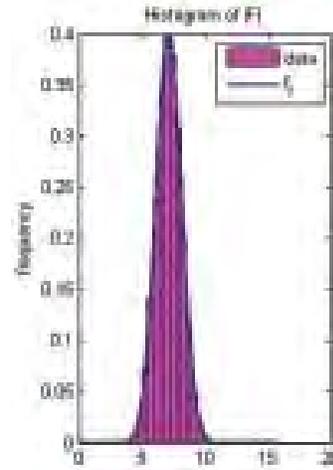


Figure 4 – A histogram of the magnitude fat signal measured at high SNR, $\sqrt{50}$. The pdf, f_f , looks approximately normal with the parameters $\mu_f = \sqrt{50}$ and $\sigma_f = 1$. Clearly the pdf perfectly fits the histogram when $\sigma_f = \sigma_{f,r} = 1$.

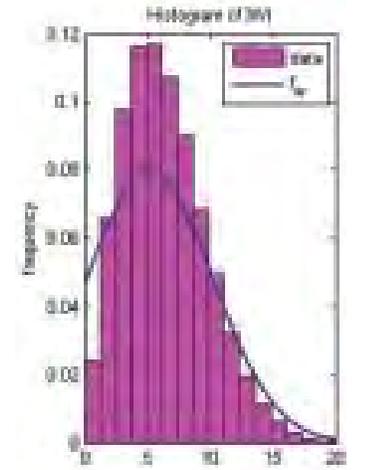


Figure 5 – A histogram of the magnitude water signal measured at low SNR, $\sqrt{\frac{2}{5}}$. The histogram is estimating the pdf of f_w with the parameters $\mu_w = \sqrt{2}$ and $\sigma_w = 5$. This is not a bad fit considering that $\sigma_w = \sigma_{w,r} = 5$.

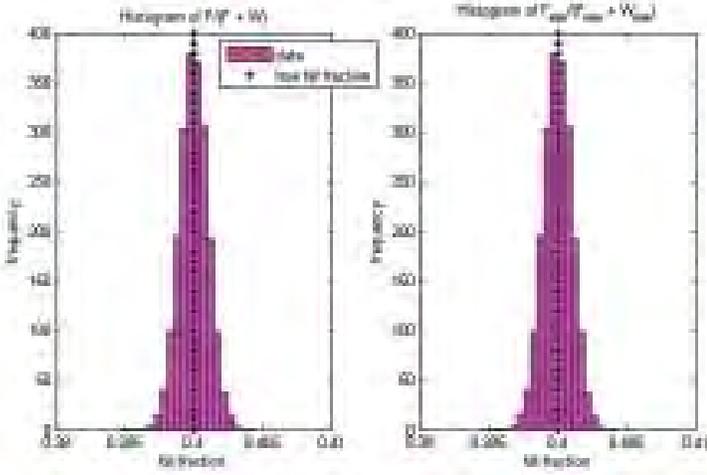


Figure 6 – The figure on the left shows the results of the fat-fraction estimates by the clinical method, and the figure on the right shows the results of the fat-fraction estimates by the MLE method. The black line of asterisks represent the true fat fraction. In Case 1 (with high SNR), both methods estimate the fat fraction with the same accuracy since the variability of the samples is very small.

the SNR is high for water and fat,

$$SNR_w = 100\sqrt{18} \text{ and } SNR_f = 100\sqrt{8}.$$

The results are

$$\rho_{true} = \frac{\sqrt{8}}{\sqrt{8} + \sqrt{18}}.$$

At high SNR, the sample MSE are approximately the same for the clinical and MLE methods,

$$\widehat{MSE}_{clinical} = 1.0412 \times 10^{-6} \text{ and } \widehat{MSE}_{MLE} = 1.0412 \times 10^{-6}.$$

These results are shown in Figure 6.

Case 2) When

$$\begin{aligned} \mu_{w,r} &= \mu_{w,i} = \sqrt{0.5}, \\ \mu_{f,r} &= \mu_{f,i} = \sqrt{0.005}, \text{ and} \\ \sigma_{w,r} &= \sqrt{0.1}, \text{ and } \sigma_{f,r} = \sqrt{0.9}, \end{aligned}$$

the water signal is measured at high SNR while the fat signal is measured at low SNR,

$$SNR_w = \sqrt{10} \text{ and } SNR_f = \frac{\sqrt{10}}{30}.$$

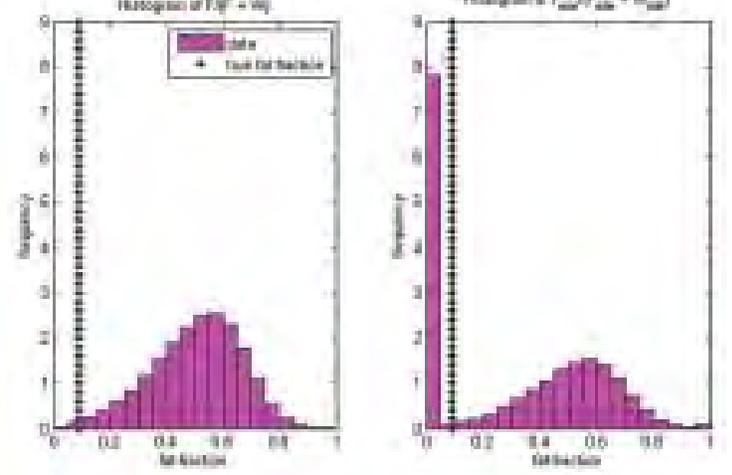


Figure 7 – The figure on the left is the result of clinical method that the CURM team produced, and the figure on the right shows the result of the fat-fraction estimates by the MLE method. The black line of asterisks represent the true fat fraction. For Case 2 (low SNR for the fat signal), the MLE method has a smaller mean squared error than the clinical method.

The results are

$$\rho_{true} = \frac{0.1}{1.1}.$$

The MLE estimates the fat fraction more accurately than the clinical method,

$$\widehat{MSE}_{clinical} = 0.1964 \text{ and } \widehat{MSE}_{MLE} = 0.1333.$$

These results are shown in Figure 7.

Case 3) When

$$\begin{aligned} \mu_{w,r} &= \mu_{w,i} = 1, \\ \mu_{f,r} &= \mu_{f,i} = 5, \text{ and} \\ \sigma_{w,r} &= 5, \text{ and } \sigma_{f,r} = 1, \end{aligned}$$

the water signal is measured at low SNR while the fat signal is measured at high SNR,

$$SNR_w = \frac{\sqrt{2}}{5} \text{ and } SNR_f = \sqrt{50}.$$

The results are

$$\rho_{true} = \frac{\sqrt{50}}{\sqrt{50} + \sqrt{2}}.$$

With these parameters the MLE method yield better results than the clinical method for the fat-fraction estimates compared to Case 2, where

$$\widehat{MSE}_{clinical} = 0.0951 \text{ and } \widehat{MSE}_{MLE} = 0.0704.$$

The results can be seen in Figure 8.

Discussion

The parameters for the water and fat signals were chosen so that the sample MSE of the MLE method will be smaller than the sample MSE of the clinical method. In addition, we assumed that the mean and the standard deviation of the real and imaginary components of the signals were equal to each other. We have made this assumption in order to understand how the MLE method works in estimating the fat fraction. If the parameters of the water and fat signals were inverses of each other as in Case 3 (low SNR for the water signal), we would achieve a smaller sample MSE of MLE than the clinical method.

From our preliminary results, we would expect the MLE method to provide a better estimate of the fat fraction than the clinical method since the MLE method uses statistical methods to generate an estimate of the fat fraction. The clinical method takes the magnitude of the negative

observations of the measured fat and water signals, so the method generates a positive estimate for the amount of fat and water in the liver. While the MLE method replaces the negative observations with the value zero. This leads to the large peaks in the histograms for the MLE method. When the mean of the fat signal is small, there is a peak at a fat-fraction estimate of zero. When the mean of the water signal is small, there is a peak at the fat-fraction estimate of one.

Consequently, the clinical method would either overestimate or underestimate the true fat fraction. This means the clinical method assumes there is either more or less fat in the liver, leading to a misdiagnosis of NAFLD. Results from the MLE method generated a more accurate estimator of the fat fraction than the clinical method, but it is still far from the true fat fraction.

Future Work

In the future, we will work towards finding an estimate for the fat fraction through the MLE method, but when the fat and water magnitude signals have a low signal-to-noise ratio. In addition, we will need to address how reliable the estimates from the MLE method is by comparing it to real samples.

Acknowledgments

I would like to thank the Minority Access to Research Careers (MARC) program for financial support and Peter T. Ho for his suggestions. Additionally, I would like to especially thank my advisor, Dr. Angel R. Pineda. This project would not be possible without his immense support and gentle guidance.

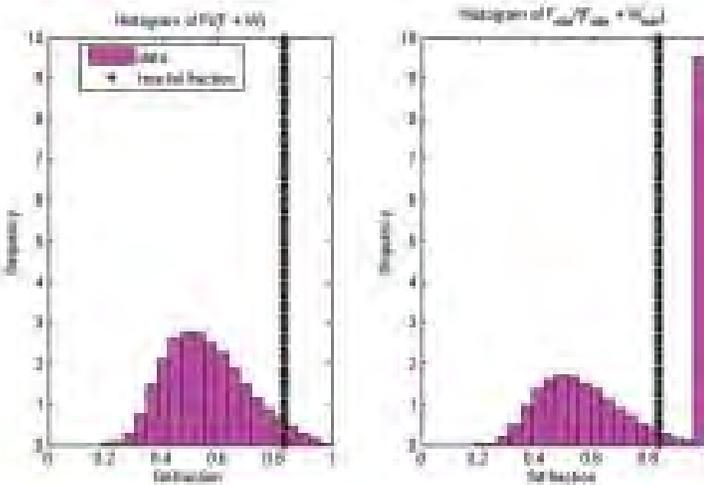


Figure 8 – The figure on the left shows the results of the fat-fraction estimates by the clinical method and the figure on the right shows the results of the fat-fraction estimates by the MLE method. The black line of asterisks represent the true fat fraction. In Case 3 (low SNR for the water signal), MLE method has a slightly smaller mean squared error than the clinical method.

Appendix I (Derivation of $\hat{\mu}_m$ by MLE)

Below is the model of the likelihood function used for both the fat and water signals:

$$L(\theta = \mu_m; \sigma_m, m) = \frac{1}{\sqrt{2\pi}\sigma_m} \cdot e^{-\left(m - \sqrt{\mu_m^2 + \sigma_m^2}\right)^2 / 2\sigma_m^2} \quad \text{for } m \geq 0. \quad [6]$$

Since most distributions are exponential, it is often easier to obtain the MLE estimate by taking the natural log first and then the derivative. Natural log is a monotonic transformation and retains the same maximum as the likelihood function. The natural log was taken from both sides,

$$\ln L(\theta) = \ln \left(\frac{1}{\sqrt{2\pi}\sigma_m} \cdot e^{-\left(m - \sqrt{\mu_m^2 + \sigma_m^2}\right)^2 / 2\sigma_m^2} \right) = - \left(m - \sqrt{\mu_m^2 + \sigma_m^2} \right)^2 / 2\sigma_m^2 + \ln \left(\frac{1}{\sqrt{2\pi}\sigma_m} \right).$$

For this derivation, we assumed that the variance is known. The derivative was taken with respect to μ_m since the fat fraction depends on the means of the fat and water signals. Then the derivative is set to zero in order to find a value of μ_m that would maximize the likelihood function.

$$\begin{aligned} \frac{d}{d\mu_m} \ln L(\theta) &= \frac{d}{d\mu_m} \left[- \left(m - \sqrt{\mu_m^2 + \sigma_m^2} \right)^2 / 2\sigma_m^2 + \ln \left(\frac{1}{\sqrt{2\pi}\sigma_m} \right) \right] = 0. \\ \frac{d}{d\mu_m} \left[- \left(m - \sqrt{\mu_m^2 + \sigma_m^2} \right)^2 / 2\sigma_m^2 \right] &+ \frac{d}{d\mu_m} \left[\ln \left(\frac{1}{\sqrt{2\pi}\sigma_m} \right) \right] = 0. \end{aligned}$$

By using quotient rule, chain rule, and the fact that $\frac{1}{\sqrt{2\pi}\sigma_m}$ is a constant, we obtain

$$\left[\frac{-2 \left(m - \sqrt{\mu_m^2 + \sigma_m^2} \right)}{2\sigma_m^2} \right] \cdot \left[\frac{-2\mu_m}{2\sqrt{\mu_m^2 + \sigma_m^2}} \right] + 0 = 0.$$

From above we have,

$$\frac{\mu_m \left(m - \sqrt{\mu_m^2 + \sigma_m^2} \right)}{\sigma_m^2 \left(\sqrt{\mu_m^2 + \sigma_m^2} \right)} = 0.$$

Since $\sigma_m^2 \neq 0$, we multiply both sides by σ_m^2 and separate the equation into two fractions,

$$\frac{\mu_m m}{\sqrt{\mu_m^2 + \sigma_m^2}} - \frac{\mu_m \sqrt{\mu_m^2 + \sigma_m^2}}{\sqrt{\mu_m^2 + \sigma_m^2}} = 0.$$

Cancel the square root terms and add μ_m to both sides,

$$\frac{\mu_m m}{\sqrt{\mu_m^2 + \sigma_m^2}} = \mu_m.$$

Divide each side by μ_m ,

$$\frac{m}{\sqrt{\mu_m^2 + \sigma_m^2}} = 1.$$

Square the left and right side,

$$m^2 = \mu_m^2 + \sigma_m^2.$$

Isolate μ_m^2 ,

$$\mu_m^2 = m^2 - \sigma_m^2.$$

Square root both sides,

$$\mu_m = \pm \sqrt{m^2 - \sigma_m^2}.$$

Since m was restricted to nonnegative values, μ_m cannot be negative.

By the method of MLE, we conclude

$$\hat{\mu}_m = \begin{cases} \sqrt{m^2 - \sigma_m^2} & \text{if } m^2 - \sigma_m^2 > 0 \\ 0 & \text{O.W.} \end{cases}$$

References

Leon A. Adams, Paul Angulo, and Keith D. Lindor. *Nonalcoholic Fatty Liver Disease*. CMAJ: Canadian Medical Association Journal, 172.7 (2005): 899-905.

Anne M. Calder, Eden A. Ellis, Li-Hsuan Huang, and Kevin Park. *Statistical Modeling through Analytical and Monte Carlo Methods of the Fat Fraction in Magnetic Resonance Imaging (MRI)*. SIAM Undergraduate Research Online (under review).

Anne M. Calder, Eden A. Ellis, Li-Hsuan Huang, and Kevin Park. *Statistical Modeling through Analytical and Monte Carlo Methods of the Fat Fraction in Magnetic Resonance Imaging (MRI)*. DIMENSIONS The Journal of Undergraduate Research in Natural Sciences and Mathematics. 13 (2011):114-123.

Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*. 4th ed. Addison-Wesley (2002).

J.K. Dowman, J.W. Tomlinson, and P.N. Newsome. *Systematic Review: The Diagnosis and Staging of Non-alcoholic Fatty Liver Disease and Non-alcoholic Steatohepatitis*. Alimentary Pharmacology & Therapeutics, 33.5 (2011): 525-540.

Há kon Gudbjartsson and Samuel Patz. *The Rician Distribution of Noisy MRI Data*. Magnetic Resonance in Medicine. 34 (1995): 910-914.

Scott B. Reeder, Philip M. Robson, Huanzhou Yu, Ann Shimakawa, Catherine DG Hines, Charles A. McKenzie, Jean H. Brittain. (2009) *Quantification of Hepatic Steatosis With MRI: The Effects of Accurate Fat Spectral Modeling*. Magn Reson Imaging. 29 (2009): 1332-1339.

Scott B. Reeder, Claude B. Sirlin. *Quantification of Liver Fat with Magnetic Resonance Imaging*. Magnetic Resonance Imaging Clinics of North America. (2010): 337-357.

Nina Schwenzer, Fabian Springer, Christina Schraml, Norbert Stefan, Jrgen Machann, and Fritz Schick. *Non-invasive Assessment and Quantification of Liver Steatosis by Ultrasound, Computed Tomography and Magnetic Resonance*. Journal of Hepatology, 51.3 (2009): 433-445.

M.K. Oh, J. Winn, and F. Poordad. Review Article: *Diagnosis and Treatment of Non-alcoholic Fatty Liver Disease*. Alimentary Pharmacology & Therapeutics, 28.5 (2008): 503-522.

Assessing Uncertainty of Clustered Neuronal Intensity Curves

Department of Mathematics, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

Mikhail Y. Popov

Advisor: Dr. Sam Behseta

Abstract

In this research, we utilize Bayesian model-based clustering with Gaussian mixtures obtained under two experimental conditions. We devise a procedure for generating confidence intervals for the mean curve of each cluster via permuting spiking data followed by implementing Bayesian curve fitting method of BARS. Throughout, we employ a measure of divergence called Total Variation Distance. Our analysis shows that in general, the proposed neuronal classification technology is far more precise when firing patterns demonstrate a trough in intensity. In contrast, the classification method yields less precision at the peak of neuronal spiking activity.

Key words: neuronal intensity functions, Bayesian curve-fitting, BARS, clustering.

Introduction

In this work we consider the problem of clustering a collection of neuronal intensity functions represented as curves. Particularly, having a sufficient number of curves, we focus on obtaining mean curves for each cluster by pointwise averaging over the curves belonging to the specific cluster. By permuting the original spike train data and refitting curves to the permuted data, we devise an algorithm which allows us to create a 95% Confidence Interval for each cluster's mean curve. This provides us with valuable information in regards to patterns and temporal location of uncertainty of neuronal firing intensity functions.

In particular, we are motivated by the neuronal data generated during a study of a group of motor cortex neurons of a Macaque monkey reaching a task (Matsuzaka et al., 2007). The monkey was trained to press a target

within 800 ms of appearance on a touch-sensitive screen. Illuminating targets were placed in a horizontal orientation on the screen. The study was performed under two experimental modes: a random mode in which the targets were illuminated in a pseudo-random order and a repeating mode in which the targets illuminated in a predetermined sequence.

Theoretically, neurons communicate through releasing electric pulses also called neuronal firings. Prior to the actual firings, there is a build up of chemicals called an action potential and once the levels of the chemicals surpass a certain threshold firing occurs. An electrode is used for sensing the firings and they are subsequently recorded as spikes (Gazzaniga et al., 1998). A sequence of these firings is referred to as a spike train. We can model the individual spikes at each millisecond time bin with a Bernoulli distribution and consequently it is possible to capture the stochastic features of the sum of a group of spiking events in a homogeneous Poisson process.

We utilize the properties of a graphical tool called Peristimulus Time Histograms (PSTH) to capture the patterns of neuronal firing intensity rates. To create PSTH we bin the experimental time window, followed by creating a histogram whose bars reflect the count of spiking occurrences in widths of 10 ms. We can adjust the PSTH so that it gives us the rates of spikes for every 10 ms bin of time across all the trials. This standardization technique yields neuronal firing intensity rates.

Histograms are prone to random fluctuations and thus one may benefit from smoothing them. In order to fit a continuous curve to the neuronal intensity rates, we employ a technique called Bayesian Adaptive Regression Splines or BARS (DiMatteo et al., 2001). This fully Bayesian process uses Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995) to

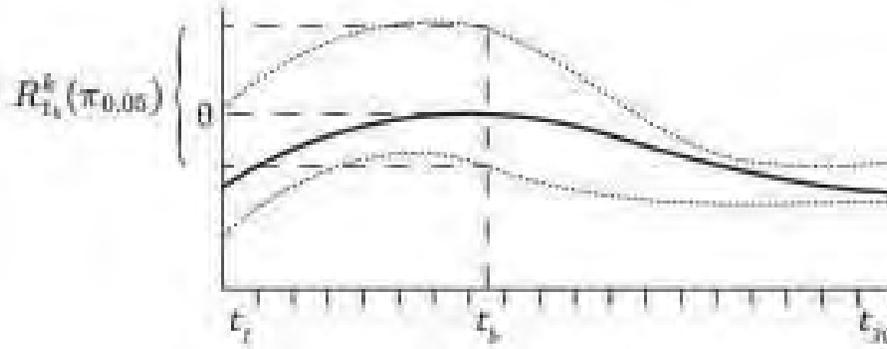


Figure 1 – Dotted curves are the upper and lower bounds of the confidence interval for $\hat{\mu}_*^k$, represented by a thick, solid curve. At each time bin t_b , the mean is θ and the bounds are formed by the region $R_{t_b}^k(\pi_{0.05})$. The sup and the inf of $R_{t_b}^k(\pi_{0.05})$ are marked by dashed horizontal lines above and below θ , respectively.

sample from a posterior distribution of curves to approximate placement of splines to smooth a discrete data set. A more rigorous description of curve-fitting process is provided in Section 2.1.

Once all PSTH are modeled with BARS, we tackle the problem of identifying common features of neuronal intensity rates through a method of classification called model-based clustering (Fraley and Raftery, 2002). Briefly, in model-based clustering, a weighted average (mixture) of multivariate normal distributions with varying vector means and covariance matrices is formed to allocate curves sharing common characteristics to the same cluster. The weights of the mixture are obtained through a Bayesian mechanism.

More precisely, model-based clustering categorizes multivariate data based on forming a mixture of Gaussian distributions whose proportions and parameters are to be determined using a Bayesian methodology. The Gaussian mixture distribution for discretized curve \vec{x} can be written in the form

$$p(\vec{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\vec{x} | \mu_k, \Sigma_k),$$

where the weights $\{\pi_1, \dots, \pi_K\}$ satisfy

$$0 \leq \pi_k \leq 1, \text{ and } \sum_{k=1}^K \pi_k = 1.$$

That is, we have K many Normal distributions (\mathcal{N}) with different parameters

and the model for the distribution of the data is a weighted, linear combination of those Normal distributions. We apply this particular clustering method with the curves fitted to the original data as well as curves fitted to the permuted data, as we shall describe in later sections. The main contribution of this work lies in assessing the uncertainty of the mean of curves within each cluster. We shall explain that in details in the following section.

Methods

Curve Fitting

We begin by fitting curves to the neuronal firing histograms. This process is done via the following:

1. Let T_m be the set of firing times for the m -th trial where $m = 1, \dots, M$ and M is the total number of trials for the neuron: $T_m = \{t : -200 \leq t \leq 100, t \text{ is a time (ms) during which a neuron spike occurred}\}$.
2. Pool the trials together and then for each time bin t_b ($b = 1, \dots, 30$) count how many spikes occurred in that particular time bin. Let $c = \{c_1, \dots, c_{30}\}$, where c_b is the count of all spikes in t_b .

Subsequently, this vector of all spike counts in the experimental window is our PSTH. Standardize c by dividing each count by the number of trials M to obtain the firing intensity rates. This ensures that a neuron with more trials recorded than another neuron will not have an advantage of counts.

Denote the set of rates as

$$F = \left\{ \frac{c_1}{M}, \dots, \frac{c_{30}}{M} \right\}.$$

	Unsupervised					Supervised		
Random	C 1	2	3	4		C 1	2	3
	N	17.27	11.33	24.44	17.98	N	17.27	11.33
	C 5	6	7			C 5	6	7
	N	6.68	9.33	12.58		N	24.44	9.33
		(a)					(b)	
Repeating	C 1	2	3	4		C 1	2	3
	N	18.71	7.33	12.89	13.62	N	26.9	7.33
	C 5	6	7			C 5	6	7
	N	15.68	9.33	7.19		N	15.68	15.68
		(c)					(d)	

Table 1 – The proportion of curves (out of 213 total) associated with cluster C under each combination of experimental mode (random, repeating) and clustering method (unsupervised, supervised).

3. Fit BARS to F , yielding a curve that may be discretized and represented as a vector of size 30. We devise the following notation for expressing this vector for neuron j :

$$\hat{u}^j = (f^j(t_1), \dots, f^j(t_{30})),$$

where f is the smooth function obtained from BARS.

The above process generates the curves $\hat{u}^1, \dots, \hat{u}^N$ since we have N many neurons.

Mean Curves and Permutations

In the second phase, we cluster the curves $\hat{u}^1, \dots, \hat{u}^N$ obtained from the original data. We then use model-based clustering on the set of those curves to create K many clusters. That is, we get a Gaussian mixture distribution with K components where K is chosen through the application of model-based clustering algorithm. This will create a partition since no curve can belong to more than one cluster and the union of all clusters is the original set of curves we started with.

We represent the mean of each cluster by averaging pointwise across all the curves belonging to that particular cluster. Let I^k be the the set containing the indices of neurons belonging to cluster k . Mathematically, we represent the mean curve as the vector

$$\tilde{\mu}^k = \left(\frac{\sum_{j \in I^k} f^j(t_1)}{n^k}, \dots, \frac{\sum_{j \in I^k} f^j(t_{30})}{n^k} \right),$$

where n^k is the number of curves there are in the k -th cluster. In other words, n^k is the cardinality of I^k . This infrastructure will facilitate the application

of an iterative procedure of data-permutation, clustering, and per-cluster-mean calculation, as explained in the following steps.

1. For the i -th permutation:

(a) For neuron j :

i. Consider the set T^j of M trials,

$$T^j = \{T_1^j, \dots, T_M^j\},$$

where each trial T_m^j is a set of firing times as described in the previous section.

ii. Sample M many trials from T^j with replacement (thus allowing for repetitions of trials) to make $T_{P_i}^j$.

iii. Put TPJ through PSTH to get the time bin spike counts c (as described in 2.1).

iv. To obtain the permutation's firing intensity rates $F_{P_i}^j$, divide the counts c_1, \dots, c_{30} by M .

v. Finally, fitting BARS to $F_{P_i}^j$ gives the curve $\tilde{\mu}_{P_i}^j$.

(b) The above procedure is performed for every neuron to yield $\hat{u}_{P_i}^1, \dots, \hat{u}_{P_i}^N$.

(c) Get K_{P_i} groups of curves by using model-based clustering on $\hat{u}_{P_i}^1, \dots, \hat{u}_{P_i}^N$.

(d) Do pointwise averaging again to obtain the mean curve of each cluster within this iteration:

$$\left\{ \tilde{\mu}_{P_i}^1, \dots, \tilde{\mu}_{P_i}^{K_{P_i}} \right\}.$$

2. Step 1 is performed B many times with B being a large number (preferably).

Assessing Similarity Between Curves

The means of clusters of original data and those obtained from permutations are of primary interest in this work. As justified later, we would like to identify those means of permuted curves whose pattern of activity are as close as possible to the one associated with the original data. To gauge “closeness” we adopt a measure of similarity called Total Variation Distance (TVD) (Tsybakov, 2009). In previous work, we have demonstrated the superiority of TVD (via Mean Squared Error analysis) as a measure of divergence of two curves in comparison to a number of other metrics (Popov et al., 2011).

Definition. Let (χ, A) be a measurable space and let P and Q be two probability measures on (χ, A) . Suppose that v is a σ -finite measure on (χ, A) satisfying $P \ll v$ and $Q \ll v$. Define $p = dP/dv, q = dQ/dv$. That is, p and q are probability density functions. Define Total Variation V Distance between P and Q as

$$V(P, Q) = \sup_{\alpha \in A} \left| \int_{\alpha} dP - dQ \right|.$$

In the context of neuronal intensity curves, let \hat{r}^1 and \hat{r}^2 be two probability mass functions. We can calculate their divergence using the algorithmic translation of TVD via

$$V(\hat{r}^1, \hat{r}^2) = \max \left\{ |\hat{r}_1^1 - \hat{r}_1^2|, \dots, |\hat{r}_{30}^1 - \hat{r}_{30}^2| \right\}.$$

Note that V will be 0 when the two probability functions are identical and 1 when they are completely different. Due to the fact that TVD is structured around probability densities, in order for us to be able to employ it we must convert curves to probabilities. This is achieved by converting the curve

$$\hat{u} = (f(t_1), \dots, f(t_{30})),$$

to its corresponding probability mass function \hat{r} via:

$$\hat{r} = \left(\frac{f(t_1)}{|\hat{u}|}, \dots, \frac{f(t_{30})}{|\hat{u}|} \right), \text{ where } |\hat{u}| = \sum_{y=1}^{30} f(t_y).$$

Recall that we have permutations $P_1, P_2, \dots, P_i, \dots, P_B$ that each i -th iteration has K_{P_i} many clusters of curves, and that each of those clusters has a mean curve. The purpose of the following algorithm is to select one of those mean curves (from each permutation) that are highly similar to $\tilde{\mu}^k$. Let us start with the first cluster $k = 1$ and its mean curve $\tilde{\mu}^1$.

1. (a) For P_1 having L clusters ($L = K_{P_1}$), obtain distances $v_{P_1}^1, \dots, v_{P_1}^L$ for that iteration's clusters by comparing the mean curves $\tilde{\mu}_{P_1}^1, \dots, \tilde{\mu}_{P_1}^L$ with

the mean curve $\tilde{\mu}^k$ using the metric V .

- (b) Ignore those mean curves whose measures would be greater than the threshold v^* . Consider the threshold 0.2 as a mark of high similarity.
- (c) Within the group of curves with TVD less than v^* , pick the the mean curve that resulted in the smallest TVD during step 1a. Therefore the mean curve with the highest similarity to $\tilde{\mu}^k$ is selected.

2. Repeat step 1 for each of the remaining permutations P_2, \dots, P_B to obtain ω^1 , the set of mean curves which are very similar to the curve $\tilde{\mu}^1$. This process is repeated for each of the remaining main clusters $2, \dots, K$ to yield $\omega^1, \dots, \omega^K$. Therefore each ω^k is a set of mean curves which are very similar to main clusters' mean curves $\tilde{\mu}^1, \dots, \tilde{\mu}^K$, respectively.

Confidence Intervals

The $\tilde{\mu}^1, \dots, \tilde{\mu}^K$ are no longer needed as the formulation of the confidence interval devised in this work solely relies on the $\omega^1, \dots, \omega^K$. We designate $\tilde{\mu}_*^k$ as the pointwise average of the mean curves found in ω^k using the same technique as before.

Definition. Let D denote data and θ be a one-dimensional parameter. Let $L(\theta|D)$ be a likelihood function given D and $\pi(\theta)$ be a prior. Then the Bayesian posterior density has the form $\pi(\theta|D) \propto L(\theta|D)\pi(\theta)$. A $100(1 - \alpha)\%$ Highest Posterior Density interval (HPDI) for θ is the region

$$R(\pi_\alpha) = \{\theta : \pi(\theta|D) \geq \pi_\alpha\},$$

where π_α is the largest constant such that $P(\theta \in R(\pi_\alpha)) \geq 1 - \alpha$. $R(\pi_\alpha)$ will be the smallest interval width among all credible intervals (Chen et al., 2000).

We construct a 95% Confidence Interval around $\tilde{\mu}_*^k$ by operating binwise on $\tilde{\mu}_*^k$. In other words, if there are many curves in ω^k then we regard each time bin as that data vector (of size n). We use HPDI with $\alpha = 0.05$ to find the upper and lower bounds around $\tilde{\mu}_*^k$ to create our confidence interval. Essentially, we let $\alpha = 0.05$ and θ be the average of all curves at time bin t_b (for $b = 1, \dots, 30$). Figure 1 demonstrates how the confidence interval for the mean curve is constructed from $R_{t_1}^k(\pi_{0.05}), \dots, R_{t_{30}}^k(\pi_{0.05})$. In R, this is achieved through using the HPDIinterval function within the lme4 package (Bates et al., 2011).

Data Analysis and Results

In the application of this procedure, we worked with 213 neurons recorded

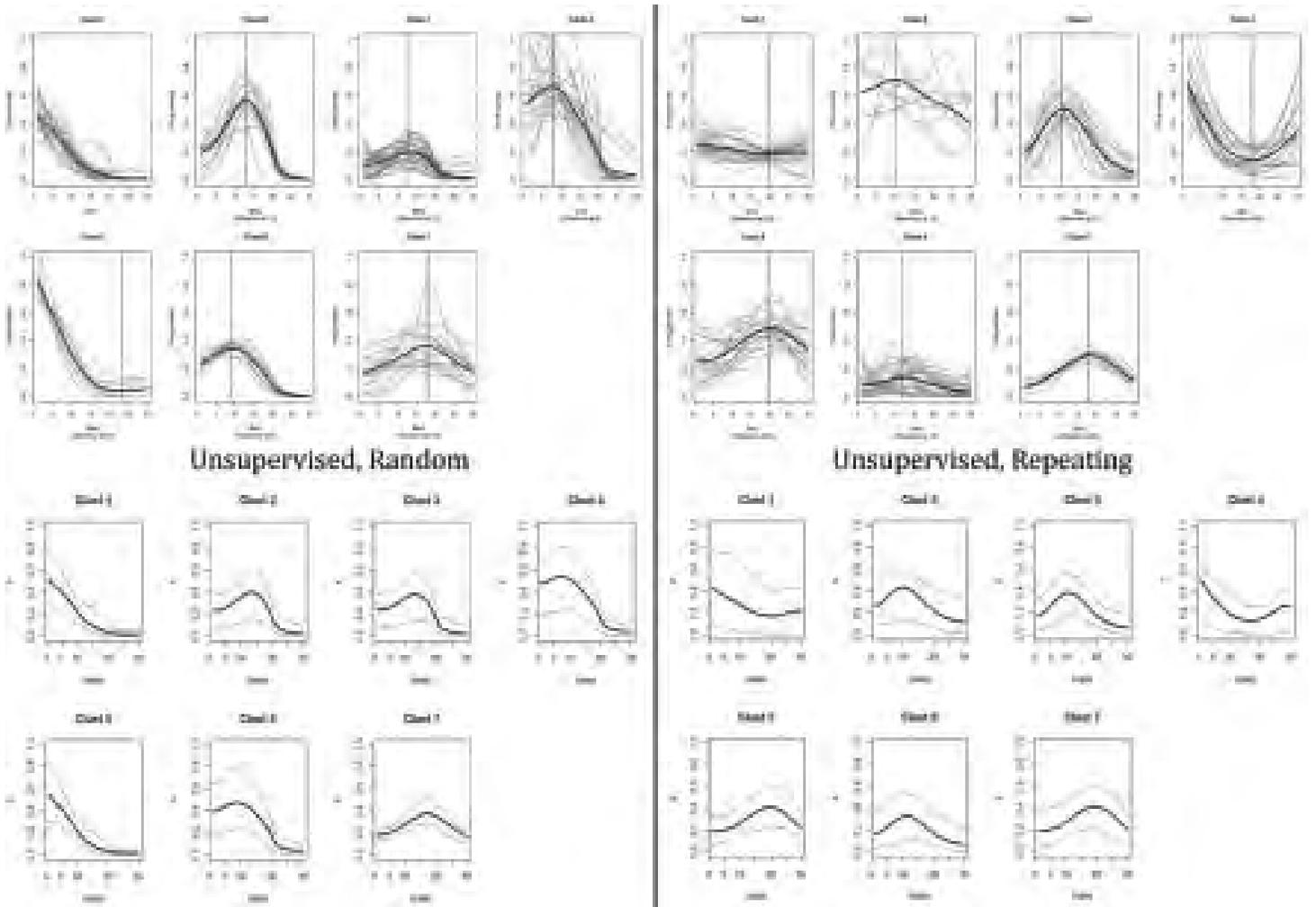


Table 2 – The optimal number of clusters ($K = 7$) was decided by *mclust*'s algorithm. The top half plots the original curves belonging to each cluster. The bottom half shows each cluster's 95% Confidence Interval as derived from the algorithm described earlier.

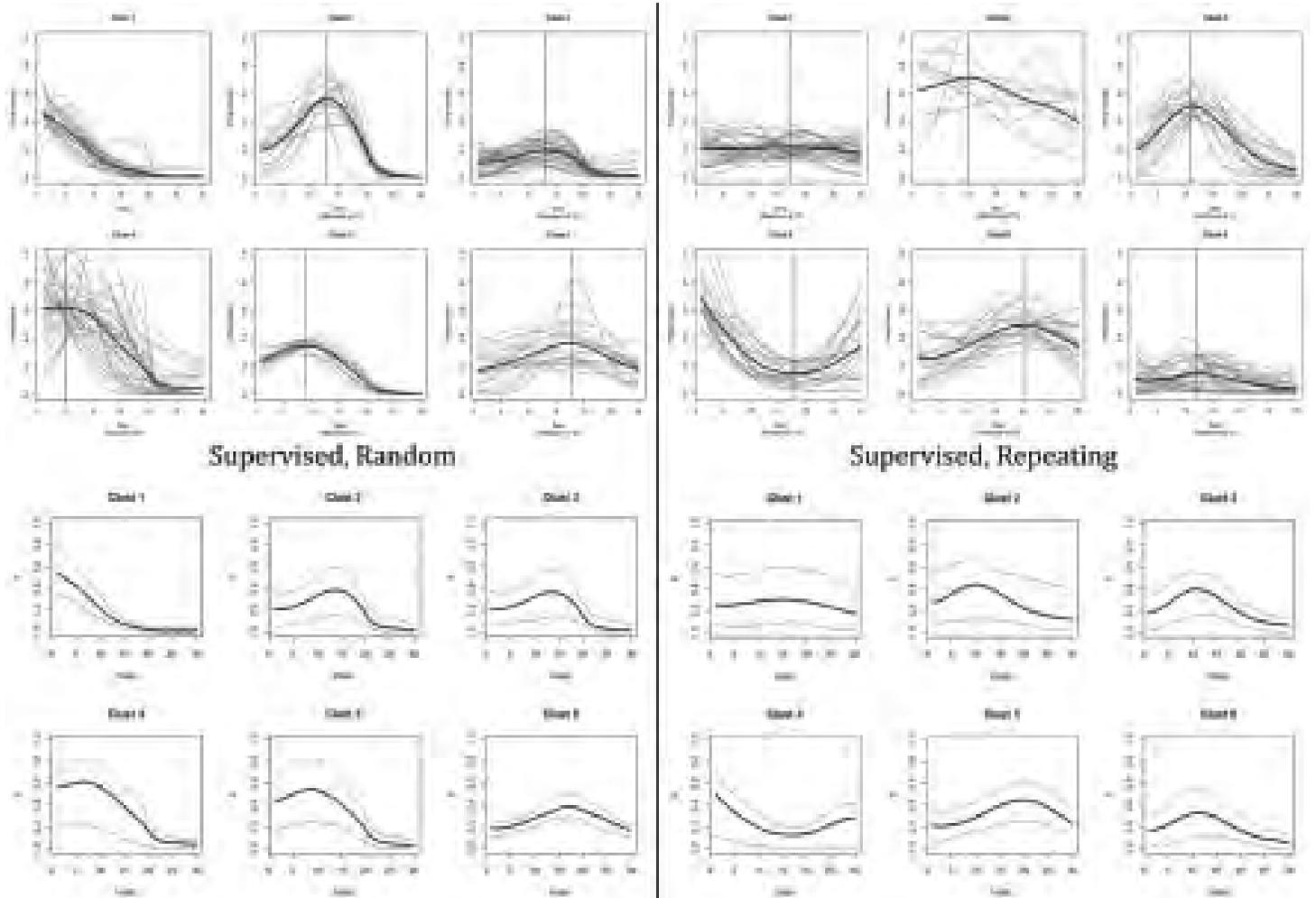


Table 3 – The optimal number of clusters ($K=6$) was set by us after seeing that $K=2, 3, 4, 5$ did not provide enough clusters for the curves. The top half plots the original curves belonging to each cluster. The bottom half shows each cluster's 95% Confidence Interval as derived from the algorithm described earlier.

under the random and repeating experimental modes. We used the permutation algorithm described in Section 2.2 to permute the spike train data, generating 1000 permutations for each of the 213 neurons for each mode.

To cluster the curves created from those neurons, we employed `mclust`, a clustering package in R (Fraley and Raftery, 2006). Two types of clustering, *supervised* and *unsupervised*, were performed on both original and permuted data. Under supervised clustering, we set the number of clusters K to be 6. We chose this number empirically, after considering the clusterings resulting from $K = 2, \dots, 6$. For each i -th permutation we then set the number of clusters, K_{P_i} , to be 6.

Under unsupervised clustering, the number of clusters K was determined by `mclust`'s internal optimization algorithm and, quite interestingly, the algorithm chose K to be 7 under both random and repeating modes. Furthermore the same optimization algorithm was applied in order to decide K_{P_i} , the number of clusters for the i -th permutation.

Table 1 should be viewed in conjunction with Figures 2 and 3 to identify clusters accounting for large and small proportion of curves. For example, from Table 1 (a), we observe that clusters 1, 3, and 4 contain majority of the curves in the random mode. It should be noted that there is no reason to believe that clusters labeled with the same number under each combination of Table 1 are identical.

Figures 2 and 3 are represented in four quadrants. In the top two quadrants we plot the curves of each cluster. The vertical line represents the location of maxima or minima of the mean of the cluster. The bottom two quadrants shows the confidence intervals generated using our procedure.

Regardless of experimental mode or clustering choice, there was always one cluster which collected miscellaneous curves unfit for other clusters. It is interesting to see that the intervals widen when firing intensity increases and then tighten when firing intensity decreases. This results in more uncertainty at the peaks and less uncertainty at the troughs.

The clustering mechanism did not just consider the behavior of the curves (i.e. where the rises and falls are) as expected but also took into account the intensity rate. This resulted in pairs of clusters (in both the random mode and the random mode) which behaved similarly but were treated as different clusters due to differing firing intensity rates.

Random Mode

Under supervised clustering, the mean curves for clusters 1 and 4 are distinguished by being monotonically decreasing whereas there is a rise and fall (a peak) in firing intensity in the remaining four clusters. Cluster 4 is clearly a set of very different curves which did not fit in any of the other five clusters and this greatly affects the uncertainty as evident by the very wide confidence interval for bins 1-20. Oddly, cluster 6 has the tightest confidence interval across the entire 300 ms timeline, especially at the peak.

Under unsupervised clustering, the addition of a seventh cluster allowed for some curves in cluster 4 to be separated into a group of their own. The new cluster 5 is very similar to cluster 1 but starts with a much greater intensity rate. This causes the mean curve of cluster 4 to gain a hill. With regards to mean curves and uncertainty, cluster 7 (which is equivalent to cluster 6 from supervised clustering) retains the feature of having the tightest interval overall. We also see three pairs of very similar confidence intervals form in clusters: 1 & 5, 2 & 3, and 4 & 6.

Repeating Mode

There are several observations to be made for the repeating mode. The uncertainty is especially large for all clusters in comparison to the uncertainty found in random mode, as evidenced by the wider confidence intervals. Cluster 4 exhibits a behavior not found in the random mode curves at all; 15.83% (Table 1c) of curves conform to this pattern.

Under supervised clustering, we see two interesting patterns emerge. Clusters 3 and 5 both have a rise and fall of firing intensity (a peak) but in cluster 3 the hill peaks at time bin 11 whereas in cluster 5 the hill peaks at time bin 20. Cluster 2 has curves which don't behave similarly at all yet have rather large intensity rate, on average. In fact, the same could be said of clusters 1 and 6 except the average intensity rates differ between two three clusters. Unlike the random mode, the clustering mechanism had a lot more trouble grouping random-mode curves by their behavior.

Under unsupervised clustering, the addition of a seventh cluster allowed for some curves in cluster 1 to be separated into a cluster of their own. The new mean curve in cluster 1 is less flat and the curves forming cluster 7 have a hill which peaks near time bin 19, much like the curves found in cluster 5. Clusters 5 and 7 may behave similarly but the greater firing intensity rate in cluster 5 prove it to be a differentiating factor when clustering curves.

Discussion

One shortfall of this work is that it limits its attention to mixtures of Gaussians only. Considering that neuronal spike trains are Poisson processes in their nature, a possible alternative would be to assume unknown distributions rather than Gaussians and then utilize Bayesian methodology for obtaining proper posterior distributions with the appropriate parameters.

An extension of this research would use other methods of classification. For example, methods associated with machine learning (hierarchical clustering, principal component analysis, etc.) could be used in conjunction with the rest of the procedure.

The entire devised procedure in this paper is computationally expensive. Even though powerful computational devices were used it took several days to complete. This is solely due to using BARS to fit a curve to each of the 213 neurons in each of the 1000 permutations under both experimental modes. We used an older, private implementation of BARS rather than a recent, publically available version which may have been slightly faster (see Wallstrom and Kass, 2006).

References

Bates, D., Maechler, M., and Bolker, B. lme4: Linear mixed-effects models using Eigen and Eigen++, 2011. URL <http://CRAN.R-project.org/package=lme4>. R package version 0.999375-42.

Behseta, S. and Chenouri, S. Comparison of two populations of curves with an application in neuronal data analysis. *Statistics in Medicine*, 30(12):1441–1454, 2011.

Bishop, C. *Pattern recognition and machine learning*. Springer, New York, 2006.

Chen, M., Shao, Q., and Ibrahim, J. *Monte Carlo methods in Bayesian computation*. Springer, 2000.

DiMatteo, I., Genovese, C., and Kass, R. Bayesian curve-fitting with free-knot splines. *Biometrika*, 88(4):1055, 2001.

Fraley, C. and Raftery, A. Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association*, 97:611–631, 2002.

Fraley, C. and Raftery, A. *MCLUST Version 3 for R: Normal Mixture Modeling and Model-based Clustering*, 2006. (revised in 2009).

Gazzaniga, M., Ivry, R., and Mangun, G. *Cognitive neuroscience: the biology of the mind*. W.W. Norton, 1998.

Green, P. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

Kass, R., Ventura, V., and Brown, E. Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology*, 94(1):8, 2005.

Matsuzaka, Y., Picard, N., and Strick, P. Skill representation in the primary motor cortex after long-term practice. *Journal of neurophysiology*, 97(2):1819–1832, 2007.

Paninski, L., Brown, E., Iyengar, S., and Kass, R. Statistical models of spike trains. *Stochastic methods in neuroscience*, pages 278–303, 2009.

Popov, M., Rock, D., and Ying, V. A comparative study of divergence metrics in the analysis of neuronal data. *Dimensions*, 13:124–137, 2011.

Puente, S. A re-sampling approach for the comparison of two random curves with an application in neuronal data analysis. *Dimensions*, 12:80–86, 2010.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL <http://www.R-project.org/>.

Tsybakov, A. *Introduction to nonparametric estimation*. Springer series in statistics. Springer, 2009.

Wallstrom, D. and Kass, R. Bayesian Adaptive Regression Splines (BARS), 2006. URL <http://www.stat.cmu.edu/kass/bars/bars.html>. R package version 1.0-1.

An Inverse-Based Analogue of the Probability That Two Elements in a Finite Group Commute

Department of Mathematics, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

Nick Blackford, Daniel Lenders, Danny Orton
Advisor: Dr. Scott Annin

Abstract

Given a finite group G , the probability that two randomly chosen elements of G commute has long been viewed in the literature as a natural measure of the degree of commutativity enjoyed by the group ([3],[8],[5]). Many variants on this probabilistic question have arisen in the literature recently, and our research introduces yet another such variation that provides somewhat different information. In particular, given a product of elements of G , we investigate the likelihood of being able to permute the order of the elements in the product and obtain the inverse of the original element. With the help of the software program Groups, Algorithms and Programming (GAP)[4], we have discovered patterns leading to interesting results about this variant. In our paper, we will examine these patterns for such familiar finite groups as cyclic, dihedral, and symmetric groups, and describe the general results we have obtained. Some of these results are at odds with analogous ones known for the commutativity measure described in ([3],[8],[5]), thereby adding further interest into our investigation.

Introduction

In April of 2011, an article was published in Mathematics Magazine entitled “Two Generalizations of the 5/8 Bound on Commutativity in Nonabelian Finite Groups” [7]. In the article, the authors revisit and generalize the question, “How close to abelian can a nonabelian finite group be?” To measure this, they defined

$$Pr(G) := \frac{|\{(a, b) \in G \times G : a \cdot b = b \cdot a\}|}{|G|^2}.$$

Quite literally, this is the probability that two elements in a finite group G

commute. Of course $Pr(G) = 1$ if and only if G is abelian. What is less obvious, however, is that if G is not abelian, $Pr(G) \leq \frac{5}{8}$. This bound is shown to be sharp by D_4 , among other groups [1]. From this result, we can see that there is a large gap between $\frac{5}{8}$ and 1. However, the main contribution of the article [7] was to introduce a concept similar to $Pr(G)$, but considering products of more than two elements. They denoted this new measure $P_n(G)$, where $P_n(G)$ measures the probability that a string of n elements is equal to its reverse. That is, what is the probability that for $a_1, a_2, \dots, a_n \in G$, we have $a_1 a_2 \cdots a_n = a_n \cdots a_2 a_1$? More precisely, they defined

$$P_n(G) := \frac{|\{(a_1, a_2, \dots, a_n) \in G^n : a_1 a_2 \cdots a_n = a_n \cdots a_2 a_1\}|}{|G|^n}.$$

Note here that $P_2(G) = Pr(G)$. In general, the authors proved that for even $n \geq 2$,

$$P_n(G) \leq \frac{1}{2} + \frac{1}{2^{n+1}} \text{ and } P_{n+1}(G) = P_n(G),$$

and they also address a number of other interesting variants on $P_n(G)$.

A New Measure Involving Group-theoretic Inverses

Motivated by the above work, our paper seeks to create a new numerical measure for a finite group G that involves a prominent use of inverses. More precisely, we will examine the probability that the inverse of a product of n elements is equal to the product of some permutation σ of the same n elements: $(a_1 a_2 \cdots a_n)^{-1} = (a_1 a_2 \cdots a_n)^\sigma$. Let us begin by considering the case $n = 2$. Then, we define

$$Inv(G) := \{(a, b) \in G \times G : (ab)^{-1} = ab \text{ or } (ab)^{-1} = ba\}.$$

Note that there are two possible conditions that the ordered pair can satisfy in order to be included in $Inv(G)$. Then, we define

$$Pr^{-1}(G) := \frac{|Inv(G)|}{|G|^2}.$$

Using a program written with the software Groups, Algorithms, and Programming (GAP) [4], we compute the value of $Pr^{-1}(G)$ for many well-known groups. (The notations $\mathbb{Z}_m, D_m, S_m, A_m$ refer to the families of cyclic, dihedral, symmetric, and alternating groups. See [2], [3], and [6] for more information.) These values are shown in Table 1 below.

m	$Pr^{-1}(\mathbb{Z}_m)$	$Pr^{-1}(D_m)$	$Pr^{-1}(S_m)$	$Pr^{-1}(A_m)$
1	1.0000	1.0000	1.0000	1.0000
2	1.0000	1.0000	1.0000	1.0000
3	0.3333	0.8333	0.8333	0.3333
4	0.5000	0.8750	0.5000	0.3333
5	0.2000	0.8000	0.2500	0.3167
6	0.3333	0.8333	0.1153	0.1417
7	0.1429	0.7857	0.0482	0.0440
8	0.2500	0.8125		
9	0.1111	0.7778		
10	0.2000	0.8000		

Table 1 – $Pr^{-1}(G)$ for $G = \mathbb{Z}_m, D_m, S_m$, and A_m .

From this data, we conjecture the following formulas for $Pr^{-1}(\mathbb{Z}_m)$ and $Pr^{-1}(D_m)$.

Formula 1 For all integers $m \geq 1$, $Pr^{-1}(\mathbb{Z}_m) = \begin{cases} \frac{1}{m} & \text{if } m \text{ is odd} \\ \frac{2}{m} & \text{if } m \text{ is even} \end{cases}$

Formula 2 For all integers $m \geq 3$, $Pr^{-1}(D_m) = \begin{cases} \frac{3}{4} + \frac{1}{4m} & \text{if } m \text{ is odd} \\ \frac{3}{4} + \frac{1}{2m} & \text{if } m \text{ is even.} \end{cases}$

These formulas turn out to be correct, and they are specific cases of general formulas which will be proven later. Until then, we will postpone this verification.

Proceeding parallel to [7], we generalize $Pr^{-1}(G)$ to a consideration of products of more than two elements. To this end, we define

$$Inv_n(G) := \{(a_1, a_2, \dots, a_n) \in G^n : (a_1 a_2 \cdots a_n)^{-1} = (a_1 a_2 \cdots a_n)^\sigma, \text{ for some } \sigma \in S_n\},$$

where $(a_1 a_2 \cdots a_n)^\sigma$ refers to a permutation of the subscripts of the product $a_1 a_2 \cdots a_n$. For example, $(a_1 a_2 a_3 a_4 a_5)^{(1,3)(2,5)(4)} = a_3 a_5 a_1 a_4 a_2$ and $(a_1 a_2 a_3 a_4 a_5)^{(2,5,3,1,4)} = a_3 a_4 a_5 a_1 a_2$. We define

$$Pr_n^{-1}(G) := \frac{|Inv_n(G)|}{|G|^n}.$$

This probability measure is similar to $P_n(G)$ in that it deals with n elements of a group. However, in contrast we look for an inverse of a product of elements that is equal to the product of elements in some order. For $n = 2$, groups containing many elements of order 2 or less will have higher values of $Pr_2^{-1}(G)$ than groups that do not have many of these elements (see Table 2 below). We can ask: What does it mean if $Pr_n^{-1}(G) = 1$? To answer this question, we recall that if G is a finite group such that every element has order 2 or less, then $G \cong \prod_{i=1}^k \mathbb{Z}_2$ (see [6], Exercise 7.2.25, Theorem 8.7). Clearly, the converse holds as well. With this information, we can prove the following:

Theorem 2.1 Let G be a finite group. Then,

$$Pr_n^{-1}(G) = 1 \quad \text{if and only if} \quad G \cong \prod_{i=1}^k \mathbb{Z}_2 \text{ for some } k \geq 0.$$

That is, G must be isomorphic to a finite direct product of \mathbb{Z}_2 with itself.

Proof For the backwards direction, since G is isomorphic to a finite direct product of \mathbb{Z}_2 with itself, G is abelian and every element of G has order 2 or less. Thus, by the definition of $Pr_n^{-1}(G)$, we have $Pr_n^{-1}(G) = 1$.

Conversely, if G is a group such that $Pr_n^{-1}(G) = 1$, then for every ordered n -tuple (g_1, g_2, \dots, g_n) belonging to G^n , we have $(g_1 g_2 \cdots g_n)^\sigma = (g_1 g_2 \cdots g_n)^{-1}$ for some σ in S_n . Now considering the n -tuple (g, e, e, \dots, e) in G^n , we obtain $g = g^{-1}$ for all g in G . Therefore, every element in G has order 2 or less. Thus, for some integer $k \geq 0$,

$$G \cong \prod_{i=1}^k \mathbb{Z}_2.$$

Theorem 2.1 helps us to answer the question, “What is $Pr_2^{-1}(G)$ really measuring?” We conjecture that $Pr^{-1}(G)$ is correlated to the fraction of elements in G with order 2 or less. The following Table 2 supports this claim.

Group	Percentage of Elements with order 2 or less	$Pr_2^{-1}(G)$
\mathbb{Z}_2	100.00	1.0000
D_4	75.00	0.8750
S_3	66.66	0.8333
D_6	66.66	0.8333
D_8	62.50	0.8125
D_5	60.00	0.8000
D_7	57.14	0.7857
\mathbb{Z}_4	50.00	0.5000
S_4	41.66	0.5000
\mathbb{Z}_3	33.33	0.3333
S_5	21.67	0.2500
\mathbb{Z}_5	20.00	0.2000

Table 2 – Comparing percentage of elements of order 2 or less with $Pr^{-1}(G)$.

Relating to the elements of order 2 or less, we can prove the following theorem:

Theorem 2.2 Given a finite group G ,

$$Pr_n^{-1}(G) \geq \frac{|\{g \in G : \text{order}(g) \leq 2\}|}{|G|}.$$

The following lemma will be useful in the proof of Theorem 2.2.

Lemma 2.3 Let G be a finite group. For all $g \in G$, there exist exactly $|G|^{n-1}$ n -tuples $(g_1, g_2, g_3, \dots, g_n)$ in G^n such that

$$g_1 g_2 g_3 \cdots g_n = g.$$

Proof Suppose $g \in G$. We can choose $g_1, g_2, g_3, \dots, g_{n-1}$ to be arbitrary elements of G and then choose $g_n = (g_1 g_2 \cdots g_{n-1})^{-1} g$. Thus the total number of choices for $(g_1, g_2, g_3, \dots, g_n) \in G^n$ is $|G|^{n-1}$.

With this lemma, we are now prepared to prove Theorem 2.2.

Proof of Theorem 2.2 Suppose $\text{order}(g) \leq 2$ and $g_1 g_2 \cdots g_n = g$. Then $g_1 g_2 \cdots g_n = (g_1 g_2 \cdots g_n)^{-1}$ and thus (g_1, g_2, \dots, g_n) is an element of $Inv_n(G)$.

By Lemma 2.3, there exists $|G|^{n-1}$ such n -tuples.

Thus $|Inv_n(G)| \geq |G|^{n-1} \cdot |\{g \in G : \text{order}(g) \leq 2\}|$. Therefore,

$$Pr_n^{-1}(G) = \frac{|Inv_n(G)|}{|G|^n} \geq \frac{|G|^{n-1} \cdot |\{g \in G : \text{order}(g) \leq 2\}|}{|G|^n} = \frac{|\{g \in G : \text{order}(g) \leq 2\}|}{|G|}.$$

Our next result states, in effect, that equality holds in Theorem 2.2 for all finite cyclic groups, which is confirmed by the data in Table 1 for $n = 2$.

Theorem 2.4 For all positive integers m and n , with $n \geq 2$, we have

$$Pr_n^{-1}(\mathbb{Z}_m) = \begin{cases} \frac{1}{m}, & \text{if } m \text{ is odd,} \\ \frac{2}{m}, & \text{if } m \text{ is even.} \end{cases}$$

Proof For this family of groups, we use additive notation. If g_1, g_2, \dots, g_k is an element of $Inv_n(\mathbb{Z}_m)$ then $(g_1 + g_2 + \cdots + g_n)^\sigma = -(g_1 + g_2 + \cdots + g_n)$ for some σ in S_n . However, since \mathbb{Z}_m is abelian, $(g_1 + g_2 + \cdots + g_n)^\sigma = g_1 + g_2 + \cdots + g_n$ for all σ in S_n . So, if (g_1, g_2, \dots, g_n) is an element of $Inv_n(\mathbb{Z}_m)$ then $g_1 + g_2 + \cdots + g_n = -(g_1 + g_2 + \cdots + g_n)$. This occurs only when

$$g_1 + g_2 + \cdots + g_n = 0$$

or

$$g_1 + g_2 + \cdots + g_n = \frac{m}{2}.$$

Case 1: If m is odd, then $\frac{m}{2}$ is not an element of \mathbb{Z}_m and thus if (g_1, g_2, \dots, g_n) is an element of $Inv_n(\mathbb{Z}_m)$ then $g_1 + g_2 + \cdots + g_n = 0$. By Lemma 2.3, this occurs for exactly m^{n-1} n -tuples of $(\mathbb{Z}_m)^n$. Thus,

$$Pr_n^{-1}(\mathbb{Z}_m) = \frac{m^{n-1}}{m^n} = \frac{1}{m}.$$

Case 2: If m is even, then $\frac{m}{2}$ is an element of \mathbb{Z}_m , and thus, we must also count elements such that $g_1 + g_2 + \cdots + g_n = \frac{m}{2}$. By Lemma 2.3, this occurs for exactly m^{n-1} n -tuples of $(\mathbb{Z}_m)^n$. Therefore, the total number of elements in $Inv_n(\mathbb{Z}_m)$ is $m^{n-1} + m^{n-1} = 2m^{n-1}$. Thus,

$$Pr_n^{-1}(\mathbb{Z}_m) = \frac{2m^{n-1}}{m^n} = \frac{2}{m}.$$

We can see that for each \mathbb{Z}_m ,

$$Pr_n^{-1}(\mathbb{Z}_m) = \frac{|\{g \in \mathbb{Z}_m : \text{order}(g) \leq 2\}|}{m},$$

for all $n \geq 2$. Now for a non-abelian group, how does $Pr_n^{-1}(G)$ behave?

We start with the family of dihedral groups, D_m .

Theorem 2.5 For all positive integers m and n , with $m \geq 3$ and $n \geq 2$, we have

$$Pr_n^{-1}(D_m) = \begin{cases} \frac{2^n - 1}{2^n} + \frac{1}{2^{nm}}, & \text{if } m \text{ is odd,} \\ \frac{2^n - 1}{2^n} + \frac{2}{2^{nm}}, & \text{if } m \text{ is even.} \end{cases}$$

Proof Let $(a_1, a_2, a_3, \dots, a_n)$ be an element of $(D_m)^n$. We prove the theorem using cases.

Case 1: Assume that at least one a_i is a flip. Let j be the smallest positive integer such that a_j is a flip and $a_1 a_2 a_3 \dots a_{j-1}$ are all rotations. Thus $a_1 a_2 a_3 \dots a_{j-1}$ is a rotation, and since a_j is a flip, then $a_1 a_2 a_3 \dots a_{j-1} a_j$ is a flip. At this point we break this case into two subcases:

Subcase A: Suppose $a_{j+1} a_{j+2} \dots a_n$ is a flip. Then since the inverse of a flip is itself,

$$\begin{aligned} [(a_1 a_2 a_3 \dots a_{j-1} a_j)(a_{j+1} a_{j+2} \dots a_n)]^{-1} &= (a_{j+1} a_{j+2} \dots a_n)^{-1} (a_1 a_2 a_3 \dots a_{j-1} a_j)^{-1} \\ &= (a_{j+1} a_{j+2} \dots a_n)(a_1 a_2 a_3 \dots a_{j-1} a_j). \end{aligned}$$

Thus, $a_1, a_2, a_3, \dots, a_{n-1}, a_n$ is an element of $Inv_n(G)$.

Subcase B: Suppose $a_{j+1} a_{j+2} \dots a_n$ is a rotation.

Then $(a_1 a_2 a_3 \dots a_n)^{-1} = (a_1 a_2 a_3 \dots a_n)$ is a flip, and thus,

$(a_1 a_2 a_3 \dots a_{j-1} a_j)(a_{j+1} a_{j+2} \dots a_n) = (a_1 a_2 \dots a_n)^{-1}$. Thus every ordered n -tuple of $(D_m)^n$ which contains at least one flip is an element of $Inv_n(G)$.

In this case we represent all but one element as either a flip or rotation. Thus, for each of the n elements we have two choices, either a flip or rotation.

However, we must remove the special case where every element is a rotation. This gives us $2n - 1$ choices. It follows that we have m choices for each of

the n rotations and n flips. Thus, by the multiplication rule we find a total $(2^n - 1)m^n$ of such elements.

Case 2: If $a_1, a_2, a_3, \dots, a_n$ are all rotations, then since the rotations form an abelian subgroup of D_m , $(a_1 a_2 a_3 \dots a_n)^\sigma = (a_1 a_2 a_3 \dots a_n)$ for all σ in S_n . So $(a_1 a_2 a_3 \dots a_n)^\sigma = (a_1 a_2 a_3 \dots a_n)^{-1}$ if and only if $\text{order}(a_1 a_2 a_3 \dots a_n) \leq 2$, which only occurs if $a_1 a_2 a_3 \dots a_n = e$ or $a_1 a_2 a_3 \dots a_n = r^{\frac{m}{2}}$ where $r^{\frac{m}{2}}$ is rotation of 180 degrees, and only exists if m is even.

This case shows there are m^{n-1} such elements for an odd m , and $2m^{n-1}$ such elements for an even m by Lemma 2.3.

Summing across both cases we get a total of $(2^n - 1)m^n + 2m^{n-1}$ elements in $Inv(D_m)$ if m is odd and $(2^n - 1)m^n + 2m^{n-1}$ if m is even. Then dividing by $|D_m|^n$, we obtain our result.

These formulas for $Pr_n^{-1}(G)$ in well-known groups are useful when considering the theorems in the following section.

Further Results

Theorem 3.1 Given a finite group G and an integer $n \geq 2$, $Pr_n^{-1}(G) \leq Pr_{n+1}^{-1}(G)$.

Proof Let $(g_1, g_2, g_3, \dots, g_n)$ be an element of $Inv_n(G)$. Therefore

$$(g_1 g_2 g_3 \dots g_n)^\sigma = (g_1 g_2 g_3 \dots g_n)^{-1} \text{ for some } \sigma \text{ in } S_n.$$

Thus $(g_1, g_2, g_3, \dots, g_{n-1}, a, b)$ is an element of $Inv_{n+1}(G)$ if $ab = g_n$. By Lemma 2.3, this occurs for exactly $|G|$ ordered pairs of (a, b) in G^2 . So for all $(g_1, g_2, g_3, \dots, g_n)$ in $Inv_n(G)$, there are at least $|G|$ distinct elements of $Inv_{n+1}(G)$. Thus,

$$|G| |Inv_n(G)| \leq |Inv_{n+1}(G)|.$$

Therefore,

$$Pr_n^{-1}(G) = \frac{|Inv_n(G)|}{|G|^n} = \frac{|G| |Inv_n(G)|}{|G|^{n+1}} \leq \frac{|Inv_{n+1}(G)|}{|G|^{n+1}} = Pr_{n+1}^{-1}(G).$$

One way to create new larger groups from already established ones is to use direct products (For more information on direct products, see [6], [2]).

Using the same GAP program as before, we can collect data on the direct products shown below:

$n \setminus m$	1	2	3	4	5	6	7
1	1.0000	1.0000	0.3333	0.5000	0.2000	0.3333	0.1429
2	1.0000	1.0000	0.3333	0.5000	0.2000	0.3333	0.1429
3	0.8333	0.8333	0.2778	0.4167	0.1667	0.2778	0.1191
4	0.5000	0.5000	0.1667	0.2500	0.1000	0.1667	0.0714
5	0.2500	0.2500	0.0833	0.1250	0.0500	0.0833	0.0357
6	0.1153	0.1153	0.0384	0.0577	0.0231	0.0384	0.0165
7	0.0482	0.0482	0.0161	0.0241	0.0096	0.0161	0.0069

Table 3 – $Pr^{-1}(\mathbb{Z}_m \times S_n)$.

$n \setminus m$	1	2	3	4	5	6	7
1	1.0000	1.0000	0.8333	0.8750	0.8000	0.8333	0.7857
2	1.0000	1.0000	0.8333	0.8750	0.8000	0.8333	0.7857
3	0.8333	0.8333	0.5833	0.6458	0.5333	0.5833	0.5119
4	0.5000	0.5000	0.3403	0.3802	0.3083	0.3403	0.2946
5	0.2500	0.2500	0.1653	0.1865	0.1483	0.1653	0.1411
6	0.1153	0.1153	0.0384	0.0577	0.0231	0.0384	0.0165
7	0.0482	0.0482	0.0161	0.0241	0.0096	0.0161	0.0069

Table 3 – $Pr^{-1}(D_m \times S_n)$.

Again, a few trends are noticed in this data. These trends are confirmed by the following theorems.

Theorem 3.2 Given finite groups G and H ,

$$Pr_n^{-1}(G \times H) \leq Pr_n^{-1}(G) \cdot Pr_n^{-1}(H).$$

Proof Because $|G \times H| = |G| \cdot |H|$ we are left to show that,

$$|Inv_n(G \times H)| \leq |Inv_n(G)| \cdot |Inv_n(H)|$$

Any element of $Inv_n(G \times H)$ has the form $((g_1, h_1), (g_2, h_2), \dots, (g_n, h_n))$ with g_i in G and h_j in H and $(g_1 g_2 \cdots g_n)^\sigma = (g_1 g_2 \cdots g_n)^{-1}$ and $(h_1 h_2 \cdots h_n)^\sigma = (h_1 h_2 \cdots h_n)^{-1}$ for some σ in S_n . So in order for $((g_1, h_1), (g_2, h_2), \dots, (g_n, h_n))$ to be an element of $Inv_n(G \times H)$, (g_1, g_2, \dots, g_n) must be an element of $Inv_n(G)$ and (h_1, h_2, \dots, h_n) must be an element of $Inv_n(H)$. Thus the maximum size of $Inv_n(G \times H)$ is $|Inv_n(G)| \cdot |Inv_n(H)|$.

Theorem 3.3 Given finite groups G and H such that at least one of them is abelian,

$$Pr_n^{-1}(G \times H) = Pr_n^{-1}(G) \cdot Pr_n^{-1}(H).$$

Proof Let us assume without loss of generality that H is abelian. Because

$|G \times H| = |G| \cdot |H|$, we are left to show that

$$|Inv_n(G \times H)| = |Inv_n(G)| \cdot |Inv_n(H)|.$$

If $(h_1 h_2 \cdots h_n)^\alpha = (h_1 h_2 \cdots h_n)^{-1}$ for some α in S_n , then

$(h_1 h_2 \cdots h_n)^\sigma = (h_1 h_2 \cdots h_n)^{-1}$ for all σ in S_n , since H is abelian.

So if $(g_1 g_2 \cdots g_n)^\sigma = (g_1 g_2 \cdots g_n)^{-1}$ for some σ in S_n then

$(h_1 h_2 \cdots h_n)^\sigma = (h_1 h_2 \cdots h_n)^{-1}$ for that same σ . So we conclude that

if (g_1, g_2, \dots, g_n) is an element of $Inv_n(G)$ and (h_1, h_2, \dots, h_n) is an element of $Inv_n(H)$ then $((g_1, h_1), (g_2, h_2), \dots, (g_n, h_n))$ is an element of $Inv_n(G \times H)$.

So

$$|Inv_n(G \times H)| \geq |Inv_n(G)| \cdot |Inv_n(H)|,$$

and by Theorem 3.2,

$$|Inv_n(G \times H)| \leq |Inv_n(G)| \cdot |Inv_n(H)|.$$

Thus,

$$|Inv_n(G \times H)| = |Inv_n(G)| \cdot |Inv_n(H)|.$$

Therefore,

$$Pr_n^{-1}(G \times H) = Pr_n^{-1}(G) \cdot Pr_n^{-1}(H).$$

It remains to be seen whether it is possible for the equality in Theorem 3.3 to hold if both G and H are non-abelian. In the next section, we will describe a number of other unresolved research problems involving $Pr_n^{-1}(G)$.

Further Questions

To get a better understanding of $Pr_n^{-1}(G)$, we can draw connections to the concept of $Pr(G)$. Since there is extensive literature on the latter, we are left with no shortage of questions to ask about $Pr_n^{-1}(G)$. In [8] the author shows that under certain conditions, if N is a normal subgroup of G , then

$$Pr(G/N) = Pr(G).$$

Thus, it is natural to ask if there is any relation between $Pr_n^{-1}(G)$, $Pr_n^{-1}(N)$ and $Pr_n^{-1}(G/N)$.

The concept of $Pr(G)$ is usually associated with finite groups only. However, in [5] the author generalizes the concept of $Pr(G)$ to include compact, Hausdorff topological groups. This generalization to infinite groups even preserves the $\frac{5}{8}$ bound on commutativity discussed at the beginning of the paper for finite groups. This concept provides encouragement to develop a generalization of $Pr_n^{-1}(G)$ for infinite groups as well as to determine an upper bound for $Pr_n^{-1}(G)$.

Other questions concerning $Pr_n^{-1}(G)$ include finding formulas for $Pr_n^{-1}(S_k)$ and $Pr_n^{-1}(A_k)$ and other common families of groups. Another question is for what values of $r \in [0, 1]$ does there exist a group G such that $Pr_n^{-1}(G) = r$? One way to obtain a more conceptual understanding of $Pr_n^{-1}(G)$ is to prove that if G and H are finite groups and,

$$Pr_2^{-1}(G) \leq Pr_2^{-1}(H),$$

then H has a higher percentage of elements of order two or less than G . However, at present this is only a conjecture. Furthermore, if we know more generally that $Pr_n^{-1}(G) \leq Pr_n^{-1}(H)$ for some value(s) of n , what relationship between G and H does this suggest?

Acknowledgement

The authors would like to thank their advisor, Dr. Scott Annin, for his guidance on this problem. In addition the authors are grateful to the referee for many helpful comments and suggestions on this manuscript.

References

Clifton, C., Guichard, D., Keef, P., *How Commutative are Direct Products of Dihedral Groups?*, Math. Mag., 84 (2011), 137-140.

Dummit, D., Foote, R. *Abstract Algebra*, 3rd ed., John Wiley & Sons, Inc., Hoboken, NJ, 932 (2004).

Gallian, J. *Contemporary Abstract Algebra*, 7th ed., Brooks Cole, Belmont, CA, (2010)

GAP 2008, The GAP Group, GAP-Groups, Algorithms, and Programming, Version 4.4.12; 2008. (<http://www.gap-system.org>).

Gustafson, W.H. *What is the Probability that Two Group Elements Commute?*, Amer. Math. Monthly, 80 (1973), 1031-1034.

Hungerford, T. *Abstract Algebra: An Introduction*, 2nd ed., Brooks Cole, Florence, KY, (1996)

Langley, L., Levitt, D., Rower, J. *Two Generalizations of the 5/8 Bound on Commutativity in Nonabelian Finite Groups*, Math. Mag., 84 (2011), 128-136.

Rusin, D. *What is the Probability that Two Elements of a Finite Group Commute?*, Pacific J. Math., 82 (1979), 237-247.

To Err is Human, to Trisect Divine...

Department of Mathematics, College of Natural Science and Mathematics
California State University, Fullerton, CA, USA

Jonathan Sayre
Advisor: Dr. Scott Annin

Abstract

In this paper we will place the link between impossible geometric constructions and abstract algebra into historical context. First we develop a field of constructible numbers. Then we set up the relationship between constructions and algebra. Next, we establish a connection between algebraic equations and impossibility followed by the proofs of the three classical construction problems. Finally, we will reflect on some observations made through the exploration of this subject, including a few words about false proofs, their flaws, and a look at dissenting viewpoints regarding the possible flaws in the ‘explanation of impossibility.’

Introduction

Since its inception, the study of geometry has branched into many directions. But from the distant past of geometry’s history, there remain three famous impossibility problems, queries, or challenges as they might be called. In the nineteenth century, mathematician Pierre Laurent Wantzel (1837) [1] took the tools of abstract algebra and applied them to an impossibility problem, that of trisecting an angle. Through this application of abstract algebra, the other two problems were shown to be impossible as well. But the proofs of the three problems take an understanding of the complexities of all three problems to tie them together.

The first of these three problems stems from a legend from ancient Greece. As the legend goes, the god Apollo placed a plague on the town of Delos and the oracle there appealed to Delphi for a way to end the plague. The oracle told the people they must double the size of the Temple of Apollo, which was in the shape of a cube. The problem was put to Plato,

who explained that the citizenry of Delos should spend more time studying geometry. Initially the people thought that the solution to the problem was easy: simply double the length, width, and height of the temple. The oracle explained that the task was to double the volume of the temple. Plato then had three men work on this problem, but rebuked their solution for not using pure geometry [2]. Alas, the first of our three problems is revealed, duplication of a cube (which is equivalent to construction of $\sqrt[3]{2}$). Combine this with the problem of squaring a circle and trisecting an angle and you have a trifecta of impossible problems in geometric construction.

The main tools in geometric construction are the compass and straightedge, and seeing how these tools can be related to algebra allows one to fully comprehend why each problem is impossible to solve.

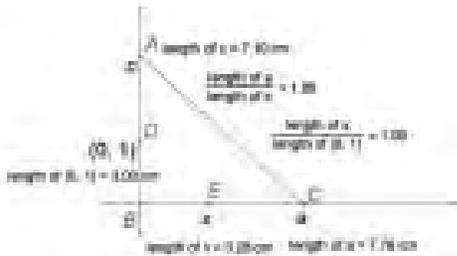
Tools and their connection

The Greeks were of the opinion that pure geometry was rendered through the use of only a compass and a straightedge. In actuality, what we call a compass today was a stick and a chord or two sticks held together with a string. The straightedge may have been the stick used in the compass. Whatever was used as a straightedge did not have any marks on it because the units of measure were arbitrary. Though limited, these tools could easily perform addition, subtraction, multiplication, division, and square roots of line segments. The Greeks also put a caveat on constructions in that the process must consist of a finite number of steps. The reason for having this caveat may seem a little silly, but it is there in order to avoid a situation similar to Zeno’s paradox. That is to say, if a method gets progressively close to an answer, but never reaches that answer exactly, then it would not qualify as a solution.

The operations of addition and subtraction begin with choosing two points. This establishes a line and a unit of length. From there, the processes of multiplying, dividing, and finding square roots are all based on these two points. In essence, a coordinate system is generated from the original two points. The relationship between the operations and the coordinates is what establishes a set of points which we can call constructible points.

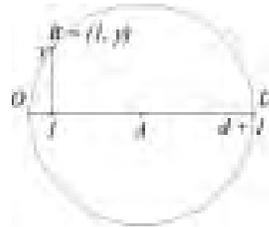
The value for the distance between points is called a constructible number. Thus constructible numbers and constructible points are interchangeable in meaning. In [3], the following theorem is stated:

Theorem 1: "Let a, b, c, d be constructible numbers with $c \neq 0$ and $d > 0$. Then each of $a + b, a - b, ab, a/c$ and \sqrt{d} is a constructible number." ([3], pg. 453). The proofs of addition and subtraction are straightforward but division, multiplication, and roots take a little more finesse:



Proof: Let a and c be line segments lengths represented on a Cartesian plane where a is along the x -axis and c is along the y -axis. Thus we have points $(a, 0)$ (labeled C) and $(0, c)$ (labeled A) on our plane. Construct the point $(0, 1)$ (labeled D) on the y -axis and a line from $(a, 0)$ to $(0, c)$. Now construct a line from $(0, 1)$ to the x -axis that is parallel to the line from $(a, 0)$ to $(0, c)$ and label this intersection $(x, 0)$ (labeled E). Claim: $x = a/c$. Note that since $\overline{AC} \parallel \overline{DE}$, then $\triangle ABC$ is similar to $\triangle DBE$. Thus the following relationship exists between the sides: $\frac{AB}{BC} \equiv \frac{DB}{BE}$.

Therefore we have $c/a = 1/x$. Solving for x we get $x = a/c$. Thus the fraction a/c has been constructed. Using a similar method with a line from $(0, 1)$ to a and a parallel line from c to the x -axis, we get the relationship of $1/a = c/x$. Solving for x we get $x = ac$.



To construct the square root of d , we construct a segment equal to $d + 1$. Find the midpoint of this segment and construct a semicircle. Draw a perpendicular up from the end on the segment equal to the length of 1. Label the points 1, D , A , and B as shown in the figure. The segment from 1 to B has length y . Claim: $y^2 = d$. By a theorem from geometry, we know that $\triangle OBD$ is right at vertex B . Now we apply the Pythagorean Theorem to the triangles to get:

$$\begin{aligned} \overline{OB}^2 + \overline{BD}^2 &= \overline{OD}^2 \\ (1^2 + y^2) + (y^2 + d^2) &= (1 + d)^2 \\ 1^2 + 2y^2 + d^2 &= 1^2 + 2d + d^2 \\ 2y^2 &= 2d \\ y^2 &= d \\ y &= \sqrt{d} \end{aligned}$$

Thus \sqrt{d} can be constructed, provided that d was constructible in the first place.

The field our tools create

With this concept established, we begin to get an idea of the structure of our set of constructible numbers. From Hadlock's *Field Theory and Its Classical Problems*:

Definition 1 : "Let F be a subset of the set \mathbb{R} of real numbers. We say that F is a field if it satisfies the following two conditions:

- (1) F is closed under the rational operations (addition, subtraction, multiplication, and division except by 0), meaning that whenever these operations are applied to elements of F , the result is an element of F .
- (2) The number 1 is an element of F ." [4]

From the (addition, subtraction, multiplication, and division) constructions above, we see that the integers, \mathbb{Z} , and the rational numbers, \mathbb{Q} , belong

to the ‘field of constructible numbers.’ Therefore the unusual elements of the constructible numbers are square roots. Since the process of constructing square roots geometrically can be iterated, any members of the field will be of the form $a + b\sqrt{c} + \sqrt{d + e\sqrt{f}} + \dots$ (with $a, b, c, d, e, f \in \mathbf{F}$). Hungerford’s *Abstract Algebra* gives the following definition and result that play a central role in what follows:

Lemma 1 : Let \mathbf{F} be a subfield of \mathbb{R} and k a positive element of \mathbf{F} such that $k \notin \mathbf{F}$. Let

$$\mathbf{F}(\sqrt{k}) := \{a + b\sqrt{k} \mid a, b \in \mathbf{F}\}.$$

- (1) $\mathbf{F}(\sqrt{k})$ is a subfield of \mathbb{R} that contains \mathbf{F} .
- (2) Every element of $\mathbf{F}(\sqrt{k})$ can be written uniquely in the form $a + b\sqrt{k}$ with $a, b \in \mathbf{F}$.

Definition 2 : The field $\mathbf{F}(\sqrt{k})$ in Lemma 1 is called a **quadratic extension** field of \mathbf{F} . (Quadratic extension fields play a crucial role in determining which numbers are constructible.) ([3], pg. 455)

By Theorem 1, we now have a field of constructible numbers that contains \mathbb{Z} , \mathbb{Q} , and the quadratic extension field, $\mathbf{F}(\sqrt{k})$. In the interest of clarity, we will denote the field of constructible numbers by \mathbf{C}_N to distinguish it from other general fields discussed. Since a straightedge is used to draw lines and a compass is used to draw circles, we need to associate the drawing of these figures with the elements of the field \mathbf{C}_N . Hungerford’s *Abstract Algebra* does this nicely in a lemma ([3], p. 454):

Lemma 2 : Let \mathbf{C}_N be the field of constructible numbers.

- (1) If a line contains two points whose coordinates are in \mathbf{C}_N , then the line has an equation of the form

$$ax + by + c = 0, \text{ where } a, b, c, \in \mathbf{C}_N.$$

- (2) If the center of a circle is a point whose coordinates are in \mathbf{C}_N and the radius of the circle is a number whose square is in \mathbf{C}_N , then the circle has an equation of the form

$$x^2 + y^2 + rx + sy + t = 0, \text{ where } r, s, t \in \mathbf{C}_N.$$

By using the equations from Lemma 2, we obtain a generalized polynomial for points that are reached with the combination of a straightedge and compass in the form

$$ax^2 + by^2 + cx + dy + e = 0,$$

where a, b, c, d , and e must be from the field, \mathbf{C}_N .

Why these equations are significant and how they relate to the impossible problems will be addressed shortly. The significance is in the solutions to the equations, which must be from the field, \mathbf{C}_N . We now present a lemma ([3], pg. 455) that builds on Lemma 2 by considering intersections of lines and circles with coefficients in \mathbf{C}_N .

Lemma 3 : Let L_1 and L_2 be lines whose equations have coefficients in \mathbf{C}_N . Let

- C_1 and C_2 be circles whose equations have coefficients in \mathbf{C}_N . Then
- (1) If L_1 intersects L_2 then the point of intersection has coordinates in \mathbf{C}_N .
- (2) If C_1 intersects C_2 then the points of intersection have coordinates in \mathbf{C}_N .
- (3) If L_1 intersects C_1 then the points of intersection have coordinates in \mathbf{C}_N .

Lemma 3 implies that the intersection points of lines and circles (whose equations have coefficients in \mathbf{C}_N) have components that are elements of \mathbf{C}_N . We know that our field contains the integers and rational numbers within it (or subsets of it) and that our field is a subset of the real numbers. Hungerford’s book then gives us a nice way to describe our field. ([3], p.457)

Theorem 2 : If a real number r is constructible, then there is a finite chain of fields

$$\mathbb{Q} = \mathbf{F}_0 \subseteq \mathbf{F}_1 \subseteq \dots \subseteq \mathbf{F}_n \subseteq \mathbf{C}_N$$

such that $r \in \mathbf{F}_n$ and each \mathbf{F}_i is a quadratic extension of the preceding field, that is,

$$\mathbf{F}_1 = \mathbb{Q}(\sqrt{c_0}), \quad \mathbf{F}_2 = \mathbf{F}_1(\sqrt{c_1}), \quad \mathbf{F}_3 = \mathbf{F}_2(\sqrt{c_2}), \dots, \mathbf{F}_n = \mathbf{F}_{n-1}(\sqrt{c_{n-1}}),$$

where $c_i \in \mathbf{F}_i$ and $c_i \notin \mathbf{F}_i$ for $i = 0, 1, 2, \dots, n-1$.

This concept of tying geometry to algebraic equations gets most of its advances beginning in the 16th century. The work of del Ferro, Tartaglia, Cardano, and Cardano’s student Viète gave us the solution to the cubic and quartic equations. In the next century, Fermat and Descartes used “algebra to solve geometric problems and vice versa.” ([5], pg.15) Toward the end of the 18th century, Gauss’ interest in constructions, which was the motivation

to his becoming a mathematician, led to “Key events in the development of modern algebra, involving insight into the hidden, abstract properties of polynomials...” ([5], pg. 16) It is these properties that lead us to the factoring of polynomials and the roots of polynomials.

The main tool from abstract algebra that we now need is sometimes referred to as the Tower Theorem for Field Extensions ([3], p.336).

Theorem 3 : If $F \subseteq E \subseteq K$ are three fields, each contained in the next such that $\deg(K/E)$ and $\deg(E/F)$ are finite, then

$$\deg K/F = (\deg K/E) (\deg E/F).$$

When Theorem 3 is tied together with the following corollary, the insight about why certain numbers are ‘out of range’ of the straightedge and compass begins to become clear.

Corollary 2 : For the field F_n in Theorem 2, we have $\deg F_n/\mathbb{Q}$ is a power of 2. ([6], pg. 242)

Proof : Adjoining a square root of an element that is not already a perfect square creates a field extension of degree 2. So, by Theorem 3, we find that $\deg F_n/\mathbb{Q} = 2^k$ for some $k \geq 1$.

The three ‘impossible to solve’ problems

Duplication of the cube: Let there be a cube with side length of 1 unit, and thus, volume 1^3 . What is the length of the side of a cube whose volume is 2^3 ? This can be written $x^3 = 2$. We want to know if $x = \sqrt[3]{2}$ is constructible. That is, is $x \in \mathbf{C}_N$?

Proof: (by way of contradiction) Suppose $x = \sqrt[3]{2}$ is constructible. It is a root of the equation $x^3 - 2 = 0$. The polynomial $x^3 - 2$ is irreducible over \mathbb{Q} . Thus, the field $\mathbb{Q}(\sqrt[3]{2})$ is an extension of degree 3 of \mathbb{Q} . By Corollary 2, since 3 is not a power of 2, we conclude $\sqrt[3]{2} \notin \mathbf{C}_N$. [Translated from Hartshorne’s ([6], pg. 243) for this paper.]

Trisection of the angle: The issue of trisecting an angle requires more discussion than the duplication of a cube. The Greeks knew that some angles could be trisected (i.e. 360° , 180° , 90° , 45° , and combinations of these). So the fact is not that trisection *cannot* be successful accomplished, but the fact is that the trisection *can not always* be accomplished. The Rational Roots Test plays a pivotal role in the proof.

Proof: (by way of contradiction/counterexample) Suppose that any angle can be trisected. Take the angle measuring 60 degrees. Then $\cos 20^\circ$ would be constructible in \mathbf{C}_N . We must then find $\cos 20^\circ$. The triple angle formula is

$$\cos 3\theta = 4\cos^3 \theta - 3\cos \theta.$$

Substituting $\theta=20^\circ$; we have $3\theta = 60^\circ$ and $\cos 3\theta = \frac{1}{2}$. Also let $x = 2 \cos \theta = 2 \cos 20^\circ$. Then

$$\frac{1}{2} = \frac{x^3}{2} - \frac{3x}{2} \text{ or } 1 = x^3 - 3x.$$

It can be shown with the Rational Roots Test that the cubic equation $x^3 - 3x - 1 = 0$ is irreducible over \mathbb{Q} . Details may be found in [7], pg. 33-34.

Before continuing to the final impossible problem, we must point out that the previous problem begins to shed some light on the solution to the next problem. To help us see why, we recall the following important definition from abstract algebra:

Definition 4 : An algebraic number over \mathbb{Q} is a complex number that is the root of some monic polynomial with rational coefficients.

The following result is well known:

Theorem 4 : If t is an algebraic number over \mathbb{Q} and is the root of a polynomial degree n in $\mathbb{Q}[x]$, then

$$\mathbb{Q}(t) = \{a_0 + a_1x + a_2t^2 + \dots + a_{n-1}t^{n-1} \mid a_i \in \mathbb{Q}\}$$

is a subfield of \mathbb{C} and every element in $\mathbb{Q}(t)$ is an algebraic number. ([5], pg. 313)

Note that any number that can be written using fractions and roots is algebraic over \mathbb{Q} . Take, for example, cosine 18° and cosine 21° . The following formulas show that these are both algebraic over \mathbb{Q} .

$$\begin{aligned} \cos \frac{\pi}{10} &= \cos 18^\circ = \frac{1}{4} \sqrt{5 + \sqrt{5}} \\ \cos \frac{7\pi}{60} &= \cos 21^\circ = \frac{1}{16} \left[2\sqrt{3} - 1 + \sqrt{3 - \sqrt{3}} + \sqrt{2} \left(\sqrt{3} + 1 \right) + \sqrt{5} \right] \end{aligned}$$

However, $\cos 20^\circ$, which we explored in the previous impossible problem, cannot be written in this form and is not constructible. Indeed, $\cos 20^\circ$ is not algebraic over \mathbb{Q} , therefore not constructible. This fact is embodied in the following theorem:

Theorem 5 : All elements of the field \mathbf{C}_N are algebraic over \mathbb{Q} .
Conversely, as 32 illustrates, not every algebraic number is constructible.
With this insight, we press forward to our last impossible problem.

Square a circle: The explanation of why the area of a circle cannot be constructed into a square stems from the fact that π is not algebraic. In 1882, Lindemann proved that π was *transcendental*. A transcendental number is the opposite of an algebraic number. There does not exist a finite combination of algebraic expressions that can be used to represent a number that is transcendental, such as π or e . The proof that π is transcendental is very complex and involves a variety of mathematical techniques ranging from calculus to number theory which is beyond the scope of this paper. We will leave Lindemann's assertion as a theorem.

The Lindemann Theorem : π is *transcendental*.

Armed with this knowledge, we have a very simple proof that a circle cannot be squared.

Proof : Given a circle, suppose we can construct a square of the same area. Let C be a circle of radius 1. We need to show the area of circle C can be constructed into the shape of a square with the same area. By the area formula for circles we know the area of circle C to be π units square. To construct a square of equal area, we need a side to be the π . By Theorem 3, π must be an algebraic number to be within our field \mathbf{C}_N . By the Lindemann Theorem, π is transcendental and not algebraic. Therefore $\pi \notin \mathbf{C}_N$ and cannot be constructed. Since π cannot be constructed, then π cannot be constructed. Therefore, a circle can not be squared.

Conclusion

Since the inception of the three famous impossible construction problems, many have claimed to have found a way to solve them. But all have been debunked in one way or another. This use of abstract algebra concepts forms a strong argument for the impossibility of the solutions. Yet there has been at least one renowned mathematician who has a viewpoint that challenges these arguments. Kronecker proposed that there are no irrational numbers. 'He is famous for saying: "God made the natural numbers, all the rest is the work of man." He is also said to have told Lindemann, á propos of the transcendence of π : "Of what use is your beautiful proof, since irrational numbers do not exist?" ([5], pg. 73)

As Kronecker probably realized, the reduction of mathematical concepts to finite sequences of integers can only be done in certain limited domains (which he happened to be interested in), such as algebraic number theory. Even then, this reduction does not necessarily make the theory easier to understand. Adopting this philosophy means, for example, working with the congruence relation rather than with congruence classes. This is one reason Kronecker's work is harder to follow (for most people) than Dedekind's. ([5], pg. 73)

A point of interest to take note of in trisecting angles, for each angle that can be constructed, there exists two more which cannot. For example, 3° can be constructed but 1° and 2° cannot, 30° can, while 10° and 20° can not. The proofs of impossibilities are founded based on the concept of two initial points, but what if a third arbitrary point was selected? Would there be a 2/3 chance that the new arbitrary point was not an element of the field currently being used, and if so, would the new arbitrary point lead to a possible solution? With this uncertainty, the dream of solving the unsolvable will have amateur 'trisectors' popping up in the future. And maybe, though unsuccessful in solving these problems, the interest will create another Gauss or Wantzel or Euler. Who knows, we can only wait and see.

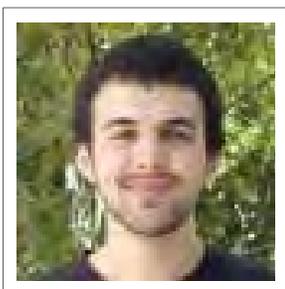
References

- "Pierre Laurent Wantzel." *Encyclopædia Britannica. Encyclopædia Britannica Online*. Encyclopædia Britannica Inc., 2012. Web. 14 Mar. 2012
- "doubling the cube." *Encyclopædia Britannica. Encyclopædia Britannica Online*. Encyclopædia Britannica Inc., 2012. Web. 14 Mar. 2012.
- Hungerford, T. (1997). *Abstract Algebra: An Introduction*. Fort Worth: Saunders College Pub.
- Hadlock, C. (1978). *Field Theory and Its Classical Problems*. [Washington]: Mathematical Association of America.
- Stillwell, J. (1994). *Elements of Algebra: Geometry, Numbers, Equations*. New York: Springer-Verlag.
- Hartshorne, R. (2000). *Geometry: Euclid and Beyond*. New York: Springer.
- Bold, B. (1969). *Famous Problems of Mathematics: A History of Constructions with Straight Edge and Compasses*. New York: Van Nostrand Reinhold Co.

AUTHOR BIOGRAPHIES



ASHLEY MELENDEZ graduated in 2011 with a B.S. degree in Geology. Her love of the outdoors and the natural world has led her to geological sciences and she will be a lifelong learner as she attempts to better understand the world around her. Under the direction of her advisor, Dr. Brandon Browne, she completed an undergraduate thesis that investigated magmatic processes recorded by deposits produced during the 2006 eruption of Augustine Volcano in southwestern Alaska. Ashley visited Augustine Volcano in the summer of 2010 with Dr. Browne and several other scientists from the US Geological Survey to complete her field work. She was awarded best undergraduate proposal poster at the 2011 CSUF Geology Research Day for her presentation on this research.



DANNY ORTON is a senior majoring in pure mathematics and physics at California State University Fullerton. He is currently working with Dr. Scott Annin researching aspects of Group Theory. Also, he is conducting research with Dr. Morty Khakoo in molecular physics. After graduating Danny will go on to pursue a Ph.D. in either pure mathematics or physics.

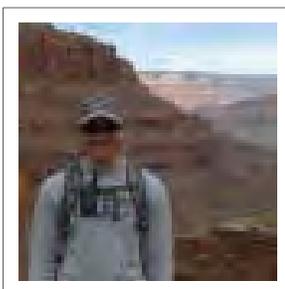


DANIEL LENDERS is a junior at CSUF majoring in Mathematics in the pure concentration. His current pursuit in mathematics is from a dream he had in second grade to become a mathematician. In his pursuit, he has been actively moving toward his goal by being involved with math club and by undertaking research with Dr. Scott Annin in finite group theory. Daniel hopes to one day attain a Ph. D in mathematics and be able to convey his knowledge and inspire others to become involved in math. In his spare time, Daniel plays video games with his friends, and he enjoys conversing about philosophy.



EMILY NGUYEN WIEBER will graduate with a B. S. in May 2012.

She emigrated to the U.S. from southern Viet Nam since 2005. Despite being the first college generation, with her ambition and tenacity, she has achieved high academic and research accomplishments at CSUF. She has won numerous scholarship awards such as Rachel Carson in Conservation Biology, David Walkington Memorial, etc... She contributed her accomplishments to her research advisors, Dr. Schenk and the directors of the McNair and SCERP program. Her undergraduate research is investigating citrus plant vulnerability to embolism. In fall 2012, she will start a Master program at CSUF and continue working with Dr. Schenk on quantifying water usage and comparative citrus fruit quality. She plans to pursue a Ph. D. in sustainable agriculture. Her ultimate goal is to become a faculty at a university.



ERIK CADARET is an undergraduate student at California State University, Fullerton pursuing a bachelor's degree of science in Geology. He is emphasizing Hydrogeology, Geochemistry, and Geological Engineering in hopes that with these emphases, he can pursue a master's degree to ultimately become a Hydro-philanthropist and a professional consulting geologist. His research on the hydrogeology and geochemical interactions in the Sheep Creek fan area will hopefully provide insight as to the potential for artificial recharge to augment the declining water levels in the area.



ERNEST NUNEZ is a geology major at California State University, Fullerton. He is currently finishing his undergraduate thesis project studying fault scarp morphology along the Northern Eureka Valley Fault Zone, Eureka Valley, Eastern California under his advisor Jeff Knott. His participation in the Louis Stokes Alliances for Minority Participation program funded his research over the summer of 2011 to obtain volcanic glass and basalt samples for analysis using analytical instruments such as an electron microprobe at Cal Tech, and an X-ray fluorescence unit at Pomona College. Ernest will present his research at the 2012 American Association of Petroleum Geologists (AAPG). Part of his undergraduate experience includes a variety of leadership positions, including serving as the first President of his junior college Math Club. He will graduate from California State University, Fullerton in the summer of 2013 with a B.S. degree in Geology, then plans on attending graduate school after graduation.



JOHN HENNESSY received his B.S. in geology from CSUF in August 2011. He had been an undergraduate student at CSUF since transferring from Cypress College in the fall of 2008. Upon arriving at CSUF, he immediately realized he was part of a uniquely great geology program and is ever-grateful for the experiences and opportunities the CSUF geology department has given him. He was advised by Dr. Brandon Browne, who provided insightful knowledge and guidance during John's undergraduate degree. John is now employed as a geologist for a geological consulting company and is currently involved with international mining projects that include regional seismic hazard assessments and geotechnical investigations. He is a member of the South Coast Geological Society and frequently attends their monthly meetings in Santa Ana, CA.



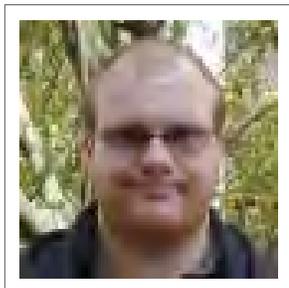
JORLY CHATOUPHONEXAY is a junior majoring in applied mathematics at California State University, Fullerton (CSUF). She is currently working with Dr. Angel R. Pineda on the Maximum Likelihood Estimation of the Fat Fraction Using Magnetic Resonance Imaging at High Signal-to-Noise Ratio. Jorly would like to thank Dr. Angel R. Pineda from the CSUF mathematics and Dr. Amybeth Cohen from the CSUF biological science departments for their mentorship. Jorly would also like to thank the Minority Access to Research Careers (MARC) program and the CSUF Honors program for their support. She wishes everyone here at CSUF success and happiness.



LAURA TOLENTINO is working on her B.S. degree in Cell and Developmental biology in conjunction with a chemistry minor. She has worked under the direction and supervision of Mike Horn for three years, along with other graduate students at California State University, Fullerton. This is Laura's first published work and is grateful to her research advisor, Dr. Horn and graduate student, Tyler Flisik, for their guidance. Financial support for this research was provided by California State University, Fullerton as a part of a Faculty-Student Research and Creative Activities Grant. Laura will be graduating from California State University Fullerton, in the spring of 2013.



ROBERT LEEPER is a senior BSc student majoring in geology at California State University, Fullerton (CSUF). He worked full-time for the United States Geological Survey (USGS) since 2007. While working at the USGS, he contributed to a variety of research projects such as seismic monitoring, post-fire debris flows, and paleotsunami investigations along the California coast. While attending CSUF, he conducted international research studying paleotsunami along the southern Thailand coast. Advancing his education while he continues to work to support his family is how he see his immediate future unfolding. After he completes the requirements for a BS in geology, he plans to apply to the Master's program at CSUF and hopefully expand on previous paleotsunami research he conducted in southern Thailand.



MIKHAIL POPOV is graduating in May with a B.A. in Mathematics (Probability & Statistics). He's an expert programmer and freelances as a web/graphic designer. Currently he's the developer of Greed: Make History, an upcoming turn-based strategy game for iPhone and iPad. Mikhail will be working on a Master's degree in Statistical Practice at Carnegie Mellon University this fall.



NICK BLACKFORD is a senior majoring in Mathematics, with a concentration in Pure Mathematics. He is currently working with Daniel Lenders and Danny Orton under the advisement of Dr. Scott Annin on Finite Group Theory. He plans on attending a Ph.D. Program in the Fall of 2012, and hopes to pursue a career in teaching. In his spare time, he enjoys spending time with friends from school, church, and high school. He is a small group leader at his church and enjoys advising students, along with his girlfriend, Katherine.

JONATHAN SAYRE is an undergraduate working on his degree in mathematics. He will graduate in the Spring of 2012 and will enter the credential program at California State University, Fullerton. He will also pursue a Masters in Mathematics Education to complete his training to be a high school math teacher. Jonathan is a returning student after a successful career as a field technician for computerized engraving machines. He is the father of three children and looks forward to being an inspirational high school math and science teacher.

EDITOR BIOGRAPHIES



AMBER SHAH (EDITOR-IN-CHIEF) is currently an undergraduate studying biology at California State University, Fullerton. She intends to graduate having achieved a Bachelor's of Science concentrating in Cell and Developmental Biology. She is a part of the University Honors Program and is intrigued by its many diverse disciplines. In addition, Amber has worked in Dr. Sacco's lab as a volunteer and wishes to further immerse herself in the field of research. Amber is also interested in pediatrics. After graduating from Cal State Fullerton, she will pursue the MS/OD dual degree at optometry school and fulfill her many aspirations within the field.



CHRISTOPHER BAKER (GEOLOGY) is an undergraduate student attending CSUF, with plans to attain a B.S. in geological sciences. Currently, he is working with Dr. Phil Armstrong on determining the exhumation rates of the southern Alaskan mountain ranges, and will be traveling to Alaska in the summer of 2012 to gather samples for fission track dating. After attaining a B.S. he plans to continue to graduate school to attain his M.S. and Ph.D. He transferred to CSUF in the fall of 2011 from Cypress and Fullerton Community Colleges, and currently is an officer in the South Coast Geological Society.



MANAR ADEM (BIOLOGY) is a Biological Sciences major with a concentration in Cell and Developmental Biology. She is the Biological Sciences section editor of New Dimensions Research Journal. She is in her last year at Cal State Fullerton and plans on pursuing a medical degree.



PETER HO (PHYSICS/MATH) is a senior majoring in physics and mathematics at California State University, Fullerton. His interests are in differential geometry and understanding the applications of geometry in physics. He will be attending graduate school in mathematics after graduating.



BRIAN HUEZO (COVER DESIGNER) is a Honduran born designer based in Fullerton, California. He is graduating with a Bachelors Degree in Fine Arts from the California State University in Fullerton.



COURTNEY KIM (LAYOUT EDITOR) is a designer based in Los Angeles, California, specializing in editorial and web design. She is graduating with a Bachelors Degree in Fine Arts from California State University, Fullerton and plans on pursuing a masters degree in the future. Her work can be viewed at courtneykimdesign.com.