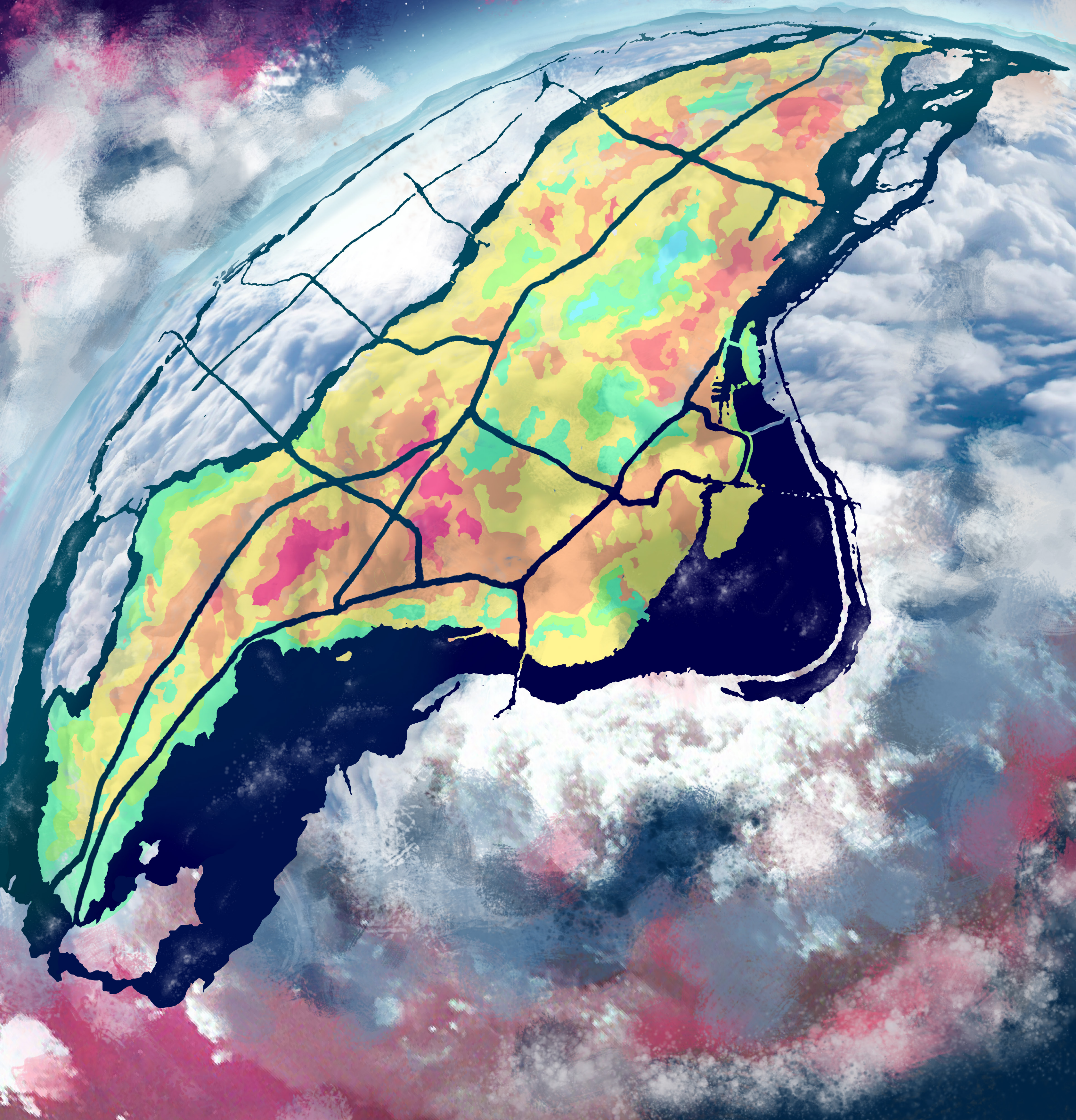
VOLUME 17 \ ISSUE 1 \ APRIL 2022

MSURJ

MCGILL SCIENCE UNDERGRADUATE
RESEARCH JOURNAL



ON THE COVER

From the glacial colds of Winter to the blistering heat of mid-July, a bird's eye view of Montreal's climate paints a picture of uncompromising yearly temperature shifts. The distribution of cities' heat islands has significant impacts on human health and well-being. On page 56, Aucoin, Briand, Duval, and Qudsi describe Montreal's micro-urban heat islands and discuss the impacts of vegetation coverage and urbanization, while on page 36, Frie, Gilmer, Buraga, and Franceschini measure the city's albedo, a property of the surface's ability to diffuse or reflect solar radiation.

The digital painting on the cover depicts a stylized Island of Montreal viewed from space, emphasizing our modeling of heat from large-scale data sets.

MCGILL SCIENCE
UNDERGRADUATE
RESEARCH JOURNAL

BOARD 2021-2022

Editors-in-Chief

John Ni
Biochemistry

Jenny Zheng
Neuroscience

Managing Editors

Hugh Deng
Microbiology and Immunology

David Derish
Medicine

Ryan Huang
Neuroscience

Sofia Reynoso
Biology

Muhammad Shahzad
Microbiology and Immunology

Senior Editors

Mina Mahdi
Biomedical Sciences

Laura Meng
Medicine

Aelis Spiller
Earth Systems Science

Aamir Wahhab
Biochemistry

Editors

Anthony Dumais
Cognitive Science

Ines Durant
Biology

Shafaq Nami
Bioengineering

Autumn Pereira
Biology and Mathematics

Nathan Pilkey
Microbiology and Immunology

Laura Reumont
Physics

Javeria Rizwan
Physics

Tiffany Spector
Biomedical and Life Sciences

Layout Designers

David Derish
Shafaq Nami
Sofia Reynoso
Aelis Spiller
Jenny Zheng

Graphic Designer

David Derish

805 Sherbrooke St. West, Room 1B22
Montreal, Quebec, H3A 2K6 Canada
Phone: (514) 398-6979
Fax: (514) 398-6766
Email: msurj.sus@mail.mcgill.ca
Website: msurj.com

TABLE OF CONTENTS

RESEARCH ARTICLES

REVIEW ARTICLES

FOREWORD

Dear Reader,

Every year, we have the honour of witnessing the products of undergraduate research across the world and the privilege of sharing these accomplishments with the wider community. The pandemic has brought countless challenges to undergraduate researchers these past two years, yet we continue to see the dedication and passion of undergraduates at McGill and beyond in engaging with the scientific process.

The twelve articles enclosed within this issue are the culmination of these students' hard work and perseverance. As such, we hope the 17th volume of the McGill Science Undergraduate Research Journal's presentation of their findings shines a spotlight on the diligence and research aptitude of these young researchers.

On behalf of our entire editorial board, thank you.

John Ni & Jenny Zheng
Editors-in-Chief

ACKNOWLEDGEMENTS

The McGill Science Undergraduate Research Journal would like to thank its generous contributors without whom this journal would not have been possible.

We thank the librarians of McGill University for their guidance. We thank the Office for Science Education for their continued support of both undergraduate research at McGill and the MSURJ.

We thank all the financial supporters in the McGill community for their generous support:

Faculty of Medicine and Health Sciences
Department of Biology
Department of Physiology
Department of Pharmacology & Therapeutics
Science Undergraduate Society of McGill University

Finally, we thank the peer reviewers, including Dr. David Bailey, Emilie Boutet, Dr. Ken Dewar, Dr. Carolina Dufour, Dr. Frédéric Dupont, Dr. Ursula Eicker, Timothy Jiang, Dr. Naomi Keena, Dr. Andrew Kirk, Hagit Malikin, John V. L. Nguyen, Dr. Elissa Penfound, Dr. Crystal Schaaf, Dr. Jan Sedláček, Dr. Steven Shaw, Dr. Iain Stewart, Dr. James Voogt, Dr. Nigel Waters, and the anonymous peer-reviewers. Their careful revision of the articles is critical to the success of the 17th issue of the MSURJ.

Submitted: 01/17/2022
Accepted: 03/19/2022

Research Article

Killian Abellon[1], Amelia Murphy[1], and Anika Anderson[1]

[1]Department of Geography, McGill University, Montreal, QC, Canada

Overblown? Analyzing Wind Speed in the Hurricane Warning Response System

Keywords

Hurricanes, Saffir-Simpson, warnng response system, recovery, resilience

Email Correspondence

killian.abellon@mail.mcgill.ca

Abstract

The role of wind speed in determining the impacts of hurricanes is examined via statistical analysis of Category 2-5 hurricanes that made landfall in the U.S. Atlantic basin coastline, including Puerto Rico's coast, from 1970-2020. The results indicate a positive yet statistically insignificant correlation between wind speed and hurricane deaths, cost of damages and federally obligated recovery aid. Other factors, such as storm surge, rainfall, and inland inundation, may be more strongly correlated with these impacts. The results are contextualized by a wealth of literature pointing to the role of social, political, and economic factors in determining the destructiveness of hurricanes. Finally, alternative indices to the popular Saffir-Simpson hurricane hazard scale – which relies on wind speed – are examined. As climate change advances and hurricanes become increasingly frequent and severe, more comprehensive hazard-rating scales may provide the basis for a more effective warning-response system, ultimately bolstering the resilience of coastal areas.

Introduction

Hurricane Katrina hit New Orleans in 2005 as a Category 3 hurricane, yet the residents had expected a Category 1. A Category 1 unexpectedly became a Category 5 in a matter of hours, then a Category 3 at landfall. New Orleans, a city used to hurricanes, was left unprepared[3]. These categories, found on the Saffir-Simpson Scale, are used to assess hurricane risk in the Western Hemisphere and communicate it to the public. However, when the scale fails to accurately predict or communicate risk, on top of other bureaucratic challenges, the public and authorities are ill-prepared. Katrina marked an important turning point in warning-response research and led to reforms in hurricane preparedness and response[4].

Almost 16 years post-Katrina, hurricane research has never been more relevant. As the climate changes due to a warming atmosphere, storms can hold more water, move more slowly and are becoming more frequent. In other words: storms are becoming increasingly damaging[5]. Since 1980, tropical cyclones in the US have caused $945.9 billion in total damages, averaging $21.5 billion per event.

2017 saw the highest costs ever from hurricanes, largely from Hurricane Harvey which made landfall in Houston, Texas. Additionally, of all weather-related disasters, hurricanes are responsible for the highest number of deaths (6,593 direct deaths since 1980)[6]. Although advancements in meteorology have greatly improved storm forecasting, as hurricanes become more damaging, it is critical that cities can anticipate and build resilience to hurricanes with accurate warning response systems[5]. Thus, the value of relying on wind speed in the Saffir-Simpson Scale comes under question.

Hurricane-related damages, costs, and deaths result primarily from water (rainfall, storm surge, etc.), though wind is the measure by which governments prepare for and respond to hurricanes through the Saffir-Simpson Scale (see Table 1). In fact, 88% of deaths are from water, not wind[7]. Furthermore, scholars such as Robert D. Bullard and Beverley Wright (2009) studied Hurricane Katrina at length and deemed it a "preventable catastrophe", driven by socio-economic factors such as discriminatory policy, poor land-use planning and failure of the warning-response bureau-organization[4,8].

Thus, two questions arise: What role does wind speed play in hurricane risk assessment? Should other factors be considered in this process? By providing a more holistic assessment of the factors driving hurricane damage and deaths, we can better anticipate vulnerabilities, build social and ecological resilience to extreme weather and align policy and financial responses with on-the-ground realities.

To develop this holistic assessment, two primary objectives are observed: [1] examine wind speed as a determinant of hurricane-related deaths, cost of damages and recovery aid for hurricanes that made landfall in the United States and Puerto Rico from 1970 to 2020, and [2] identify other factors, such as storm surge; rainfall; and social, economic, and political factors which may better predict deaths, damages, and recovery aid. The 3-part methodology involves building a database, conducting linear regression analysis, and contextually investigating the qualitative literature. This will facilitate analysis to determine if there are significant changes in deaths, damages, and recovery funding between Saffir-Simpson Categories (i.e. between different wind speeds).

We expect that wind speed as a hurricane categorization limits the effectiveness and resilience of the warning-response system, and hope to contribute to a system more capable of adaptation and emergency management planning.

Table 1. Saffir-Simpson Hurricane Scale[9]

Saffir-Simpson Hurricane Scale Category	**Maximum Sustained Wind Speed (kt)**	**Maximum Sustained Wind Speed (km/h)**
Tropical Depression	< 34	< 62
Tropical Storm	35-63	63-118
1	64-82	119-153
2	83-95	154-177
3	96-112	178-208
4	113-136	209-251
5	≥ 137	≥ 252

Methods

Firstly, we constructed a database containing wind speed, deaths, cost of damages, and federal recovery aid from the Atlantic Hurricane Database (NOAA), Historical Hurricane Tracks (NOAA/GIS), U.S. Census Data (GIS, U.S. Census Bureau, various sources), Public Assistance Funded Projects Details (FEMA), and the Emergency Management Events Database (EM-DAT)[10-15]. Our database includes hurricanes that made landfall in the United States (including Puerto Rico) from 1970-2020 at Category 2 or higher. The first step in building the database was to reference the Atlantic Hurricane Database to identify landfall events (marked by system indicator L) of Category 2 or higher and filter for the timeframe 1970-2020. Then, we used the Historical Hurricane Tracks GIS application, overlaid on a U.S. Census map for the decade the hurricane occurred in, to identify the counties (municipalities in Puerto Rico and parishes in Louisiana) intersected by the corresponding hurricane's track on the date of landfall. The U.S. Census data was then used to calculate the intersected counties' average population density.

To ensure data accuracy, hurricanes were divided by the decade in which they made landfall to refer to the corresponding U.S. Census decade. For example, hurricanes that made landfall in 1995-2005 rely on 2000 census data, whereas hurricanes that made landfall in 2005-2015 use 2010 census data. For hurricanes before 1995, there were no GIS maps with population data, so the intersected counties were identified using a map and the county population data were found in U.S. Census reports from 1990, 1980, and 1970.

From the Historical Hurricane Tracks GIS application, we downloaded the NOAA hurricane reports to identify the deaths and damages in the intersected counties or states. In the instances where county-level data were unavailable, NOAA state-wide data were used (county-level data were used 29 times and were unavailable 24 times). Of note is that the hurricane reports greatly differ in format depending on the year, and there is no centralized federal database containing this information. For example, deaths and damages data for the older hurricanes were in scanned reports, making it in some cases difficult to decipher the data.

Furthermore, we used EM-DAT's Consumer Price Index (CPI) to normalize NOAA damages to 2021 dollars. The EM-DAT dataset was built by selecting the categories "Tropical storm" under "Natural" disasters from 1900 to 2021 in the United States and Puerto Rico. The data were downloaded as an Excel document containing the event's name, time, number of deaths, number of injuries; an estimate of the damages; and a Consumer Price Index (CPI) with a base 100 in 2021. Notably, an event is created on EM-DAT if it has 10 or more deaths, 100 or more affected individuals or resulted in an international appeal for assistance. As a result, some events present on NOAA's Atlantic Hurricane Database were not on EM-DAT, restricting the number of observations available for our research. Due to these limitations, NOAA data were used for deaths and damages in the statistical analysis, while EM-DAT's Consumer Price Index was used to normalize the NOAA damages to 2021 dollars.

To identify the amount of federal recovery aid—the grant that the federal government is obligated to pay—we filtered the Public Assistance Funded Projects Details (FEMA) data by hurricane landfall date and intersected counties and summed the "federal share obligated." In instances where the hurricane emergency was not declared on the date of landfall, reports from 1-2 days post-landfall were used. However, the FEMA database extends from 1998 to the present, and data were missing for several hurricanes within that period. Hurricanes where FEMA recovery aid data were unavailable are marked by a period (.) in the database.

Damages and deaths were normalized using the average population density of the nearest decennial census for counties intersected by the hurricane track. They are hence represented in the database as deaths, or damages, per person per square mile. The recovery aid data is presented as a ratio of the obligated federal aid against the total project amount, to demonstrate the proportion of aid paid by the federal government. The total project amount is the total cost of recovery based on FEMA's damage survey.

Once the database was complete, we statistically analyzed it using the Stata application[16]. We conducted three primary statistical analyses regarding: [1] the relationship between hurricane wind speed and deaths as a ratio of population density, [2] hurricane wind speed and damages as a ratio of population density, and [3] hurricane wind speed and recovery funding as a proportion of damages.

For all analyses, we assumed authorities prepared for the maximum sustained wind speed at landfall since hurricanes decay over land[17]. The analysis (linear regression) entailed creating two-way scatter plots of maximum sustained wind speed at landfall and each dependent variable. Then, we added regression lines and overlaid vertical lines demarcating each Saffir-Simpson category. In general, for data where there were significant outliers, we conducted additional analysis, generating graphs and regression lines that excluded the outliers.

Additionally, for damages and deaths, because EM-DAT data had a country-level scope while NOAA data had a county or state-level scope, we could compare the deaths and damages that occurred in counties or states intersected by the hurricane track with country-wide deaths and damages. This comparison revealed that many of the hurricane-related deaths and damages actually occur outside counties intersected by the hurricane track; we analyze this in the Discussion section, though the comparison is outside the main scope of this paper. Furthermore, the EM-DAT database had only 42 data points that could be compared with the NOAA reports.

After conducting statistical analysis with Stata and identifying statistical trends (or lack thereof) between sustained maximum wind speeds and deaths, cost of damages or recovery aid, several patterns in the data warranted further investigation. As a result, we contextualized the data with a qualitative investigation of the literature on hurricanes and their social impacts. While gathering data and processing results, a set of questions arose regarding why some hurricanes of the same category had such different impacts even after accounting for population density and inflation. Such questions are partially answered in the Discussion section. However, further research is needed as this secondary research theme is outside the scope of our primary objectives.

Results

The Data

From 1970 to 2020, there were 53 landfall events at Category 2 or higher in the United States and Puerto Rico for 40 different hurricanes. Hurricane Georges (1998) had the highest number of landfall events (4) at Category 2 or higher, while 31 hurricanes only had one recorded landfall event at Category 2 or higher. In some years, there were 0 Category 2-5 recorded landfall events in the United States and Puerto Rico.

Often, we found that this was because hurricanes made landfall in other countries in the Atlantic Basin, such as Cuba or the Bahamas, and reached the United States only as a tropical depression. There was an average of 2.3 landfall events per year, a median of 2, and a mode of 1. It is noteworthy that observations of four and five landfall events per year only occur post-1996, which may indicate that hurricanes made landfall outside of the U.S. in the past, or point to the increasing frequency of Category 2 and higher hurricanes as the climate warms.

There were only 2 observations in Category 5, 9 in Category 4, and 21 each in Categories 2 and 3. The Category 5 landfall events were Hurricane Andrew (1998), which made landfall at Category 5 in Miami-Dade County, Florida; and Hurricane Michael (2018), which made landfall at Category 5 in Florida and maintained Category 5 wind speeds into Georgia.

The average maximum sustained wind speed at landfall of all observations is 102.55 kt or 189.9 km/h (equivalent to a Category 3), with a standard deviation of 15.40 kt or 28.5 km/h.

Hurricane path/track: We refer to 'best track' in the Historical Hurricane Tracks GIS application, which is a "representation of a tropical cyclone's location and intensity over its lifetime"[1,2].

Landfall: We refer to "the intersection of the surface center of a tropical cyclone with a coastline"[1].

Maximum Sustained Wind Speed at Landfall and NOAA Deaths Normalized

Figure 1 shows a regression plot of wind speed at landfall (x-axis) and normalized NOAA deaths (y-axis). Red lines are thresholds between Saffir-Simpson categories. 53 observations were computed. Among them, 2 are largely outside the 95% confidence interval of the fitted values computed by Stata. Both outliers are Katrina landfall events (11.80 and 6.32). The coefficient of the fitted values' regression line is β=0.0084779.

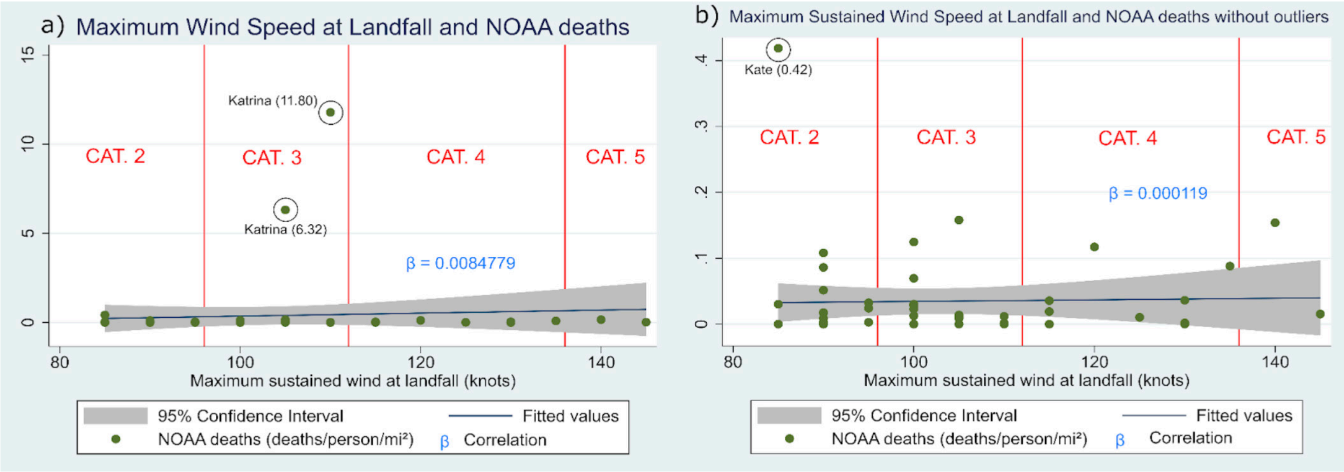


Figure 1. Maximum Sustained Wind Speed at Landfall vs NOAA Deaths Normalized. Figure 1a includes outliers while Figure 1b excludes outliers. The unit of the x-axis is knots, which is commonly used by atmospheric researchers and is equal to 1.852 kilometers per hour. The unit of the y-axis is deaths per average population density of intersected counties. Each point is an observation of a hurricane that made landfall.

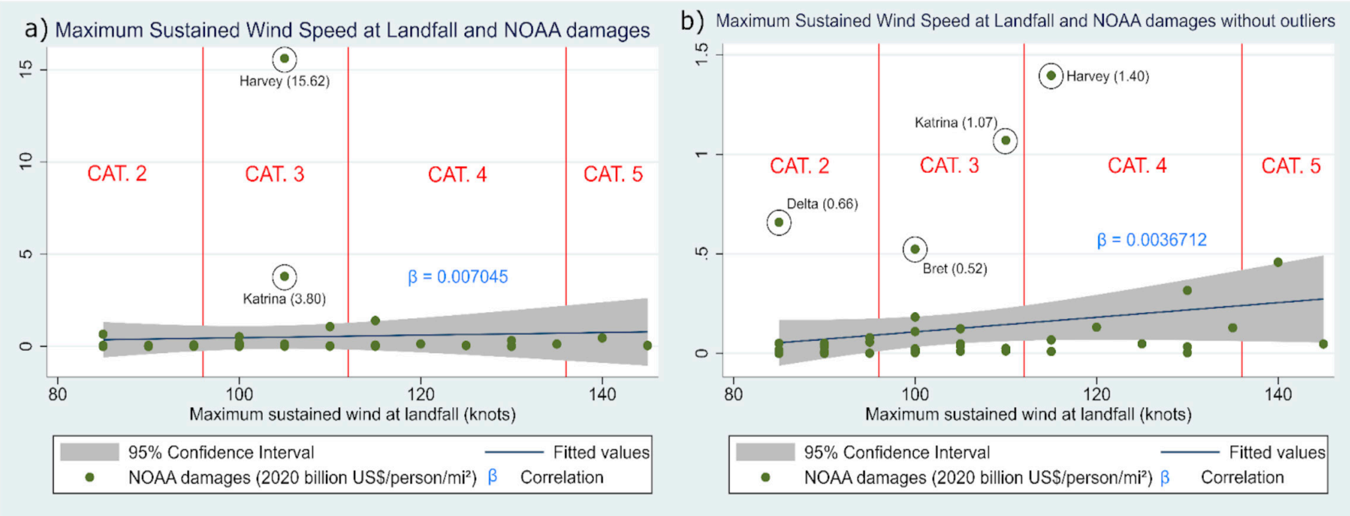


Figure 2. Maximum Sustained Wind Speed at Landfall vs NOAA Damages Normalized. Figure 2a includes outliers while Figure 2b excludes outliers. The unit of the x-axis is the same as in the previous regression plot. For the y-axis, the unit is billions of US dollars per average population density of intersected counties.

To analyze trends, especially between Saffir-Simpson categories, a new graph was computed by removing the two outliers. 21 observations or 41.1% of total observations are outside the 95% confidence interval of fitted values computed by Stata. The top left outlier is from Hurricane Kate (0.42) The coefficient of the fitted values' regression line is β=0.000119.

Maximum Sustained Wind Speed at Landfall and NOAA Damages Normalized

Figure 2 shows a regression plot of Maximum Sustained Wind Speed at Landfall (x-axis) and standardized NOAA Cost of Damages (y-axis). 52 observations were computed. Among them, 1 is an outlier and does not allow the analysis of other observations. 2 other observations seem to be outside the 95% confidence interval on the fitted values computed by Stata. The largest outlier is Hurricane Harvey (15.62), the other being Hurricane Katrina (3.80). The coefficient of the fitted values' regression line is β=0.007045.

A new graph was computed to remove the outliers (Hurricane Harvey and Hurricane Katrina landfall events). In this new graph, 4 hurricanes were above the 95% confidence interval. All were Category 2, 3 or 4 events. Some events were below the 95% confidence interval but closer than the previous 4. The 4 points far above the confidence interval are from left to right, Hurricane Delta (0.66), Hurricane Bret (0.52), Hurricane Katrina (1.07), and Hurricane Harvey (1.40). Interestingly, these outliers are in Category 2-4 hurricanes, with none in Category 5, perhaps suggesting that their lower Category designation caused an underestimation of the risks of the hurricane that resulted in the high damages. The coefficient of the fitted values' regression line is β=0.0036712.

Maximum Sustained Wind Speed at Landfall and FEMA Recovery Aid Normalized

Figure 3 has maximum sustained wind speed at landfall (x-axis) and normalized recovery aid from FEMA (y-axis). 18 observations were computed, all between a ratio of 0.75 and 1.00. At least seven observations are outside the 95% confidence interval of fitted values. There were no extreme outliers, and with only 18 observations, only one graph was generated. The coefficient of the fitted values' regression line is β=0.0000424.

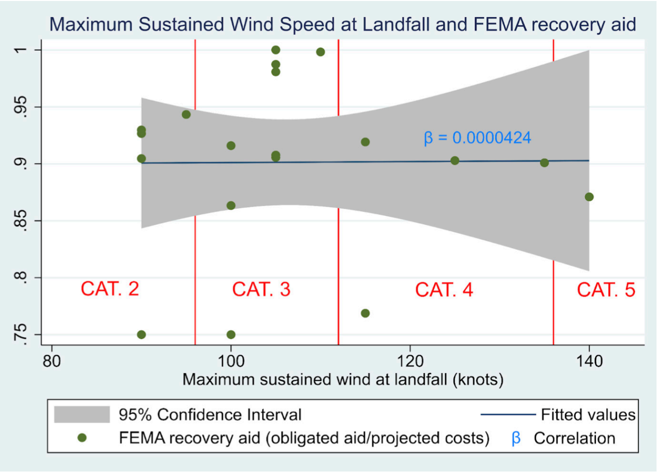


Figure 3. Maximum Sustained Wind Speed at Landfall vs. FEMA Recovery Aid Normalized. The unit of the x-axis is knots while the y-axis is a ratio of federally-obligated aid over the total project amount.

Discussion

Interpreting the Statistics

For each regression, the R-squared values, coefficients, and P values provide insight into the relationship between wind and hurricane outcome variables (see Table 2). The R-squared values suggest that the independent variable (maximum sustained wind speed at landfall) cannot explain the variation in the dependent variables (deaths, cost of damages, recovery aid). For the relationship between wind speed and deaths including outliers, the R-squared of 0.0052 was greater than the R-squared excluding outliers (0.0007). For the relationship between wind speed and cost of damages, the opposite was observed: R-squared for the relationship including outliers (0.0024) was less than the R-squared for the relationship excluding outliers (0.0457). The lower R-squared for the relationship between wind speed and deaths including outliers may point to the fact that in the case of outliers, wind speed variance explains less variance in outcomes than in hurricanes within the 95% confidence interval. The R-squared value for the relationship between wind speed and recovery aid was 0.0001.

Table 2: Regression Output Statistics

Analysis	**Correlation (β)**	**R-Squared**
Deaths (deaths/person/mi^2)	0.0084779*	0.0052
Deaths without outliers	0.00012*	0.0007
Damages (2020 billion US$/person/$mi^2$)	0.007045*	0.0024
Damages without outliers	0.0036712**	0.0457
Recovery (obligated aid/projected costs)	0.0000424*	0.0001

Level of significance: $* = 1 \geq p \geq 0.5$, $** = 0.5 > p \geq 0.1$

Low coefficient values (under 0.01) for each variable further suggest that there is a positive but weak correlation between wind speed and deaths and damages. This may in part be explained by the fact that some hurricanes weaken faster than others and thus correlation for wind speed at landfall may be lower than correlation to surface wind speed for all non-landfall counties. However, this analysis used wind speed at landfall to reflect the Saffir-Simpson scale, which is how counties assess hurricane intensity. For recovery aid, the correlation was so small that it can be considered flat (0.00004). P values (all ≥ 0.1) from deaths, damages, and recovery aid analysis suggest that the findings are not statistically significant, indicating strong evidence for the null hypothesis – in this case, that there is little relationship between the variables studied. We suspect that this is in part due to the limited sample size (n = 53), compared to the total number of hurricanes that occurred from 1970-2020. However, this may also suggest that using wind speed as the determining factor for hurricane risk does not reflect the real extent of the correlation between wind speed and hurricane impacts.

Spatial Variation

During data collection, we observed that many of the deaths caused by the hurricanes occurred outside intersected counties or states. This is demonstrated by the large disparity between the NOAA deaths, which are specific to county/state, and the EM-DAT deaths, which are country-wide. The disparity is visible in Fig. 4, which plots EM-DAT deaths against NOAA deaths, with a 1:1 ratio notated by the line x=y. Because EM-DAT has fewer recorded hurricane events than the NOAA database, this comparison has a somewhat smaller sample size than our main analysis, at 42 data points compared to 53.

Many of the external deaths were due to flooding caused by storm surge or heavy rainfall, following the trend that most hurricane deaths are caused by water (i.e. flooding and offshore deaths), in line with previous research which found that 80% of U.S. hurricane deaths occur in non-landfall counties[7,18]. Since the bulk of fatalities occur due to water, the hurricane does not have to directly hit an area for it to cause deaths. Hurricanes cause heavy rainfall across a large area, leading to flooding even if the area is not experiencing high winds, and the wind can kick up waves that reach a long distance to cause storm surges. The National Hurricane Center lists heavy rainfall and inland flooding, storm surge, rip currents, and tornadoes as the primary hazards of hurricanes, along with high winds[19]. While these factors are correlated with wind speed, Irish et al. (2008) demonstrate that for a given Saffir-Simpson intensity, storm surge can vary as much as 30%[20]. Between our results and existing research that shows the role of water in deaths from hurricanes, a hurricane scale that includes rainfall predictions may be more effective for hurricane warning and preparedness systems. Risk-prediction models that account for various hazards, such as wind and rain, storm surge, and freshwater inundation have been proposed by authors such as Baradaranshoraka et al. (2017)[21].

Outliers

Both NOAA deaths and damages had outliers that were two orders of magnitude higher than most other values, notably Hurricane Katrina and Hurricane Harvey. Neither hurricane was a Category 5, making landfall at Category 3 and 4 respectively. The magnitude of differences between the fitted line and observed deaths and damages, particularly in outliers, points to the critical human/social/political influence on storm outcomes. Much literature has analyzed the role of socioeconomic and demographic factors in determining hurricane impact.

For example, Parker et al. (2009) point to the psychological, organizational, and political factors that preconditioned the Katrina catastrophe[4]. Bullard (2007) additionally illustrates the "racial divide in the way the U.S. government responds to natural and man-made disasters, such as hurricanes and floods, and public health threats," and Bullard and Wright (2009) detail "the role of race and place and how unequal protection and unequal treatment make some populations more vulnerable in the rebuilding and recovery process"[8,22]. Addressing the role that social, political, and economic factors play in determining the outcomes of hurricanes will likely become increasingly important as climate change develops, repeating existing patterns of inequality and environmental injustices[23].

Alternative Scales

Although the scope of our research was limited, there is significant qualitative research on the importance of scale in perceiving and communicating hurricane risk[24,25]. Given the likelihood of increasingly frequent 'outlier' events as climate change worsens, a scale that does not correlate strongly to deaths and damages may not be an adequate tool for communicating risk. Kantha (2006, 2013) examines how the Saffir-Simpson Scale "was devised principally to predict the expected intensity of hurricane wind damage to structures," and while it has served this purpose, Kantha proposes a more comprehensive Hurricane Intensity Index used in conjunction with a Hurricane Hazard Index which captures the risk posed by storm surge, to prepare prior to landfall and for relief efforts, respectively[26,27]. Senkbeil and Sheridan (2006), noting that the Saffir-Simpson scale fails to accurately "account for observed impacts over land," also propose a Hurricane Classification System which also accounts for surges[28]. More recently, Rezapour and Baldock (2014) propose a hurricane hazard index that demonstrates a stronger correlation to deaths and damages for recent hurricanes (2003-2012) and includes a rainfall subindex[29], while Klotzbach et al. (2020) propose that indices account for surface pressure as well[30].

Nonetheless, this proliferation of new indices might undermine their acceptance and understanding by target users. Indeed, while the push for research in this area might develop powerful risk assessment tools, public authorities and citizens might become confused about which one to rely on, ultimately undermining hurricane hazard mitigation strategies. Lastly, residents of areas where hurricanes strike are deeply attached to their homes, and the decision to leave in the face of a hurricane is rarely solely a calculation of risk[31]. Future research that controls for how many people evacuate an area rather than total population density may provide deeper insight into the potential risk of hurricanes and success of emergency management planning. Further, while the adoption of a new scale may better capture hurricane risk, policies such as hurricane scales often take years to change, and more immediate strategies to improve the resilience of coastal areas are critical in the meantime.

Conclusion

In addition to accounting for more recent hurricanes, our research aligns with existing literature to conclude that the Saffir-Simpson Scale does not strongly correlate with hurricane impacts, namely the cost of damages incurred, deaths, and recovery aid. Discrepancies in the deaths and damages in counties intersected by hurricanes tracks compared to country-wide deaths and damages point to the importance of factors other than wind speed, such as storm surge or rainfall. Observed outliers further suggest the importance of social, political, and economic factors in determining hurricane impact. Given the likelihood that hurricanes will become more frequent and severe due to climate change, our research highlights the need for more comprehensive hurricane risk calculation and communication to minimize their impacts.

Additionally, our research has identified several areas for future research or action. Notably, the lack of a centralized federal database recording hurricane characteristics, deaths, cost of damages, federally obligated recovery aid, and emergency evacuation reports may present a barrier to future research. Constructing such a database could facilitate a better understanding of hurricanes and in turn improve hurricane preparation and response by federal agencies such as NOAA and FEMA. While we identified discrepancies in deaths and damages that occurred in counties or states intersected by hurricane tracks versus country-wide deaths and damages, another critical area of future research may include comparisons of indirect and direct deaths. For example, Rappaport & Blanchard (2016) found that since 1995, the ratio of indirect deaths to direct has increased from 1:2 to 2:1, partly due to differences in how deaths are categorized, but also potentially due to changes in electricity infrastructure and evacuation procedures[32]. The causes of indirect and direct deaths in intersected counties and counties outside the immediate hurricane track may be an area where research could be life-saving. Finally, while literature has already suggested new indices for evaluating hurricane risk and hazard, further research on combining social, political and economic vulnerability indices with storm surge, rainfall, and inland inundation risk indices will be critical in improving the resilience of coastal areas.

Acknowledgements

This project was a part of GEOG 460: Research in Sustainability. We would like to thank first and foremost our professor, Dr. Brian Robinson, for his mentorship and guidance. Additionally, we thank our classmates for their continued support and feedback.

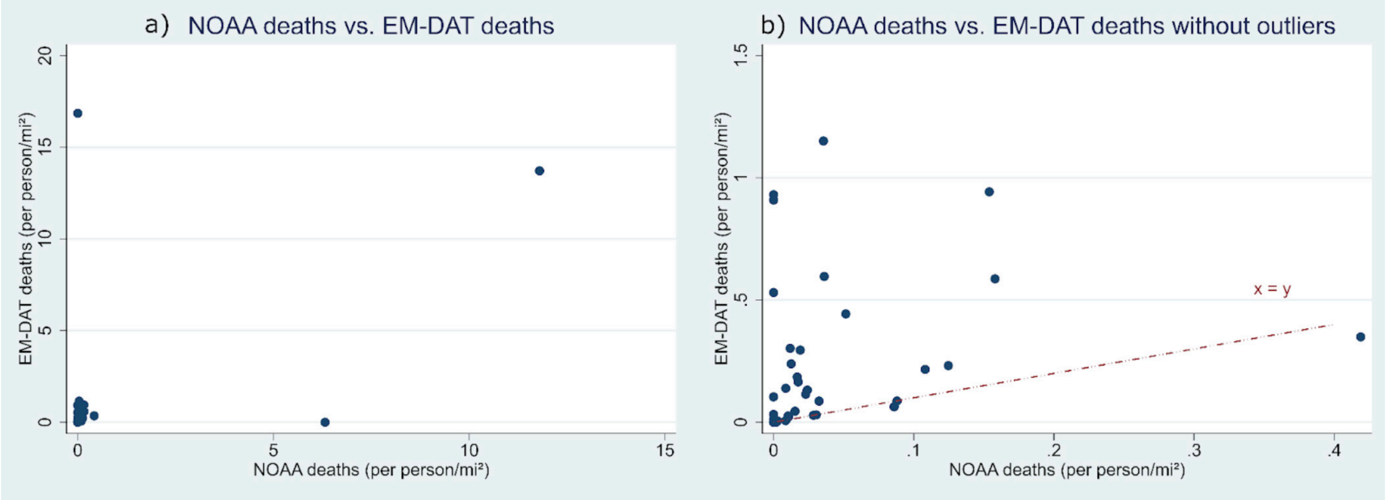


Figure 4. NOAA Deaths vs. EM-DAT Deaths. Figure 4a includes outliers while Figure 4b excludes outliers. The x-axis is NOAA deaths per average population density of intersected counties, and the y-axis is EM-DAT deaths per average population density of intersected counties.

References

1. National Oceanic and Atmospheric Administration. Glossary of NHC Terms. National Hurricane Center and Central Pacific Hurricane Center https://www.nhc.noaa.gov/aboutgloss.shtml.

2. ESRI & National Oceanic and Atmospheric Administration. Historical Hurricanes Tracks 1842-2020. https://www.arcgis.com/home/item.html?id=d053e72aabfd4c5ab4139c3829c1e11c.

3. Evans, M. M. & Merrett, R. Hurricane Katrina 15 Years Later: 10 Survivors on the Storm's Impact | PEOPLE.com. People https://people.com/human-interest/hurricane-katrina-survivors-15th-anniversary/ (2020).

4. Parker, C. F., Stern, E. K., Paglia, E. & Brown, C. Preventable Catastrophe? The Hurricane Katrina Disaster Revisited. J. Contingencies Crisis Manag. 17, 206–220 (2009).

5. Verbeten, E. & Wisconsin-Madison, U. of. Trends in hurricane behavior show stronger, slower and farther-reaching storms. https://phys.org/news/2020-10-trends-hurricane-behavior-stronger-slower.html.

6. Hurricane Costs. https://coast.noaa.gov/states/fast-facts/hurricane-costs.html (2021).

7. Erdman, J. 88% of U.S. Deaths From Hurricanes, Tropical Storms Are From Water, Not Wind | The Weather Channel - Articles from The Weather Channel | weather.com. The Weather Channel https://weather.com/safety/hurricane/news/hurricanes-tropical-storms-us-deaths-surge-flooding.

8. D. Bullard, R. & Wright, B. Race, Place, and Environmental Justice after Hurricane Katrina: Struggles to Reclaim, Rebuild, and Revitalize New Orleans and the Gulf Coast. (Taylor & Francis Group, 2009).

9. Saffir-Simpson Hurricane Wind Scale. https://www.nhc.noaa.gov/aboutsshws.php.

10. Emergency Management Events Database & Center for Research on the Epidemiology of Disasters. Custom request. EM-DAT https://public.emdat.be/ (2021).

11. ESRI & United States Census Bureau. 2020 USA Population Density. https://www.arcgis.com/home/item.html?id=b9095ebdf5e8442588ab3f269dc7ee5e (2021).

12. Federal Emergency Management Agency. OpenFEMA Dataset: Public Assistance Funded Project Details- v1. https://www.fema.gov/openfema-data-page/public-assistance-funded-projects-details-v.

13. National Oceanic and Atmospheric Administration. Historical Hurricane Tracks. https://coast.noaa.gov/hurricanes/#map=2/53.27/-18.05&search=eyJzZWFyY2hTdHJpbmciOiJOb3J0aCBBdGxhbnRpYyBPY2VhbiBCYXNpbiIsInNlYXJjaFR5cGUiOiJiYXNpbiIsImNhdGVnb3JpZXMiOlsiSDUiXSwieWVhcnMiOltdLCJtb250aHMiOltdLCJlbnNvIjpbXSwicHJlc3N1cmUiOnsicmFuZ2UiOlswLDExNTBdLCJpbmNsdWRlVW5rbm93blByZXNzdXJlIjp0cnVlfSwic2VsZWN0ZWRTdG9ybUlEIjoiMTkzMjI0NE4xOTI5NiIsImJ1ZmZlclVuaXQiOlsiTWlsZXMiXSwic29ydFNlbGVjdGlvbiI6eyJ2YWx1ZSI6InllYXJzX25ld2VzdCIsImxhYmVsIjoiWWVhciAoTmV3ZXN0KSJ9LCJhcHBseVRvQU9JIjpmYWxzZSwiaXNTdG9ybUxhYmVsc1Zpc2libGUiOnRydWV9.

14. National Oceanic and Atmospheric Administration. National Hurricane Center Data Archive. https://www.nhc.noaa.gov/data/?fbclid=IwAR2aQY4iXnq2fc7uJLkEMxqJhCE5d-qYfz3zhu4FLzw6a8EaB_VoGTBjabM#tcr.

15. National Oceanic and Atmospheric Administration. Storm Events Database | National Centers for Environmental Information. https://www.ncdc.noaa.gov/stormevents/ (2021).

16. Why Stata. Stata https://www.stata.com/why-use-stata/ (2022).

17. University of Rhode Island Graduate School of Oceanography. Hurricane Decay: Demise of a Hurricane. Hurricanes: Science and Society http://www.hurricanescience.org/science/science/hurricanedecay/#:~:text=The%20roughness%20of%20the%20land,in%20the%20first%2024%20hours (2020).

18. Czajkowski, J., Simmons, K. & Sutter, D. An analysis of coastal and inland fatalities in landfalling US hurricanes. Nat. Hazards 59, 1513–1531 (2011).

19. Hurricane Preparedness - Hazards. https://www.nhc.noaa.gov/prepare/hazards.php.

20. Irish, J. L., Resio, D. T. & Ratcliff, J. J. The Influence of Storm Size on Hurricane Surge. J. Phys. Oceanogr. 38, 2003–2013 (2008).

21. Baradaranshoraka, M., Pinelli, J.-P., Gurley, K., Peng, X. & Zhao, M. Hurricane Wind versus Storm Surge Damage in the Context of a Risk Prediction Model. J. Struct. Eng. 143, 04017103 (2017).

22. Bullard, R. D. Equity, unnatural man-made disasters, and race: why environmental justice matters. in Equity and the Environment (eds. C. Wilkinson, R. & R. Freudenburg, W.) vol. 15 51–85 (Emerald Group Publishing Limited, 2007).

23. García-López, G. A. The Multiple Layers of Environmental Injustice in Contexts of (Un)natural Disasters: The Case of Puerto Rico Post-Hurricane Maria. Environ. Justice 11, 101–108 (2018).

24. Hauser, D. J. & Fleming, M. E. Mother Nature's Fury: Antagonist Metaphors for Natural Disasters Increase Forecasts of Their Severity and Encourage Evacuation. Sci. Commun. 43, 570–596 (2021).

25. Tierney, K., Bevc, C. & Kuligowski, E. Metaphors Matter: Disaster Myths, Media Frames, and Their Consequences in Hurricane Katrina. Ann. Am. Acad. Pol. Soc. Sci. 604, 57–81 (2006).

26. Kantha, L. Time to replace the Saffir-Simpson hurricane scale? Eos Trans. Am. Geophys. Union 87, 3–6 (2006).

27. Kantha, L. Classification of hurricanes: Lessons from Katrina, Ike, Irene, Isaac and Sandy. Ocean Eng. 70, 124–128 (2013).

28. Senkbeil, J. C. & Sheridan, S. C. A Postlandfall Hurricane Classification System for the United States. J. Coast. Res. 22, 1025–1034 (2006).

29. Rezapour, M. & Baldock, T. E. Classification of Hurricane Hazards: The Importance of Rainfall. Weather Forecast. 29, 1319–1331 (2014).

30. Klotzbach, P. J. et al. Surface Pressure a More Skillful Predictor of Normalized Hurricane Damage than Maximum Sustained Wind. Bull. Am. Meteorol. Soc. 101, E830–E846 (2020).

31. Heglar, M. A. Before the Storm: How Do You Know When to Go? (2021).

32. Rappaport, E. N. & Blanchard, B. W. Fatalities in the United States Indirectly Associated with Atlantic Tropical Cyclones. Bull. Am. Meteorol. Soc. 97, 1139–1148 (2016).

Submitted: 01/24/2022
Accepted: 04/01/2022

Research Article

Samuel Aucoin[1], Bruno Tremblay[1,2], Robert Newton[2]

Coastal Sea-Ice Break-Up Events in Beringia

[1] Department of Atmospheric and Oceanic Sciences, McGill University, Montreal, QC, Canada

[2] Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

Abstract

We quantify changes in break-up events of landfast ice in the transition from a perennial to a seasonal sea ice cover in the Arctic. A break-up event is defined as a time when coastal sea ice concentration drops below 95% after a minimum period of 10 days of stable ice conditions. To this end we analyze output diagnostics from the Community Earth System Model (Version 1) – Large Ensemble from 1920 to 2080, focusing on six coastal communities of Alaska, Chukotka, and the Kamtchatka Peninsula: Utqiaġvik, Point Hope, Gambell, Novoye Chaplino, Sireniki, and Pakhachi. Model results generally agree with the satellite record with open water formation along the coastline associated with sustained offshore winds, although the sensitivity of CESM1-LE is higher than that of observations due to the absence of a landfast ice parameterization in CESM1-LE. Specifically, we see a linear relationship between the magnitude of the opening and offshore surface wind stresses integrated over the 10 days prior to the opening event, (p-value < 0.01). While the break-up event frequency increases (5.53×10^{-5} events/day/year for Utqiagvik) in the 21st century due to the thinning, or weakening, of the landfast ice cover, the total number of winter break-up events decreases due to a shortening of the winter season (mean of -5.3 days/decade).

Keywords

Arctic, sea ice dynamics, break-up events, climate change.

Email Correspondence

samuel.aucoin@mail.mcgill.ca

Introduction

The Arctic is home to approximately 4 million inhabitants[1], spread out across seven countries. Landfast ice, defined as sea ice that is fastened to the shore or ocean floor, is an important part of Arctic coastal ecosystems[2]. Arctic communities often depend upon local landfast ice for daily activities such as hunting and fishing[3], transportation, and recreational activities and thus have an intimate relationship with sea ice. While large-scale mean changes in Arctic sea ice have been studied extensively (e.g.[4–8] etc.), of more immediate use and concern to local peoples[9] are small-scale spatial and temporal changes, which have received less attention. Understanding the changes in the stability and predictability of landfast ice for instance[10]. The Arctic has had perennial ice cover since at least the mid Holocene Climactic Optimum[11], and is expected to transition to seasonal cover before the mid 21st century[4,5]. It is therefore important to better understand this transition, which is likely too rapid for local communities to adapt to without severe social and economic stress.

Despite its name, landfast ice is very much a dynamic system, and understanding those dynamics is key to ensuring the safety and well-being of those who depend on it. In this study, we investigate break-up events[10,12–14], which is the phenomenon where a portion of landfast ice breaks away from the coast during winter, forming an area of open water, followed by the closing of the pack ice several days later. These events are important to understand because they can create unsafe conditions for residents engaging in activities on the landfast ice, including becoming stranded on ice that has become a drifting floe. There has been some previous work investigating break-up events, sometimes named breakout events, but much of this work has either been focused on spring break-up rather than mid-winter events (e.g.[15,16]), or single events[10,12–14], and mostly using a qualitative definition. While these studies characterized the general mechanisms behind break-up events such as wind forcing, ocean currents, and sea level change among other things, the overall understanding is by no means comprehensive. There has not been a study with the same goals as ours: exploring the use of models to study break-up events, and investigating future trends.

Few communities have local landfast ice monitoring systems in place (e.g. Utqiagvik, Alaska[10]), but such monitoring systems do not provide information on future events or trends. In this study we look at decadal projections in the number of break-up events in coastal communities of Beringia (see Fig. 1), and at physical mechanisms responsible for break-up events. Models are an invaluable tool in climate science since they allow us to predict future trends, but some argue that the results involving landfast ice should be interpreted with caution[17,18] as landfast ice is difficult to model. However, since ground and satellite data on landfast ice is severely limited in terms of both spatial and temporal scale, the use of models is important to understand this phenomenon.

To this end, we analyse changes in simulated sea ice concentrations from the Community Earth System Model (Version 1) – Large Ensemble (CESM1-LE) and passive microwave observations.

Data

Passive Microwave Data

We use daily sea ice concentration (SIC) data from the NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration (CDR), Version 4[19]. The CDR database is based on the SICs from both the NASA Team algorithm[20], and the NASA Bootstrap algorithm[21]. The database uses the NSIDC Polar Stereographic projection, which has a spatial resolution of 25 km × 25 km, and provides nearly continuous daily SIC estimates from October 25, 1978 to December 31, 2020. In the following, we omit 1978 and 2020 to have only complete years, and to be temporally consistent with ERA5 winds. Specific grid cells adjacent to the coast are chosen for analysis. We also use daily 10 m wind components from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5th Generation (ERA5)[22]. ERA5 reanalysis, published by the ECMWF, provides hourly estimates of many Earth system components. We use 10 m wind components from the ECMWF ERA5 reanalysis interpolated on the Equal-Area Scalable Earth (EASE) Grid, at a spatial resolution of 25 km × 25 km. Grid cells located nearest to those chosen from the CDR dataset are used for analysis.

CESM1–LE

We use daily sea ice concentration, radiative surface temperature, and surface wind diagnostics from the Community Earth System Model Large Ensemble (CESM1-LE)[23] . CESM1-LE uses a rotated grid that puts the North Pole over Greenland with a nominal 1-degree latitude and longitude spatial resolution for the ocean and ice components. The atmospheric grid uses a standard spherical polar grid, also with a nominal 1-degree latitude and longitude spatial resolution. In this latitude band, 1-degree equates to approximately 50-100 km, which is large compared to a typical landfast ice extent in this region of approximately 20 km[24], but still comparable. CESM1-LE is comprised of 5 individual components (atmosphere, ocean, land, sea ice, runoff), and a coupler to transfer the variables between the atmospheric and ocean/ice components. The model does not resolve, nor does it have a parameterization of, landfast ice, which makes the ice more susceptible to dynamical forcings, and is therefore less representative of reality. CESM1-LE has 40 ensemble members, each starting with slightly differing initial atmospheric conditions. For this study, only one Ensemble Member (EM2) is analyzed, from January 1, 1920, to December 31, 2080. The results are robust to the exact choice of member. All ensemble members are forced with historical carbon dioxide (CO2) emissions up to 2005, then follow the Representative Concentration Pathway 8.5 (RCP8.5) emissions scenario to the end of the simulation. RCP8.5 represents the scenario where current trends in population and economic growth continue in an absence of climate change policies, and has the highest greenhouse gas emissions of the RCPs[25].

Methodology

The analysis is conducted for communities in the Beringia region which are study sites of the Arctic Robust Communities Navigating Adaption to Variability (ARCNAV) project (see Fig. 1). These include three communities in Alaska (USA) and three communities on the Kamchatka Peninsula (Russia): Utqiagvik (AK), Point Hope (AK), Gambell (AK), Novoye Chaplino (RU), Sireniki (RU), and Pakhachi (RU).

We identify the closest grid cell of each data set to each community's location to complete our landfast ice analysis (see Fig. 1). The grid cells from each of the CESM1-LE sea ice and atmosphere components are chosen to most closely match each other geographically, and likewise for the CDR and ERA5 datasets. Each grid cell is chosen based on its latitudinal and longitudinal proximity to its corresponding community in combination with minimal overlap with the land.

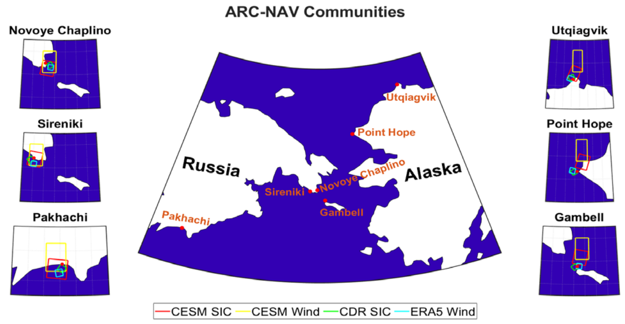


Figure 1. Map of Beringia (area comprising Alaska and Eastern Siberia) and locations of each key Arctic Robust Communities Navigating Adaption to Variability (ARCNAV) community: Utqiaġvik, Point Hope, Novoye Chaplino, Sireniki, Gambell, Pakhachi (red dots) with the closest grid cells for each data set used for analysis; CESM1-SEAICE (red), CESM1-ATMOSPHERE (yellow), CDR (green), and ERA5 (blue).

We define the start of the freeze season as the first day of a series of 10 days with a daily mean greater than 95% SIC and the start of the melt season as the first day of a series of 30 consecutive days with SIC below 95%. For some southern communities, these conditions are no longer met later in the 21st century, and the length of the winter season is set to zero. We define a break-up event as a period of at least 2 days with a mean SIC below 95%, following a period of 10 days with a mean SIC above 95%. We consider a 10-day period with a mean SIC $\geq$95% as a stable ice cover. Since a break-up event can last multiple days using this definition (up until the stability condition is no longer met), the date on which the SIC of the event reaches a minimum is used for date-specific analysis. We choose a 95% SIC threshold for a break-up event because higher thresholds will trigger too many false events due to noise in the signal, and lower thresholds will fail to capture smaller-magnitude events. The exact number chosen is somewhat arbitrary, but thresholds in the range of 90%-98% show similar results (not shown). A different threshold may impact the number and average magnitude of events, but likely would not have a major effect on the qualitative properties of the observed trends. We therefore choose 95% SIC as our threshold, understanding that the trends are more important than exact values in this analysis.

Results and Discussion

Historical Record

We describe the general behaviour of the landfast ice cover and break-up events for a typical year in the mid 20th century in Utqiagvik (Fig. 2). The other communities follow the same pattern with the exception of Pakhachi, where winter sea ice cover rarely reaches stability. A qualitative analysis shows that most break-up events are preceded by a large spike in both wind speed perpendicular to the coastline, (meridional in Utqiagvik's case) and temperature.

While temperature is not explored further due to relative unimportance compared to wind forcing and lack of observational data, the relationship observed in Fig. 2 is explained by further analysis; Both CESM1-LE and CDR observations indicate that they often last a period of 3-10 days (see Fig. 3). The magnitude of a break-up event is weakly correlated with the 4-day integrated off-shore wind (max. $r^2 \approx 0.3$)(see Fig. 4) and the three to ten day timescale of break-up events points toward synoptic-scale (O ~ 1000 km) weather phenomena, such as storms, that bring strong winds and relatively warm temperatures from the south. We infer that atmospheric forcings, specifically storms situated in a geographical position to apply off-shore wind stress, play an important role in causing and modulating the strength of break-up events.

The model and satellite observations, however, do not always agree on the nature of the correlation (see Fig. 4). For instance, meridional (southerly) and zonal (easterly) wind anomalies have the largest impact on the magnitude of break-up events in Utqiagvik in the model and observations, respectively. In CESM1-LE, the sandspit on which Utqiagvik is located is not resolved, and the hamlet faces directly north to the Arctic Ocean, whereas in reality it faces north-west. All other communities considered, except Pakhachi, are located on geographical features (e.g. a peninsula) that are not resolved in the model For example, Gambell is on an island that does not exist in CESM1-LE (i.e. Gambell is in open water instead of on land). In observations (Fig. 4), we find a wide range of zonal wind velocities for SIC $\approx$ 95% (i.e. near-stable landfast ice), reflecting differences in the internal ice stress and the presence of anchor points on the ocean floor on the shallow shelf[12]. In CESM1-LE, however, we see a much tighter SIC-meridional wind relationship even at near-stable SIC. This is likely due to the fact that CESM1-LE does not have a landfast ice parameterization (e.g. [26]), which causes it to be more susceptible to changes in wind forcing since there is no frictional force keeping the ice attached to shore, as well as a possible underestimation ice thickness near the shore due to a lack of an anchor point.

The addition of such a parameter, such as the one presented in [26], would improve the model's representation of reality, and would likely reduce both the number and magnitude of events observed in this study. Moreover, there is uncertainty associated with SIC measurements from passive microwave satellites near the coast due to land contamination[27], where satellite have difficulty differentiating between sea ice and land snow. For this analysis we accepted that uncertainty since passive microwave satellite measurements are still the most consistent data source for SIC. ERA5 wind components have also been shown to be less reliable near coastlines, due mainly to the differences in surface roughness and solar irradiation[28].

ERA5 is, as well, less reliable in areas of heterogeneous topography[28], which could affect our results in Novoye Chaplino, Sireniki, and Pakhachi, as they are located near mountain ranges. Despite this, ERA5 has been found to usable in Europe[29] and the South China Sea[30], so we therefore believe ERA5 reasonably represents meteorological conditions in Beringia, within a degree of uncertainty.

Additionally, there could be an introduction of errors from downscaling the 1-degree model resolution for use at these locations, as the resolution is larger than the phenomena. For example, a break-up event could be detected in CESM1-LE that occurs far enough offshore to no longer be considered landfast ice. As a consequence of this, we expect CESM1-LE to overestimate break-up events. Despite these caveats, we find that CESM1-LE generally represents observed break-up events statistically well in terms of number, magnitude, length, frequency, and mechanisms of events. Based on these findings, we use CESM1-LE to predict long-term trends in break-up events.

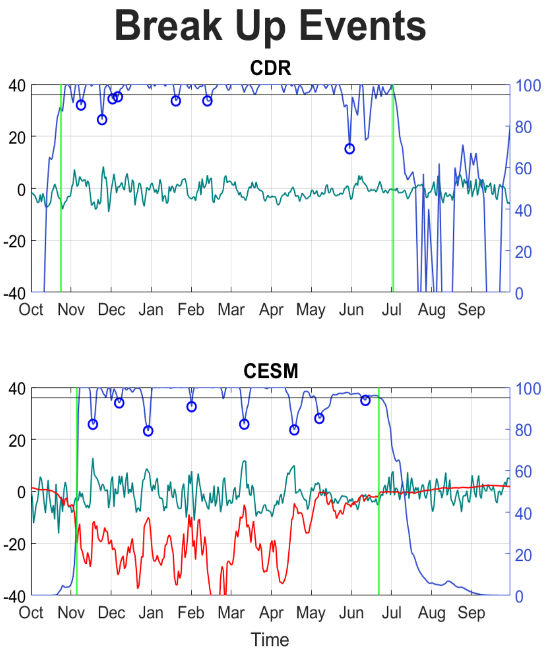


Figure 2. Sea-ice concentration (blue), surface ice temperature (red), and meridional wind speed (dark green) from the CDR, ERA5 and CESM1-LE EM 2 in Utqiagvik. The left-hand side vertical axis is both wind speed (m/s) and temperature (°C). Winter 1980 is shown for both CESM1-LE and CDR to represent typical mid-20th century years. A break-up event (blue circles) is defined as SIC falling below 95% following stable ice cover for 10 days. The start of the freeze season (green lines) is defined as the first day of a series of 10 days with a daily mean greater than 95% SIC, and the start of the melt season (green lines) is defined as the first day of a series of 30 consecutive days with SIC below 95% SIC.

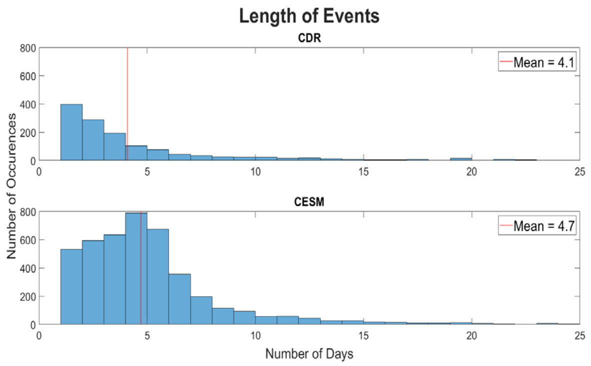


Figure 3. Histogram of the length of all events for one grid cell offshore of each location from the CDR and CESM1-LE datasets. Pakhachi is not included as there are too few events to analyse. The mean of each distribution is included as a vertical red line.

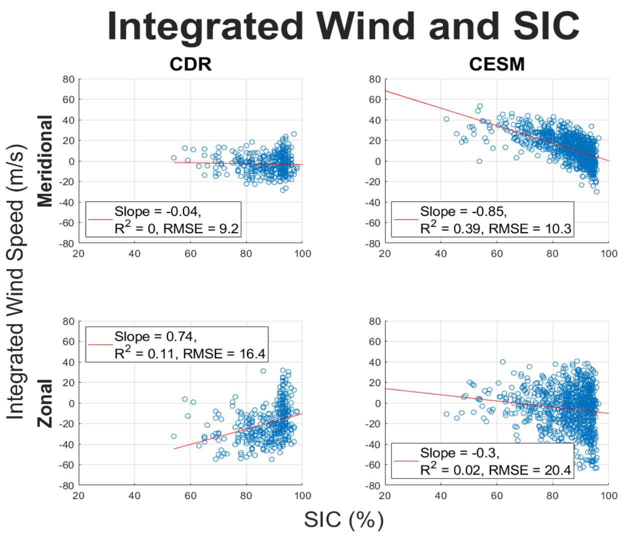


Figure 4. Scatter plot and linear regression of sea-ice break-up event magnitude, defined as minimum SIC of the event, and a four-day moving sum of meridional/zonal wind velocity preceding each event for one grid cell offshore Utqiagvik from the CDR, ERA5 databases and EM 2 of CESM1-LE.

Projections

The number of break-up events in a season can be impacted by two opposing effects. The first effect is the shortening of the winter season, which reduces the number of break-up events by limiting the amount of time to produce them. The second effect is the thinning of the ice cover, which reduces the ice strength and increases the probability of a break-up event for a given wind forcing. Results show a negative trend in the mean number of break-up events per winter for all locations in both CDR (not significant) and CESM1-LE (p-value < 0.01) (see Fig. 5 A.). The negative trend is primarily the result of the decrease in the length of the winter season (see Fig. 6) as opposed to an increase in the time between events (results not shown).

In Utqiagvik and Point Hope, the model predicts a stable landfast ice cover being present until the end of the 21st century, contrary to the other four locations, in which the landfast ice cover becomes too unstable to meet our defined criteria at some point in the 21st century (see Fig. 6). In Gambell, Novoye Chaplino, Sireniki, and Pakhachi, we already see a complete absence of stable landfast ice cover, meaning no freeze date is identified, in the satellite record in 1, 2, 4, and 14 years at each location respectively. This is in accordance with the observational record which shows signs of a transition to a more unstable landfast ice coverage[6,7,31]. Primarily, the absence of a stable ice cover occurs late in the observational record (approximately 2015-2019). Whether this is due to long-term variability, or a trend, cannot be inferred from the short record. Nonetheless, it is consistent with model results.

The dominant effect changing the number of break-up events over the course of a winter (shortening of season versus thinning ice cover) depends on location (see Fig. 5 B.). In both the satellite observations and the model simulation, the more northern locations with a stable landfast ice regime (Utqiagvik and Point Hope) show almost no correlation between the number of events and the length of season; the more southern locations show a significant correlation with the length of the winter season. In CESM1-LE, we observe that Utqiagvik and Point Hope have positive slopes for break-up event frequency, which can be interpreted as 1 event every x days during the winter season (expressed in events per day per year), while Gambell, Novoye Chaplino, and Sireniki have negative slopes (all p-values < 0.01 and mean across 6 locations of -5.80×10^{-5} events/day/year, p-value < 0.01) (results not shown). All locations have decreasing season length (mean of -5.3 days/decade). The combination of these two measures results in negative slopes at all locations for break-up events per year (all p-values < 0.01, mean of −0.020 events/year). In the CDR, we ob- serve that Utqiagvik (p-value < 0.01), Point Hope (not significant), and Novoye Chaplino (p-value < 0.01) have positive slopes for break-up event frequency, Sireniki has a negative slope (p-value < 0.01), and Gambell has a slope of 0 (not significant) (mean of 1.27×10^{-4} events/day/year, not significant). In accordance with the model, all locations have decreasing season length (mean of −9.8 days/decade, p-value < 0.01), resulting in non-significant slopes for break-up events per year at all locations, except for Gambell (p-value ≈ 0.08) and Sireniki (p-value < 0.01) with negative slopes (mean across 6 locations of −0.006 events/year, not significant). In general, the number of events per year decreases significantly everywhere within the CESM1-LE, whereas in satellite observations, the trends are generally not significant, with the exception of winter season length, which decreases in both model and satellite observations significantly.

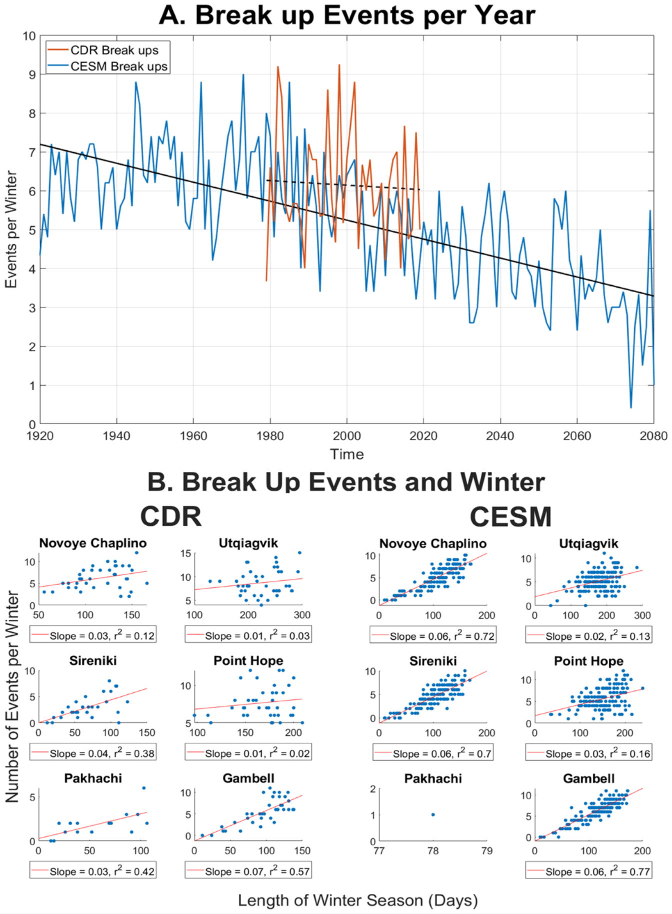


Figure 5. A. Mean Number of break-up events per year for the 6 considered locations identified from CDR (solid orange line) and CESM1-LE (solid blue line) with their linear regressions (black dotted and solid lines respectively). B. Scatter plot between the length of the winter season (days) and number of events (blue dots) with linear regressions (red lines). Pakhachi shows only one break-up event, and therefore a correlation is not possible. Model component is from ensemble member 2 of the CESM1-LE.

Conclusions

The evolution of mid-winter coastal ice break-up events is an important factor to monitor for ensuring the safety of residents within coastal Arctic communities. An analysis of several Arctic communities with the CESM1-LE shows that its sea ice component, despite not having a landfast ice parameter, reasonably simulates break-up event magnitude and frequency. Moreover, in this analysis we demonstrate that these events are associated with synoptic-scale sustained offshore wind, and that their frequency under climate change depends on the two opposing effects of decreasing winter length, and decreasing ice stability. We show that the magnitude of each effect is associated with the minimum winter season length, which is approximated by latitude. This analysis also reveals that some communities have already begun a transition to a new, less stable sea ice regime, where break-up events are more frequent, and the winter season shorter. Model simulations indicate that most coastal Alaskan communities will pass through such a transition in the 21st century. The transition will not be easily predicted from local observations, and is likely to present increased risk for residents making their livelihood on the ice. We hope that understanding these trends can augment local and indigenous knowledge and improve people's capacity to manage the risks.

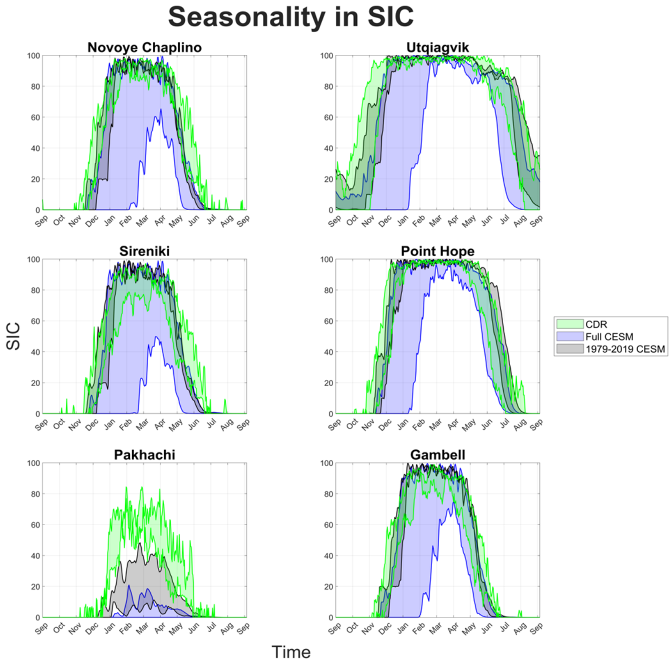


Figure 6. Decadal mean SIC lost between the start and end of each dataset (shaded regions); CDR satellite observations (1979-88 & 2010-19, green), and CESM1-LE simulation (1920-29 & 2071-80, blue, and 1979-88 & 2010-19, black). The full CESM area does not cover the later 1979-2019 CESM area for Pakhachi because the first decade of its time series has anomalously low SIC.

Acknowledgments

This study is a contribution to NSF proposal number 1928126. We acknowledge the CESM Large Ensemble Community Project and supercomputing resources provided by NSF/CISL/Yellowstone[23]. We acknowledge the NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration provided by the National Snow and Ice Data Center[19,22] was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results contain modified Copernicus Climate Change Service information 2021. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. S. A. acknowledges funding from the NSERC USRA bursary program.

References

1. Larsen, J. N. & Fondahl, G. Arctic human development report: regional processes and global linkages (Nordic Council of Ministers, 2014).

2. Arrigo, K. R. in Sea Ice: An Introduction to its Physics, Chemistry, Biology and Geology (eds Thomas, D. N. & Dieckmann, G. S.) 143–183 (Chichester : John Wiley & Sons, 2008).

3. Nelson, R. K. Alaskan Eskimo Exploitation of the Sea Ice Environment (Arctic Aeromedical Laboratory, 1966).

4. Comiso, J. C. Abrupt decline in the Arctic winter sea ice cover. Geophysical Research Letters 33 (Sept. 2006).

5. Holland, M. M., Bitz, C. M. & Tremblay, B. Future abrupt reductions in the summer Arctic sea ice. Geo- physical Research Letters 33 (2006).

6. Stroeve, J. C. et al. The Arctic's rapidly shrinking sea ice cover: a research synthesis. Climatic change 110, 1005–1027 (2012).

7. Comiso, J. C. & Hall, D. K. Climate trends in the Arctic as observed from space. Wiley Interdisciplinary Reviews: Climate Change 5, 389–409 (2014).

8. Serreze, M. C. & Stroeve, J. Arctic sea ice trends, variability and implications for seasonal ice forecasting. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 373, 20140159 (2015).

9. Mioduszewski, J. R., Vavrus, S., Wang, M., Holland, M. & Landrum, L. Past and future interannual vari- ability in Arctic sea ice in coupled climate models. The Cryosphere 13, 113–124 (2019).

10. Druckenmiller, M. L., Eicken, H., Johnson, M. A., Pringle, D. J. & Williams, C. C. Toward an integrated coastal sea-ice observatory: System components and a case study at Barrow, Alaska. Cold Regions Sci- ence and Technology 56, 61–72 (2009).

11. Dyke, A. S., England, J., Reimnitz, E. & Jetté, H. Changes in driftwood delivery to the Canadian Arctic Archipelago: The hypothesis of postglacial oscillations of the Transpolar Drift. Arctic, 1–16 (1997).

12. Mahoney, A., Eicken, H., Gaylord, A. G. & Shapiro, L. Alaska landfast sea ice: Links with bathymetry and atmospheric circulation. Journal of Geophysical Research 112 (2007).

13. Jones, J. et al. Landfast sea ice breakouts: Stabilizing ice features, oceanic and atmospheric forcing at Barrow, Alaska. Continental Shelf Research 126, 50–63 (Sept. 2016).

14. John, C. et al. Observations on shorefast ice dynamics in Arctic Alaska and the responses of the Iñupiat hunting community. Arctic, 363–374 (2004).

15. Loewen, A. Landfast Sea Ice Break-up Processes in Admiralty Inlet, NU PhD thesis (Carleton University, 2020).

16. Antonova, S. Spatial and temporal variability of landfast ice in the Laptev Sea PhD thesis (Alfred Wegener Institute, 2012).

17. Laliberté, F., Howell, S. E., Lemieux, J.-F., Dupont, F. & Lei, J. What historical landfast ice observations tell us about projected ice conditions in Arctic archipelagoes and marginal seas under anthropogenic forcing. The Cryosphere 12, 3577–3588 (2018).

18. Howell, S. E., Laliberté, F., Kwok, R., Derksen, C. & King, J. Landfast ice thickness in the Canadian Arctic Archipelago from observations and models. The Cryosphere 10, 1463–1475 (2016).

19. Meier, W., Fetterer, F., Windagel, A. K. & Stewart, J. S. NOAA/NSIDC Climate Data Record of Passive Mi- crowave Sea Ice Concentration, Version 4. NSIDC: National Snow and Ice Data Center (2021).

20. Cavalieri, D. J., Gloersen, P. & Campbell, W. J. Determination of sea ice parameters with the Nimbus 7 SMMR. Journal of Geophysical Research: Atmospheres 89, 5355–5369 (1984).

21. Comiso, J. C. Characteristics of Arctic winter sea ice from satellite multispectral microwave observations. Journal of Geophysical Research: Oceans 91, 975–994 (1986).

22. Hersbach, H. et al. ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (2018).

23. Kay, J. E. et al. The Community Earth System Model (CESM) Large Ensemble Project: A Community Re- source for Studying Climate Change in the Presence of Internal Climate Variability. Bulletin of the Amer- ican Meteorological Society 96, 1333–1349 (2015).

24. Mahoney, A. R., Eicken, H., Gaylord, A. G. & Gens, R. Landfast sea ice extent in the Chukchi and Beaufort Seas: The annual cycle and decadal variability. Cold Regions Science and Technology 103, 41–56 (2014).

25. Riahi, K. et al. RCP 8.5—A scenario of comparatively high greenhouse gas emissions. Climatic change 109, 33–57 (2011).

26. Lemieux, J.-F. et al. A basal stress parameterization for modeling landfast ice. Journal of Geophysical Research: Oceans 120, 3157–3173 (2015).

27. Bennartz, R. On the use of SSM/I measurements in coastal regions. Journal of atmospheric and oceanic technology 16, 417–431 (1999).

28. Gualtieri, G. Reliability of ERA5 Reanalysis Data for Wind Resource Assessment: A Comparison against Tall Towers. Energies 14, 4169 (2021).

29. Molina, M. O., Gutiérrez, C. & Sánchez, E. Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. International Journal of Climatology 41, 4864–4878 (2021).

30. Jiang, Y. et al. Evaluation of HRCLDAS and ERA5 Datasets for Near-Surface Wind over Hainan Island and South China Sea. Atmosphere 12, 766 (2021).

31. Meier, W. N. et al. Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. Reviews of Geophysics 52, 185–217 (2014).

Submitted: 01/24/2022
Accepted: 03/27/2022

Research Article

[1]Department of Microbiology and Immunology McGill University, Montreal, QC, Canada.

Keywords

Bacteriophage (phage), phage cluster, genome annotation, SEA-PHAGES, *Mycobacterium smegmatis mc² 155*

Email Correspondence

morgane.brouillard-galipeau@mail.mcgill.ca

Morgane Brouillard-Galipeau[1], Bao-An Chau[1], Jamie Cyr[1], Rafael Intrevado[1], Sunu (Isabel) Kim[1], Cal Koger-Pease[1], Elizabeth Lapshina[1], Alexandra Mircescu[1], Daniella Serrador[1], Michael Slattery[1], Benjamin Vonniessen[1], Michael Shamash[1], Dr. Jasmin Chahal[1]

Genome Annotation of Novel K1 Subcluster Mycobacteriophage Blizzard

Abstract

The evolution of antimicrobial resistant pathogens constitutes a significant global public health threat. Combined with the lack of incentive for pharmaceutical companies to invest in developing new antibiotics, it is clear alternative treatments are needed. Bacteriophages present one possible avenue as they harness the diversity and specificity of a microorganism that has coevolved with bacteria. However, little is known about these bacterial viruses. The SEA-PHAGES program was designed to identify and characterize novel bacteriophages and their associated gene functions. Herein, we report the genome annotation of one such novel phage: Mycobacteriophage Blizzard (GenBank accession number MW712733). Blizzard's gene content was functionally annotated using bioinformatic tools including DNA Master, Phamerator, and NCBI BLAST, to call start sites as well as predict gene function. Overall, 96 genes were identified, including a tRNA and a translational frameshift, using highly similar reference phages BEEST, Belladonna, and CREW. From the 96 genes identified, 46 were functionally annotated. The remaining 50 genes have unknown functions due to the lack of significant matches in the databases. Our results demonstrate a novel annotated phage, whose genome serves to expand the understanding of phage biology and potential implications as alternative treatment to antibiotics.

Introduction

Bacteriophages (phages) are the most abundant, ubiquitous, and diverse microorganisms on Earth[2]. Phages are viruses that infect bacteria and have been isolated from every biome where their bacterial hosts are found[1]. Their host range can span from a considerable breadth of numerous strains across bacterial species or genera to a narrow specificity of a single strain within a bacterial species. Phages that bind to a unique receptor are prone to show a narrow host range, while those that bind to multiple receptors tend to have a larger range[3].

Phages are differentially classified according to their physical structure. The largest of these classifications, the *Caudovirales* order, represents over 96% of the phages known to date. *Caudovirales* are characterized as non-enveloped, tailed phages with a double-stranded DNA (dsDNA) genome contained in an icosahedral protein capsid[4, 5]. A phage that is shown to infect pathogenic bacteria but does not kill commensal organisms can be employed to develop phage therapies, which use phages to treat bacterial infections[3]. This therapy presents an alternative option relevant to the antibiotic resistance crisis.

With the rise in antimicrobial-resistant infections and the pipeline for new antibiotics growing dry, phage therapy has become a more relevant solution. It has been shown to successfully clear multidrug-resistant mycobacteria both *in vivo* and *in vitro*[6, 7, 8]. Interestingly, the selected phages used to target these bacteria were isolated through the SEA-PHAGES program. SEA-PHAGES (Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science) is a program dedicated to cataloguing novel bacteriophages in the public Actinobacteriophage Database (PhagesDB)[9]. To isolate phages, the SEA-PHAGES program uses Actinobacteria, such as mc² 155, as bacterial hosts. To date, 17,000 actinobacteriophages have been isolated and 3,000 have been sequenced by SEA-PHAGES[10]. This expansion of phage gene sequencing has necessitated the grouping of mycobacteriophages into clusters and subclusters according to their nucleotide similarity[11]. Following the sequencing of novel phages, the genome is annotated. Annotated genomes improve efficiency when developing phage therapies, which is essential when treating patients with critical bacterial infections[12]. The more phages are characterized, the more options there are for researchers attempting phage therapy, and the more rapidly we are able to identify and gather phages to target specific bacteria[12].

Figure 1. Phylogenetic tree comparing the full genome of Blizzard (Query_44707 - red box) to the full genome of similar mycobacteriophages. BEEST (green box) appeared as the closest relative to Blizzard, sharing the closest ancestor. CREW (blue box) and Belladonna (purple box) appeared further along the branches. Obtained from BLASTn[14].

Genome annotation utilizes bioinformatics, a multidisciplinary field of science that focuses on computational analysis of biological data[13]. To annotate a phage genome, various tools including DNA Master[14], NCBI BLAST[15], Phamerator[16], and Aragorn[17] are used to align DNA and predicted protein sequences of interest to reference phages. This comparison is used to subsequently infer gene start coordinates, gene functions, as well as genetic elements such as frameshifts and tRNAs. The annotated genome must then be reviewed manually in accordance with guidelines set by GenBank to standardize annotation and generate admissible data.

Our chosen novel phage, Blizzard, was discovered in 2013 by Jean Klonchko Bull at Hope College. It was isolated from an enriched soil sample using the host *Mycobacterium smegmatis* mc^2 155, a nonvirulent relative of *M. tuberculosis*, and sequenced via Illumina sequencing[18]. We aim to annotate the Blizzard genome by finding putative protein coding and tRNA genes as well as predicting their start sites and functions to contribute to the PhagesDB database. Blizzard has over 97% homology with multiple members of its K1 subcluster[15]. We hypothesize that upon analysis, according to the annotation guidelines set by GenBank, Blizzard's genome will reveal gene functions, start sites, and tRNAs similar to many subcluster K1 phages, some of which are promising therapeutic agents.

Methodology

Retrieving the Complete Blizzard Sequence

The discovery and sequencing output data of Blizzard was obtained from PhagesDB[18]. From here, we extracted the fasta file of the complete Blizzard sequence as well as other characterization information.

DNA Master version 5.23.5 was used as the primary program for the genome annotation[14]. DNA Master setting preferences were updated to reflect the SEA-PHAGES Bioinformatics Guide recommendations[19]. The auto-annotate function of DNA Master was used to add gene predictions to the draft annotation of the phage genome[14]. We then performed a BLAST search on all the genes against the NCBI public database[15].

Gene Calling

To determine the validity of the gene calls made by DNA Master, we assessed which auto-annotation programs called the gene, the evidence of coding potential, and the presence of the same gene in other phages from the same cluster. The two auto-annotation programs used were Glimmer and GeneMark[20, 21]. Both systems specialize in determining the gene start sites of microbial genomes and can distinguish between coding and non-coding sequences[20, 21]. Coding potential was also determined using GeneMarkS graphs. The bioinformatic tool Phamerator was used to compare the Blizzard genome to BEEST (Accession number: MH509444.1, 9), CREW (Accession number: KY380102.1, 9), and Belladonna (Accession number: MH697578.1, GenBank direct submission), three annotated phages from the same cluster as Blizzard[16].

These phages share 99.95%, 96.36%, and 95.34% in nucleotide identity, respectively, with Blizzard[15]. Percent identity and e-value, indicators of quality and similarity to the query sequence, were also recorded for each gene. Additionally, every gene was assessed on their adherence to the major guiding principles as defined by the SEA-PHAGES bioinformatics guide. An overlap of 4 base pairs (bp) between genes is optimal, while over 30 bp may be unfavorable. The length of a gene is usually above 200 bp, but they can still be considered if they are over 120 bp[19].

Gene Start Site

To determine the start site of the genes, the suggested calls of Glimmer and GeneMark were noted[20, 21]. Next, using the Ribosome Binding Site (RBS) or Shine-Dalgarno Sequence Finder within DNA Master, the start sites with the most favorable scores were selected[14]. The RBS Sequence Finder is an algorithm assessing the upstream sequence of start codons, evaluating multiple variables such as the distance between this sequence and the start codon. The Z-Score and final score were used to determine the caliber of each start call[14]. The program Starterator was then used to examine all the genes within the same cluster as Blizzard. Starterator examines the alignment of the longest open reading frame (ORF) of each gene in related phages and assists in determining which start is present in most annotated genomes[19]. Then, the local BLASTp results were assessed to determine if the start site was conserved in other phage genomes[15]. The consensus of the percentage alignment, percentage similarity, e-value, and start position alignment were used to call the start sites.

Gene Function

To determine the putative function of genes, the amino acid sequences were compared by sequence alignment via PhagesDB, GenBank, and NCBI databases[15, 18, 22]. Only proteins with an e-value of 10^{-4} or less, with an appropriate query coverage, were selected. HHpred was then used to analyze the best match to the selected database sources (PDB, SCOPe7-A, Pfam-A, and NCBI CD) with a high probability score (>90%)[23]. Synteny, the use of the location of a gene to ascribe its function, was evaluated using Phamerator with reference phages BEEST, Belladonna, and CREW[16]. The general consensus of these tools determined the function of the genes.

Identifying tRNA in the genome

Aragorn and tRNAscan-SE were used to assess the presence of tRNAs through the prediction programs[17, 24]. Both programs were run during manual annotation to call for the stop site of the tRNA, if present, and scan the phage genome for a conserved region in the tRNA. Results from Aragorn embedded in DNA Master were compared to those from tRNAscan-SE and the updated online version of Aragorn.

Evaluating the Presence of Frameshifts

Phamerator and BLAST were used to determine the location of the shift in the protein sequences[15, 16]. Then, the direction of the frameshift and the coordinates of the slippery sequence were identified[19]. To edit frameshifts for Blizzard, we used the Six Frame Translation window in DNA Master[14]. The gene slippage was identified by observing a 5'-GGAAAA-3' sequence common to all K1 phages[25].

Adding Genes

After manual annotation of genes called by DNA Master, large gaps (>120 bp) between genes were evaluated using the same protocol employed for the manual verifications, to see whether any genes were missed during auto-annotation.

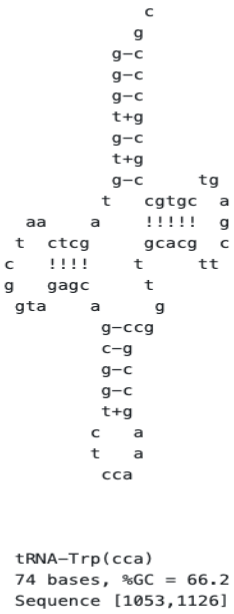
Results

Overall Genome Characteristics

Blizzard is part of the K1 subcluster. Its genome is 59,905 bp and has a G/C content of 66.6%[18]. The auto-annotation resulted in a list of 97 protein-coding genes and one tRNA sequence. After the manual annotation, a total of 95 protein-coding genes and one tRNA^{Trp} were confirmed. Functions were assigned to 46 genes, including all the genes that qualified for functional assignment via synteny. Additionally, functions were called for all required genes as outlined by the SEA-PHAGES Bioinformatics Guide.

tRNA

The tRNA gene identified in the original auto-annotation was located at 1052-1128 bp, and was identified as a tRNA^{Trp}, carrying a CCA anticodon (Figure 2). This gene was 76 bp long, within the normal range for a tRNA. However, its 3' end did not include the necessary terminal sequence. The Aragorn software output resulted in a correctly trimmed tRNA^{Trp} sequence at 1053-1126 bp, containing one C base of the conserved 3' terminal sequence.



```
              c
             g
           g-c
           g-c
           g-c
           t+g
           g-c
           t+g
           g-c      tg
          t   cgtgc  a
   aa    a    !!!!!  g
  t ctcg      gcacg  c
 c  !!!!     t     tt
 g  gagc      t
  gta    a      g
          g-ccg
          c-g
          g-c
          g-c
          t+g
         c  a
         t  a
          cca


 tRNA-Trp(cca)
 74 bases, %GC = 66.2
 Sequence [1053,1126]
```

Figure 2. The predicted structure of the trp-tRNA. Structure predicted by Aragorn 1.2.38[22].

Modifications to the Auto-Annotation

Deleted Genes

Predicted genes 5, 18, and 42 were determined to not be true protein-encoding genes as they all had only one or zero BLASTp matches, were not called by GeneMark, and had no similar genes in Belladonna or CREW (Figure 3). The genes were all shorter than 200 bp and genes 18 and 42 faced the reverse direction of those flanking them, which is uncommon for putative protein-encoding genes[16].

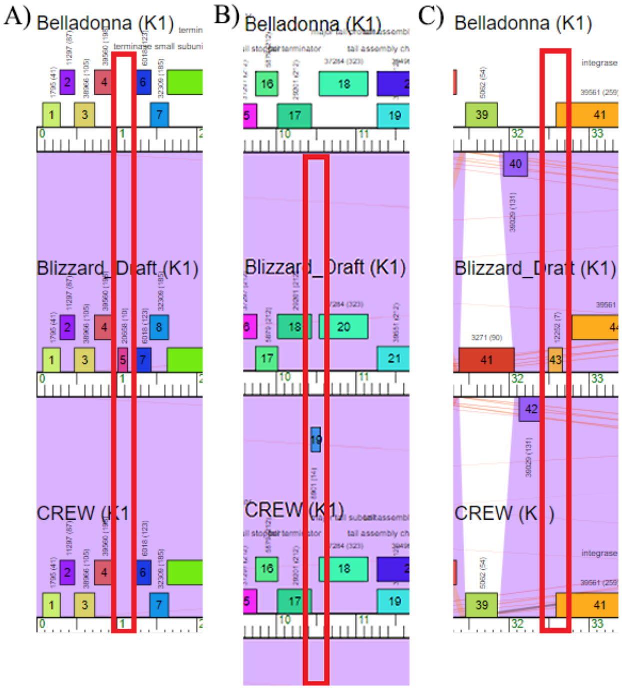


Figure 3. Phamerator maps comparing (A) Blizzard's auto-annotated gene 5 (labelled 5) to gaps in Belladonna and CREW, (B) Blizzard's auto-annotated gene 18 (labelled 19) to gaps in Belladonna and CREW, and (C) Blizzard's auto-annotated gene 42 (labelled 43) to gaps in Belladonna and CREW. Each box represents a gene, and the scale indicates kbp. The colour of the genes indicates their gene family, and background purple indicates sequence similarity, while background white indicates dissimilarity. Map generated by Phamerator[19].

Frameshift

Genes 20 and 21 encode the tail assembly chaperone genes. These genes contain a programmed -1 translational frameshift across a slippery sequence (5'-GGAAAA-3'). Such a frameshift is conserved across K1 phages, including BEEST, Belladonna, and CREW (Figure 4)[16, 18, 25]. The slippery sequence was identified with the first "A" nucleotide located at 11,670 bp (Figure 5). The annotation for gene 21 was modified to include two regions. The first region began at the start site for gene 20 (11,629 bp) and ended at the first "A" (11,670 bp) while the second region began at 11,629 bp and ended at the gene 21 stop site, 12,116 bp.

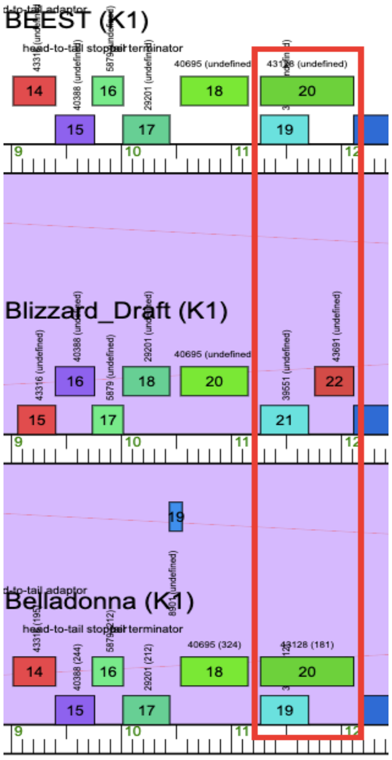
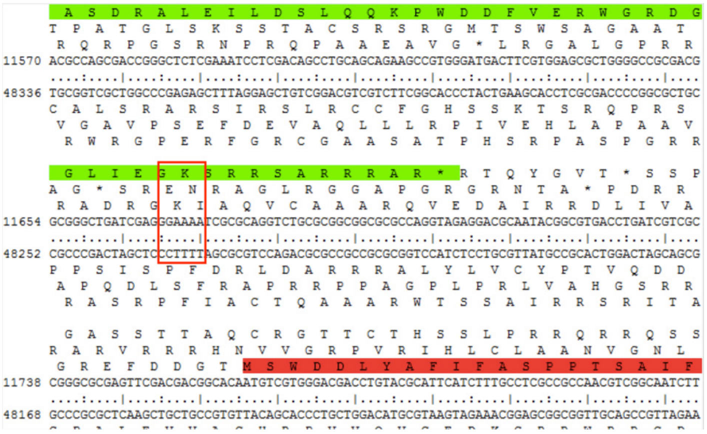


Figure 4. Phamerator genome map showing tail assembly genes in Blizzard (auto-annotated), Belladonna, and BEEST. The programmed translational frameshift is visible in Belladonna and BEEST genes 19 and 20 but is not yet edited in Blizzard_Draft genes 21 and 22 (within the red box). Each box represents a gene, and the scale indicates kbp. The colour of each gene indicates their gene family and background purple indicates sequence similarity between the two genomes. Map generated by Phamerator[19].



```
       A   S   D   R   A   L   E   I   L   D   S   L   Q   Q   K   P   W   D   D   F   V   E   R   W   G   R   D   G
      T   P   A   T   G   L   S   K   S   S   T   A   C   S   R   S   R   G   M   T   S   W   S   A   G   A   A   T
       R   Q   R   P   G   S   R   N   P   R   Q   P   A   A   E   A   V   G   *   L   R   G   A   L   G   P   R   R
11570 ACGCCAGCGACCGGGCTCTCGAAATCCTCGACAGCCTGCAGCAGAAGCCGTGGGATGACTTCGTGGAGCGCTGGGGCCGCGACG
      ....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....
48336 TGCGGTCGCTGGCCCGAGAGCTTTAGGAGCTGTCGGACGTCGTCTTCGGCACCCTACTGAAGCACCTCGCGACCCCGGCGCTGC
      C   A   L   S   R   A   R   S   I   R   S   L   R   C   C   F   G   H   S   S   K   T   S   R   Q   P   R   S
       V   G   A   V   P   S   E   F   D   E   V   A   Q   L   L   L   R   P   I   V   E   H   L   A   P   A   A   V
        R   W   R   G   P   E   R   F   G   R   C   G   A   A   S   A   T   P   H   S   R   P   A   S   P   G   R   R

       G   L   I   E   G   K   S   R   R   S   A   R   R   R   A   R   *   R   T   Q   Y   G   V   T   *   S   S   P
      A   G   *   S   R   E   N   R   A   G   L   R   G   G   A   P   G   R   G   R   N   T   A   *   P   D   R   R
       R   A   D   R   G   K   I   A   Q   V   C   A   A   A   R   Q   V   E   D   A   I   R   R   D   L   I   V   A
11654 GCGGGCTGATCGAGGGAAAATCGCGCAGGTCTGCGGCGGCGCGCCAGGTAGAGGACGCAATACGGCGTGACCTGATCGTCGC
      ....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....
48252 CGCCCGACTAGCTCCCTTTTAGCGCGTCCAGACGCGCCGCCGCGCGGTCCATCTCCTGCGTTATGCCGCACTGGACTAGCAGCG
      P   P   S   I   S   P   F   D   R   L   D   A   R   R   R   A   L   Y   L   V   C   Y   P   T   V   Q   D   D
       A   P   Q   D   L   S   F   R   A   P   R   R   P   P   A   G   P   L   P   R   L   V   A   H   G   S   R   R
        R   A   S   R   P   F   I   A   C   T   Q   A   A   A   R   W   T   S   S   A   I   R   R   S   R   I   T   A

       G   A   S   S   T   T   A   Q   C   R   G   T   T   C   T   H   S   S   L   P   R   R   Q   R   R   Q   S   S
      R   A   R   V   R   R   R   H   N   V   V   G   R   P   V   R   I   H   L   C   L   A   A   N   V   G   N   L
       G   R   E   F   D   D   G   T   M   S   W   D   D   L   Y   A   F   I   F   A   S   P   P   T   S   A   I   F
11738 CGGGCGCGAGTTCGACGACGGCACAATGTCGTGGGACGACCTGTACGCATTCATCTTTGCCTCGCCGCCAACGTCGGCAATCTT
      ....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....
48168 GCCCGCGCTCAAGCTGCTGCCGTGTTACAGCACCCTGCTGGACATGCGTAAGTAGAAACGGAGCGGCGGTTGCAGCCGTTAGAA
```

Figure 5. DNA Master Six frame view showing gene 20 in green and gene 21 in red, with the slippery sequence in the red box. Figure from DNA Master, version 5.23.5[15].

Added Gene

Large gaps are uncommon in phage genomes, such that stretches of unassigned DNA greater than 120 bp were examined for putative genes[19]. An additional gene not called by the auto-annotation was identified between auto-called genes 71 and 72. This region of 409 bp did not show coding potential on the GeneMark maps (Figure 6), but it had well-aligned BLASTp matches to several putative proteins. There are also homologous genes in Belladonna and BEEST, but not in CREW (Figure 7). According to the Starterator report, this protein was annotated in 46 of subcluster K1's 83 non-draft members, excluding Blizzard, in 1 of subcluster K3's members, and in 10 of subcluster K6's members[26]. However, few BLASTp matches had a percent identity greater than 90% with this protein's sequence[15], indicating that this protein sequence is weakly conserved.

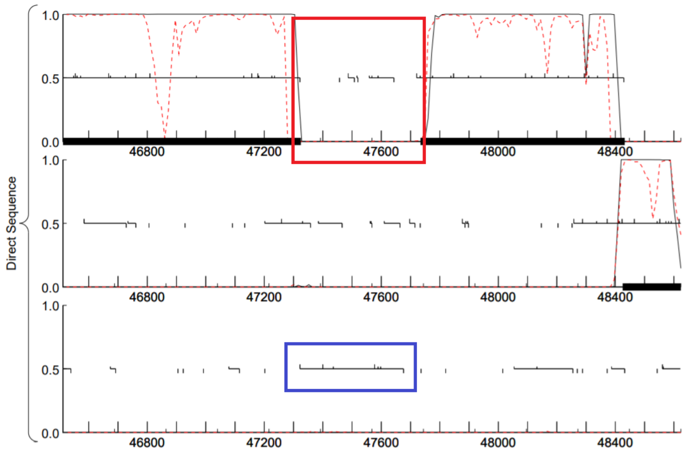


Figure 6. GeneMarkS graph of the gap between genes 71 and 72 (red box) and the ORF where gene 71.5 was inserted (green box). Horizontal lines indicate ORFs, with upward ticks indicating potential start sites and downward ticks indicating stop sites. Black curves indicate typical coding potential, white red dotted curves indicate atypical coding potential. Scale is measured in base pairs. Figure from GeneMarkS version 2.5p[18].

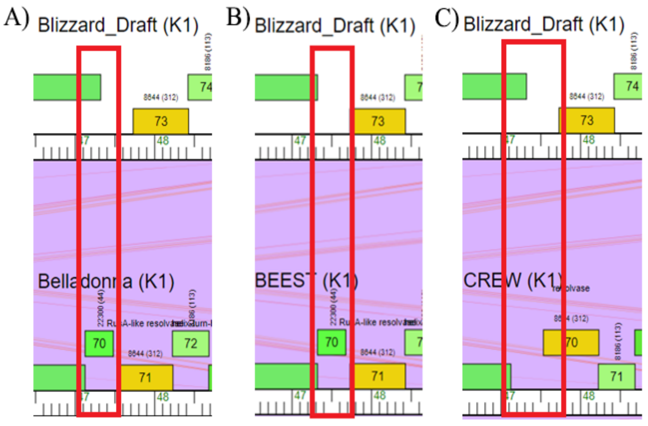


Figure 7. Phamerator maps comparing the gap in Blizzard auto-annotation where gene 71.5 was inserted with (A) gene 70 in Belladonna, (B) gene 70 in BEEST, and (C) a gap in CREW. Each box represents a gene, and the scale indicates kbp. The colour of the genes indicates their gene family, and background purple indicates sequence similarity. Map generated by Phamerator[19].

Ambiguous Gene Calls

Several genes had conflicting evidence for different potential start sites, such as gene 46. The gene was not called by GeneMark's auto-annotation. However, it was called by Glimmer, had atypical coding potential in the GeneMarkS graph (Figure 8), and had strong BLASTp similarity to the well characterized Cro protein, and therefore it was determined to be a gene. Glimmer called the start site at 35,432 bp, but the start site with the best RBS score was 35,456 bp. Starterator called a third start site at 35,474 bp as the consensus, present in 20.3% of annotated homologues. The 35,432 bp start resulted in the longest reasonable ORF. In the K1 phage BEEST, a similar gene had a start site at 35,474 bp (Figure 9). Gene sequences with the different start sites were analyzed by BLASTp, and the sequence starting at 35,432 bp had the strongest results. Ultimately, the Glimmer call, BLASTp results, and the ORF length were determined to be the stronger pieces of evidence, and the start site was called at 35,432 bp.

Another ambiguous start site call was gene 62. Glimmer called the start at 41,471 bp, while GeneMark called it at 41,453 bp. The Starterator report called the consensus start at 41,447 bp. 41,447 bp also had the highest RBS score and longest possible ORF. Additionally, it resulted in the ideal 4 bp overlap with the previous gene (bioinformatics guide). For these reasons, 41,447 bp was called as the start site for gene 62.

Some gene functions were difficult to call due to conflicting evidence. Gene 71's strongest BLASTp match was DNA primase from Adephagia, however, further analysis suggested the presence of additional domains: there were BLASTp matches to DNA primase/helicase and DNA primase/polymerase. The HHPred data support the presence of additional domains, as the 5' end sequence of gene 71 best matched a DNA primase, and the 3' end matched a helicase. We sought to confirm a helicase domain running only the suspected domain's sequence in BLASTp but obtained no different results than for the entire gene 71. Thus, there was insufficient evidence to conclude that gene 71 had a helicase domain, and it was called a DNA primase.

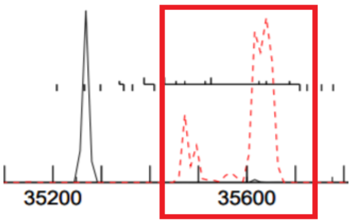


Figure 8. GeneMarkS graph of gene 46 (35432-35713 bp). Horizontal lines indicate ORFs, with upward ticks indicating potential start sites and downward ticks indicating stop sites. Black curves indicate typical coding potential, white red dotted curves indicate atypical coding potential. Scale is measured in base pairs. Figure from GeneMarkS version 2.5p[18].

Gene 92's function was predicted to be "antitoxin in a toxin/antitoxin system, HicB-like". BLASTp results for this protein showed primarily hypothetical protein matches, with some helix-turn-helix DNA binding domain and HicB-like antitoxin matches. HHPred had multiple matches to bacterial HicB antitoxin. Although non-phage matches are generally less convincing, here the gene function in question was HicB-like, indicating similarity to HicB. Given that the data for gene 93 was confidently annotated as HicA-like toxin, synteny suggests that its obligatory counterpart HicB-like antitoxin should be the next gene, despite poor BLASTp data.

Discussion

In our annotation of Blizzard, we identified 96 putative genes, including one tRNA. We were able to assign known functions to 46 of the genes. In accordance with our hypothesis, the called genes and their locations were like other K1 phages. Our manual annotation was revised by the SEA-PHAGES review board, who completed the annotation of gene 1 and added gene 94, bringing the total gene count to 97. Blizzard's 5' end showed a conserved sequence of structural genes and many hypothetical proteins directly downstream of that region. Closer to the 3' end of the genome, non-structural proteins were identified. These are required for the phage's lysogenic life cycle and DNA replication.

Blizzard has one more gene than its closest relative BEEST but otherwise contains the same gene families in the same locations[16]. Adephagia and Blizzard are also nearly identical; Blizzard has two additional genes, genes 70 and 94, as well as a non-homologous gene number 64[16]. The prominent presence of gene 70 in the K1 subcluster compared to its sparse occurrence in the other K subclusters despite its low BLAST matches could indicate that the gene is likely not involved in a critical phage function but that its function may still be somewhat advantageous to K1 phages specifically. However, it should be noted that there is a possibility for this gene to be an artifact; *in vitro* experiments would be needed to confirm the annotation's validity.

Notably, we identified a translational frameshift in the 5' region, in the tail assembly chaperone genes. This corresponds with the highly conserved translational frameshift found in dsDNA bacteriophage tail assembly genes[25]. Although other phage genes can undergo translational frameshifts, we only considered the one in the tail assembly chaperone genes, as it is the only programmed translational frameshift in phage genomes supported by *in vitro* experiments[25].

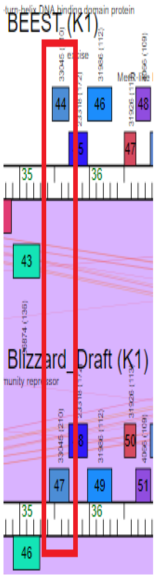


Figure 9. Phamerator map comparing Blizzard's auto-annotated gene 46 (labelled 47), starting at 35432 bp, with BEEST's gene 44 (labelled 44), which starts at 35474 bp. Each box represents a gene, and the scale indicates kbp. The colour of the genes indicates their gene family, and background purple indicates sequence similarity. Map generated by Phamerator[19].

Blizzard's genome contains a single tRNA^{Trp} gene, as do most phages in sub-cluster K1[16, 18]. Phages use host translation machinery, so they do not typically encode translational proteins, although some phages encode tRNA molecules[27]. Phages may encode tRNAs for amino acids used commonly in phage proteins but not host proteins in order to make translation more efficient[27]. This would suggest that Blizzard's proteins contain more tryptophan codons than its host(s). This result aligns with phage biology as Blizzard is a temperate phage and thus integrates its genome within that of the host. Temperate phages encode fewer tRNAs than lytic phages, which replicate in and lyse their host cell, as temperate phages have lower codon biases and few compositional differences from their host genome, facilitating genome integration[27].

We called the function of gene 71 as DNA primase, but HHPred data suggests that this gene contains a helicase domain as well. Gene 71 could contain a primase/helicase polyprotein that is separated post-translationally via proteolytic cleavage. If gene 71 codes for a polyprotein, there is likely a protease gene present elsewhere in Blizzard's genome. Alternatively, this gene could encode a single bifunctional primase/helicase enzyme, as observed in other *Caudovirales* phages such as T7[28].

Blizzard's gene 40 did not have a homologue in many K1 phages, however there are homologues present in some K3 and K6 phages. This implies that the gene may have been present in the last common ancestor of K cluster phages and was lost in most K1 phages. Alternatively, Blizzard and other cluster K phages could have acquired it horizontally[29]. *In vitro* experimentation involving gene 40 knockout strain of Blizzard could determine the function and necessity of this gene.

While manual annotation allows for integration of results from various bioinformatics and auto-annotation software, the methods can still be perfected as the estimates of gene functions are mainly based on synteny and comparison with closely related phage genomes. Since many of the tools used also rely on similar algorithms, it is also possible that if one tool provides erroneous data, other tools may also have made the same error. This limitation illustrates the importance of using a variety of types of tools with differing algorithms to come to a well-supported decision when making a conclusion. Phage-specific auto-annotation software, PHANOTATE and Prodigal, can be used to confirm our results[30]. Another limitation of this genome analysis is that it is not possible to determine the functions of many of the proteins present in Blizzard. To further confirm our predictions and function assignments and determine the function of hypothetical proteins, in vitro proteomic studies should be performed.

Figure 10. The top HHPred search results for gene 71's product. These results are visualized as thick, coloured bars. Red-coloured bars indicate a strong match to the query sequence. The length of the bar corresponds to the section of the query (the thin green bar at the top) that the result matches with. Figure from HHPred[21].

Significance of Genome Annotation

As phages play a crucial role in ecological processes, the characterization of isolates contributes to the understanding of the diversity of phages in the environment. By studying genome organization and gene conservation, we can better understand phage biology. Furthermore, the screening and identification of phages adds to the library of characterized genomes required for phage therapy[31].

In fact, some phages isolated and annotated by SEA-PHAGES have been used for the treatment of drug-resistant mycobacterial strains[7]. Recently, Guerrero-Bustamante et al. created a five-phage cocktail that efficiently kills several strains of *M. tuberculosis* while minimizing phage resistance[8]. One of the five phages was an engineered strain of Adephagia, a phage discovered through the SEA-PHAGES program. The tyrosine integrase and immunity repressor genes necessary for lysogeny were deleted in the strain AdephagiaΔ41Δ43, creating a lytic derivative[8]. Adephagia and Blizzard are both K1 phages with similar gene functions, as previously discussed. This makes Blizzard an ideal candidate to test against *M. tuberculosis*. Our annotation contains the putative locations and sequences of the immunity repressor and integrase genes and could be used to target these genes to engineer a lytic mutant. Furthermore, Blizzard has first been isolated using the *Mycobacterium smegmatis* mc^2 155 host, a laboratory strain of a non-pathogenic relative of *M. tuberculosis*[18, 32, 33]. This host has been used to isolate 11,752 phages thus far, across at least 31 different clusters, yet despite this wide diversity and large number, only phages of cluster K and subclusters A2 and A3 are known to also infect *M. tuberculosis* efficiently[18, 34]. Little is currently known about receptors phages utilize to invade *Mycobacterium* hosts[34], but it can be hypothesized that *M. tuberculosis* and *M. smegmatis* share a surface protein that is used as a receptor by K-cluster phages. Because Blizzard is contained in subcluster K1, it is possible that Blizzard can infect *M. tuberculosis* in a similar way it can penetrate the *M. smegmatis* host. Thus, creating a lytic Blizzard derivative and testing it for activity against *M. tuberculosis* through plaque assays is a promising future direction, as this could provide another potential phage for use in a cocktail against tuberculosis.

Finally, when choosing bacteriophages to use in a cocktail, it is important to determine whether the phages in question contain any genes that would promote the virulence of the bacteria, such as antibiotic resistance genes (ARGs). Current data suggests that phages, especially phages with lytic activity, rarely encode ARGs[35, 36]: a recent study showed that of 5295 viral sequences reconstructed from river samples, only 25 encoded ARGs[35]. This study also found that in viromes of various habitats such as soil and ocean water, only 0.001 to 0.440% of viral reads were predicted to be ARGs[35]. Because so few ARGs were annotated from phage genomes, and because our phage annotation depends on previously annotated phages, it is not possible to identify ARGs in Blizzard based on bioinformatic annotation alone; wet lab experiments would need to be performed. However, even if it is found that Blizzard contains an ARG, it would be possible to modify its genome to remove the gene for use in therapeutics.

Conclusion

Through this project, we annotated the genome of Blizzard, a temperate K1 phage. Blizzard has 96 genes; protein functions were identified for 46 of these genes, including several genes characteristic of temperate phages, as well as a tRNA^{Trp}. Though the annotation relies on a limited number of tools and techniques, further *in silico* or *in vitro* experiments could be performed to confirm the gene positions and functions. The annotated genome of Blizzard furthers our understanding of phage biology and allows better characterization of the phage for use in therapy against antibiotic-resistant bacterial infections. Additionally, the annotation facilitates the creation of targeted knockouts of Blizzard, which allows for creation of a lytic derivative, as was done with Adephagia[7]. Blizzard or its lytic derivative could be tested against drug-resistant pathogens to examine their possible use in phage therapy. Phages such as Blizzard provide an alternative treatment for bacterial infections that could prove invaluable in the fight against the antibiotic-resistance crisis.

Acknowledgments

We would like to thank our supervisors Dr. Jasmin Chahal and Michael Shamash for their guidance. We would also like to thank the SEA-PHAGES program, Dr. Samantha Gruenheid, and the Microbiology and Immunology Department at McGill University for this research opportunity. Additionally, thank you to Drs. Jesse Shapiro and Ken Dewar for their contributions to the course. Finally, we would like to extend our thanks to Hope College for isolating Blizzard and to Pittsburgh Bacteriophage Institute for sequencing the phage genome.

References

1. Salmond, G.P.C. & Fineran, P.C. A century of the phage: past, present and future. Nat. Rev. Microbiol. 13, 777-786 (2015).

2. Keen, E.C. A century of phage research: bacteriophages and the shaping of modern biology. Bioessays 37, 6-9 (2015).

3. Ross, A., Ward, S. & Hyman, P. More Is Better: Selecting for Broad Host Range Bacteriophages. Front. Microbiol. 7 (2016).

4. de Jonge, P.A., Nobrega, F.L., Brouns, S.J.J. & Dutilh, B.E. Molecular and Evolutionary Determinants of Bacteriophage Host Range. Trends Microbiol. 27, 51-63 (2019).

5. Ackermann, H.W. Bacteriophage taxonomy. Microbiol. Aust. 32, 90-94 (2011).

6. World Health Organization. Antimicrobial resistance (2021); https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

7. Dedrick, R.M. et al. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. Nat. Med. 25, 730-733 (2019).

8. Guerrero-Bustamante, C.A. et al. Toward a Phage Cocktail for Tuberculosis: Susceptibility and Tuberculocidal Action of Mycobacteriophages against Diverse Mycobacterium tuberculosis Strains. mBio 12, e00973-00921 (2021).

9. Jordan, T.C. et al. A Broadly Implementable Research Course in Phage Discovery and Genomics for First-Year Undergraduate Students. mBio 5, e01051-01013 (2014).

10. Hatfull, G.F. Actinobacteriophages: Genomics, Dynamics, and Applications. Annu. Rev. Virol. 7, 37-61 (2020).

11. Hatfull, G.F. et al. Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. J. Mol. Biol. 397, 119-143 (2010).

12. Bárdy, P., Pantůček, R., Benešík, M. & Doškař, J. Genetically modified bacteriophages in applied microbiology. J. Appl. Microbiol. 121, 618-633 (2016).

13. Third international conference "Bioinformatics: from Algorithms to Applications" (BiATA 2019). BMC Bioinformatics 20, 516 (2019).

14. Lawrence, J. DNA Master. 5.23.5 ed. University of Pittsburgh, Pittsburgh, PA; 2020.

15. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. 215, 403-410 (1990).

16. Cresawn, S.G. et al. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12, 395 (2011).

17. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32, 11-16 (2004).

18. Russell, D.A. & Hatfull, G.F. PhagesDB: the actinobacteriophage database. Bioinformatics 33, 784-786 (2020).

19. Pope, W.H., Jacobs-Sera, D., Russel, D.A., Cresawn, S.G. & Hatfull, G.F. SEA-PHAGES Bioinformatics Guide. University of Pittsburgh; 2017.

20. Delcher, A.L., Bratke, K.A., Powers, E.C. & Salzberg, S.L. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23, 673-679 (2007).

21. Besemer, J. & Borodovsky, M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 33, W451-W454 (2005).

22. Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. GenBank. Nucleic Acids Res. 44, D67-D72 (2016).

23. Söding, J., Biegert, A. & Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244-W248 (2005).

24. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. 25, 955-964 (1997).

25. Xu, J., Hendrix, R.W. & Duda, R.L. Conserved Translational Frameshift in dsDNA Bacteriophage Tail Assembly Genes. Mol. Cell 16, 11-21 (2004).

26. Starterator. http://phages.wustl.edu/starterator/.

27. Bailly-Bechet, M., Vergassola, M. & Rocha, E. Causes for the intriguing presence of tRNAs in phages. Genome Res. 17, 1486-1495 (2007).

28. Toth, E.A., Li, Y., Sawaya, M.R., Cheng, Y. & Ellenberger, T. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. Mol. Cell 12, 1113-1123 (2003).

29. Hendrix, R.W., Hatfull, G.F. & Smith, M.C.M. Bacteriophages with tails: chasing their origins and evolution. Res. Microbiol. 154, 253-257 (2003).

30. Salisbury, A. & Tsourkas, P.K. A Method for Improving the Accuracy and Efficiency of Bacteriophage Genome Annotation. Int. J. Mol. Sci. 20, 3391 (2019).

31. Gibson, S.B. et al. Constructing and Characterizing Bacteriophage

Libraries for Phage Therapy of Human Infections. Front. Microbiol. 10 (2019).

32. Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T. & Jacobs Jr, W.R. Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. Mol. Microbiol. 4, 1911-1919 (1990).

33. Li, Q., Ge, F., Tan, Y., Zhang, G. & Li, W. Genome-Wide Transcriptome Profiling of Mycobacterium smegmatis MC^2 155 Cultivated in Minimal Media Supplemented with Cholesterol, Androstenedione or Glycerol. Int. J. Mol. Sci. 17, 689 (2016).

34. Allué-Guardia, A., Saranathan, R., Chan, J. & Torrelles, J.B. Mycobacteriophages as Potential Therapeutic Agents against Drug-Resistant Tuberculosis. Int. J. Mol. Sci. 22, 735 (2021).

35. Moon, K. et al. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. Microbiome 8, 75 (2020).

36. Enault, F. et al. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. ISME J. 11, 237-247 (2017).

Submitted: 01/25/2022
Accepted: 04/05/2055

Research Article

Emily Chen[1]

[1]Department of Geography, McGill University, Montreal, QC, Canada

Assessing Early Public Response to COVID-19-Related Restrictions in New York City Using Spatial Analysis of Urban Mobility Data

Keywords

Spatial mobility, COVID-19, demographic, geographic information science (GIS), pandemic

Email Correspondence

emily.chen3@mail.mcgill.ca

Abstract

The rapid spread of COVID-19 in the United States initiated shelter-in-place policies that significantly impacted human mobility and daily routines starting in March 2020. Prior literature has examined the differences in lockdown policy efficacy and compliance with government orders[1-6], as well as the effect of mobility changes on case counts[7-12]. However, less attention has been placed on the connection between mobility and socio-demographics after the onset of COVID-19 within a city's borders. This paper focused on how human mobility patterns in New York City during the first three months of the pandemic differed based on socio-demographic factors like age, household income, and method of transportation to work. A secondary analysis determined if the four measurements of mobility used, namely distance traveled from home, home dwell time, non-home dwell time, and percentage time home, yielded significantly different findings. A mobility ratio representing the change in mobility between the first two weeks of February and April 2020 was created using aggregated and anonymized cellphone mobility data from SafeGraph. A Global Moran's Index was calculated for each mobility ratio to test for the presence of spatial autocorrelation, and then two spatial lag models were applied to account for the existence of autocorrelation. That there existed significant differences in mobility patterns based on socio-demographics reinforced the need for physical distancing policies that acknowledge the demographic diversity present not only between but also within cities.

Introduction

Since the United States detected its first case of the 2019 novel coronavirus in January 2020, efforts to contain the virus, such as stay-at-home policies, have greatly restricted human mobility and upended daily routines and momentous occasions alike. This retroactive analysis of the interaction between human mobility patterns during the COVID-19 pandemic, particularly after the implementation of state-level shelter-in-place orders and the socio-demographic differences within a city, contributes to a rapidly growing body of literature examining the effectiveness of lockdown policies. Prior work has investigated the effect of virus mitigation measures[1-6], mobility[7-12], and public gatherings[13-15] on the COVID-19 case positivity growth rate at various geographical scales. This paper focuses specifically on the relationships between average weekly levels of mobility and population demographics within New York City census block groups (CBGs) from February to April 2020. This work aims to provide fine-grained analysis on the socio-demographic effects of lockdown measures for policymakers and inform future strategies for infection mitigation and safe re-opening. To accomplish this goal, this paper raises two research questions:

- ***Research Question 1:*** Which socio-demographic factors have the greatest effect on the change in population mobility in New York City (NYC) before and after the implementation of COVID-19-related lockdown measures in March 2020?
- ***Research Question 2:*** Of the variables measuring the change in population mobility in this research, which one(s) act(s) most robustly as a proxy for physical distancing adherence?

Analyzing population movement to glean human behavior patterns from aggregated smartphone data became increasingly common leading up to the outbreak of COVID-19[16,17]. In the earliest months of the pandemic, several researchers advocated for the analysis of mobile phone surveillance data to predict the spread of COVID-19 and to understand population mobility trends[18]. Academic and industry researchers from wide-ranging disciplines and around the world acted upon these sentiments, producing a staggering number of analyses on spatial mobility trends during the COVID-19 pandemic.

Several studies examined the effects of mobility reduction on case counts outside of the United States[11,19]. In a comprehensive review focused on the geospatial and spatial-statistical analysis of the pandemic, Franch-Pardo et al.[20] evaluated 63 scientific articles on the subject and concluded that interdisciplinary action, proactive planning, and international solidarity were of utmost importance in controlling the spread of COVID-19. One notable paper by Pullano et al.[21] provided a robust overview of the demographic, socioeconomic, and behavioral factors associated with decreased mobility in France prior to and during the early lockdown period in March 2020 based on data from aggregated cellphones.

Several studies have focused on the relationship between mobility and the spread of COVID-19 in the United States[3, 22-24]. Chang et al.[8] sought to understand how the COVID-19 spread in ten of the largest U.S. metropolitan areas by constructing fine-grained dynamic mobility networks derived from geolocation data that mapped the hourly movements of 98 million people from neighborhoods to points of interests between March and May 2020. The authors found that their model simulating the spread of COVID-19 accurately predicted that higher infection rates occurred during the first two months of the pandemic amongst disadvantaged racial and socioeconomic groups because of differences in mobility[8]. Work by Badr et al.[7] investigated the effect of large-scale social distancing adherence on the spread of COVID-19 in 25 U.S. counties with the highest number of confirmed cases as of mid-April 2020. In their analysis, the authors concluded that social distancing had a significant effect on the spread of COVID-19 and that their findings could translate to other U.S. locations, given the geographical diversity of the counties in their sample set.

Within NYC, Lamb et al.[12] conducted an ecological study of residents using data for the number of daily visits to points of interest (POIs). The authors found that the proportion of the population living in households with more than three inhabitants, the proportion of uninsured 18-64-year-olds, the proportion of the population self-identifying as White, and median household income were the four aggregate markers of socioeconomic status that yielded the highest R^2 value across four time periods in April 2020. Their analyses revealed that changes in mobility and SES markers explained 56% of the variability in case positivity through 1 April 2020, but then dropped to a rate of explanation for case positivity variability of just 18% by 30 April 2020.

These findings suggest that after COVID-19 cases peaked on 6 April 2020 in NYC, the SES markers became less predictive due to greater testing capacity, higher SES areas having lower case positivity due to potentially greater engagement with unwarranted testing, and lower SES areas containing a higher number of infections. The authors also found that increased case positivity were independently associated with greater reductions in mobility on 10 April and 20 April but not on 1 April and 30 April. They attributed these mixed findings to the correlation between time and a city-wide decrease in case positivity as testing capacities increased.

Methods

Data

This paper's area of interest is NYC because it was the epicenter of the COVID-19 outbreak in the United States, with approximately 203,000 laboratory-confirmed cases reported between 1 March and 31 May 2020[25]. On 16 March 2020, the NYC school system, gyms, and casinos closed, and restaurants and bars were restricted to take-out and delivery services[26]. On 22 March 2020, all non-essential businesses closed, and the NYC on Pause Program's stay-at-home orders went into effect[27]. Building off these key dates, February 2020 was identified as the "before" time period and April 2020 as the "after" time period for analysis. Mobility patterns were retrieved from the "Social Distancing Metrics" dataset provided by the place-based data collection platform SafeGraph[28]. SafeGraph collects data using GPS pings from 20 million anonymous cellphone devices across the US. To calculate a mobile device's home, SafeGraph determines the device's common nighttime location to a Geohash-7 granularity of about 153 meters by 153 meters, and then groups devices into "home" CBGs. It also provides aggregated data, every 24-hours, for each CBG[28]. Table 1 describes the mobility variables and how they are reported by SafeGraph. To compare the differences in distance traveled from home before and after the onset of COVID-19 in NYC, the median distance traveled from home in the first two weeks of February 2020 for each CBG was divided by the median mobility value in the first two weeks of April 2020 for the equivalent CBG to create a mobility ratio (MR). The process was repeated to compare the differences in median home dwell time, median non-home dwell time, and median percentage time home. The latest available socioeconomic and demographic data was accessed from the 2016 5-year estimates in the American Community Survey (ACS)[29]. ACS data at the CBG level were the highest resolution available for the selection of socioeconomic and demographic variables. These data were cleaned to remove null and erroneous values.

Analysis

Global and local regression models were computed the same way for all four mobility ratios. Table 2 describes the explanatory variables used in all regression models. First, an ordinary least squares (OLS) linear regression model was fitted to the data to determine the global relations between mobility and socio-demographic factors. Next, the Global Moran's Index correlation test for regression residuals was used to check for spatial autocorrelation.

Table 1: Dependent Variables Used in Regression Models

Variable Name	Metadata
distance_traveled_from_home	Reported as an Integer. The value represents the median distance (in meters) of the median distance per device in a CBG traveled from the device's calculated "home" (i.e. Geohash-7 common nighttime location) within a 24-hour period. SafeGraph excluded distances equal to 0.
median_home_dwell_time	Reported as an Integer. The value represents the median time (in minutes) of the sum of all total time per device in a CBG spent at the device's Geohash-7 common nighttime location within a 24-hour period. Included in the total time are time ranges that may or may not have stopped or started within the 24-hour period.
median_non_home_dwell_time	Reported as an Integer. The value represents the median time (in minutes) of the sum of all total time per device in a CBG spent outside of the device's Geohash-7 common nighttime location within a 24-hour period.
median_percentage_time_home	Reported as an Integer. The value represents the ratio between median percentage of time spent at "home" for all devices in a CBG and the median total time observed within a 24-hour period.

Table 2: Explanatory Variables Used in Regression Models

Variable Name	Metadata
age	Estimated median age of the population.
race	Estimated number of people who identify as only White.
transport	Estimated number of workers 16 years and older who use public transportation (excluding taxicabs) to travel to work.
female_workers	Estimated number of female workers 16 years and older
housing_occupancy_rent	Estimated number of renter occupied housing units with over 1.5 occupants per room.
min_wage	Estimated number of households that earned less than $25,000 a year in 2016 (accounting for inflation) *Note: The base minimum wage in New York City from 12/31/15 to 12/31/16 was $9.00/hour, which is equivalent to $18,000/year [17].*
children	Estimated number of families with children under the age of 18.
education	Estimated number of people 25 years and older with a regular high school diploma.
health_insurance	Estimated number of people from the civilian non-institutionalized population with no health insurance coverage.

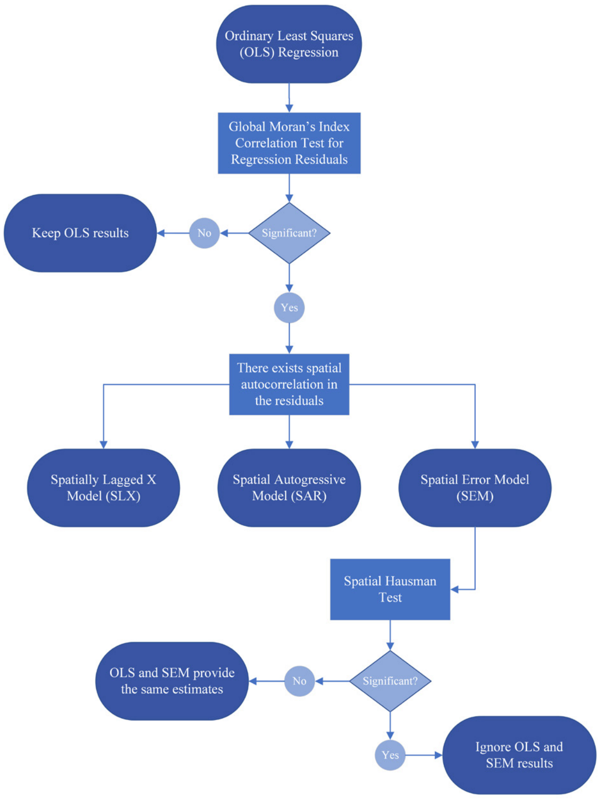


Figure 1: Spatial Regression Decision Process Flowchart

Two spatial regression models were run that determined whether the mobility patterns in surrounding CBGs affected the mobility pattern in one CBG[33]. The Spatially Lagged X (SLX) model tested local spatial relations, which meant that surrounding CBGs were those immediately adjacent to a CBG. The spatial autoregressive (SAR) Spatial Lag model tested global spatial relations, which meant that surrounding CBGs were all the observations in the data. To summarize the impacts from the SAR models, the number of simulations was set to 5,000 to compute distributions for the impact measures.

Lastly, a Spatial Error Model (SEM) and a spatial Hausman test were used to detect predictor variables in a regression model and were run to determine if differences existed between the OLS and SEM coefficients. A significant result suggested that neither OLS nor SEM yielded regression parameter estimates that matched the linear model parameters with independent identically distributed disturbances[34]. Thus, if a significant result was obtained from the spatial Hausman test, the OLS and SEM results were not considered. Figure 1 illustrates the analysis flow explained in this section. All analyses were performed using the 3.6.2 version of the R programming language[35] in version 1.2.5033 of RStudio[36].

Mapping

To create a map for each mobility ratio, R was used to remove outlier data by excluding the CBGs whose change in median home dwell time were greater than 2. Since “Null” values were changed to -999 for data parsing purposes, ratios that were less than 0 were also excluded. The categories were delineated by natural bins.

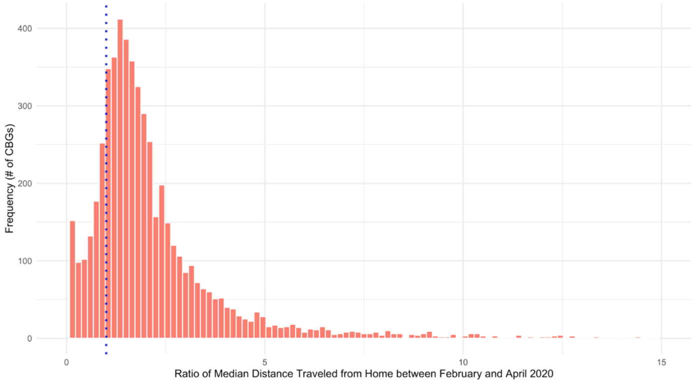


Figure 2a: Histogram for the Mobility Ratio of Median Distance Traveled from Home in NYC between February and April 2020 (bin size = 0.15). CBGs to the right of the dotted blue line at x = 1 indicate those with residents who traveled greater distances from home in February compared to April.

Figure 2b: Histogram for the Mobility Ratio of Median Home Dwell Times in NYC between February and April 2020 (bin size = 0.15). CBGs to the right of the dotted blue line at x = 1 indicate those with residents who stayed at home for longer in February compared to April.

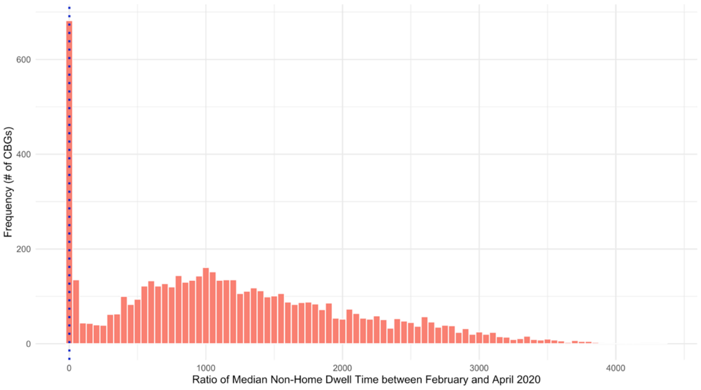


Figure 2c: Histogram for the Mobility Ratio of Median Non-Home Dwell Times in NYC between February and April 2020 (bin size = 50). CBGs to the right of the dotted blue line at x = 1 indicate those with residents who spent more time away from home in February compared to April.

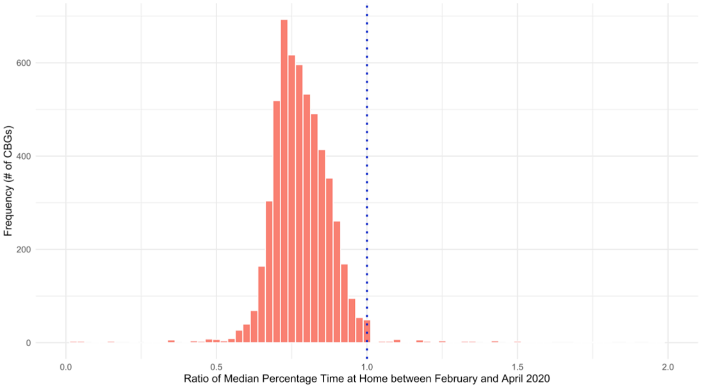


Figure 2d: Histogram for the Mobility Ratio of Median Percentage Time at Home in NYC between February and April 2020 (bin size = 0.025). CBGs to the right of the dotted blue line at x = 1 indicate those with residents who stayed at home for a higher percentage of time in February compared to April.

Shapefiles from the United States Census Bureau have cartographic boundary levels at the 2020 CBG level for each state. However, the NYC Department of City Planning provides shapefiles for the NYC boundary at only the 2010 census block level, which is at an even higher resolution than the CBG level. To obtain a shapefile with NYC CBGs, ArcMap v.10.7.1[37] was used to reproject both the NYC 2010 census block shapefile and the NY 2020 CBG shapefile to the WGS 1984 UTM Zone 18N coordinate system. The NYC census block shapefile was dissolved into census block groups and then intersected with the NY 2020 CBG shapefile. Lastly, R was used to link the shapefile with the CSV file containing SafeGraph and ACS data.

Results

Descriptive Statistics and Spatial Visualization

The frequency distributions for all four mobility ratios indicate that, overall, most NYC CBGs experienced decreased mobility and more time spent at home in the first two weeks of April 2020 compared to the first two weeks of February 2020. Based on the mobility ratio calculations, a ratio value less than 1 for a CBG suggested that people in that CBG traveled farther if they left home, stayed at home for longer, spent less time outside

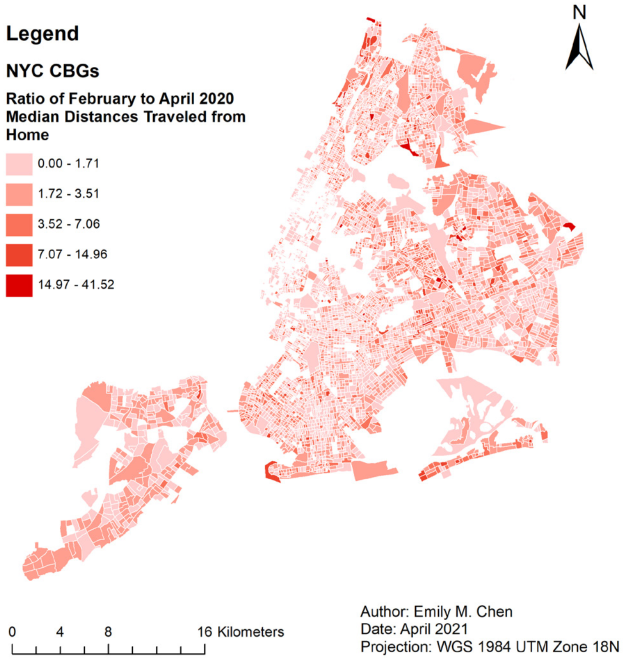


Figure 3a: Change in Median Distance Traveled from Home between February and April 2020 in New York City

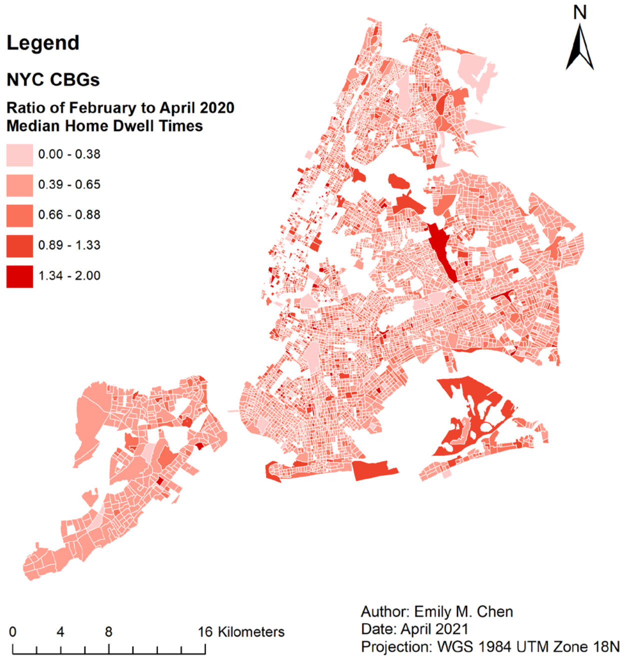


Figure 3b: Change in Median Home Dwell Time between February and April 2020 in New York City

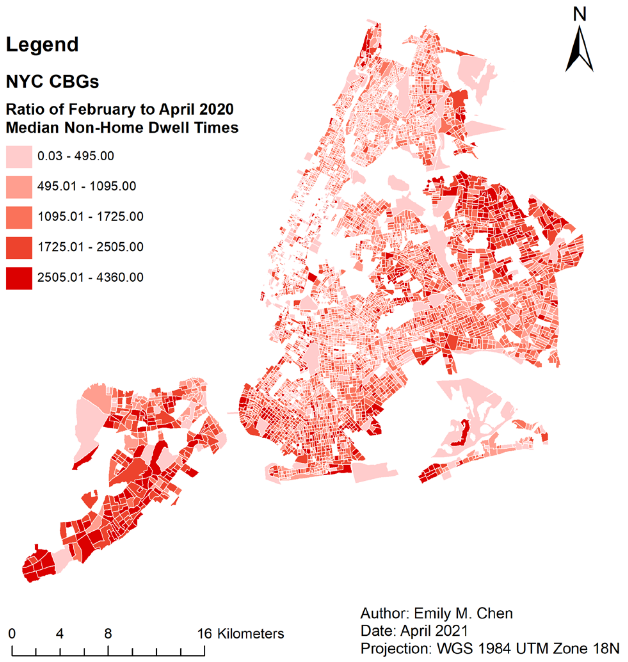


Figure 3c: Change in Median Non-Home Dwell Time between February and April 2020 in New York City

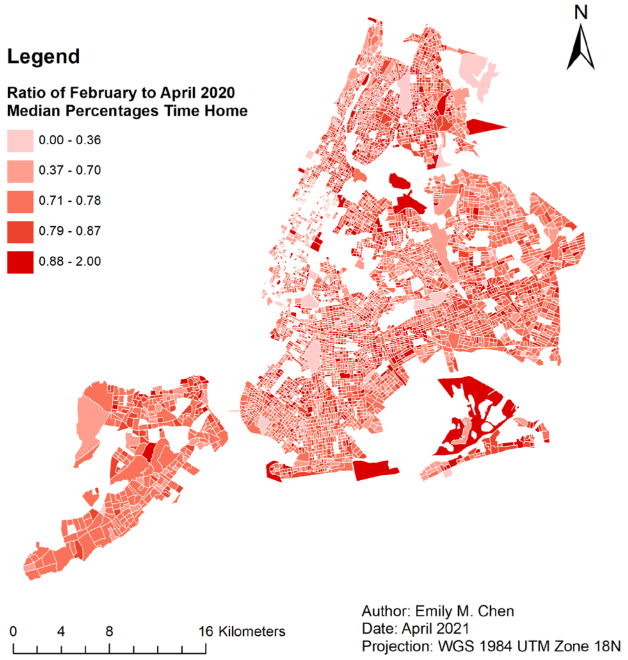


Figure 3d: Change in Median Percentage Time at Home between February and April 2020 in NYC 2020 in New York City

of home, or spent a greater percentage of their time at home in April than in February, and vice versa for a ratio value greater than 1. For distance traveled from home (Figure 2a), the histogram shows that most of the ratio values fall between 0 and 5, but with more values greater than 1 (M_{ratio}=1.6178). Thus, distances traveled from home dropped in April 2020. For home dwell time (Figure 2b), almost all ratio values fall between 0 and 1 (M_{ratio}=0.6170), so median home dwell times across all CBGs were mostly greater in April than in February. Conversely, for non-home dwell time (Figure 2c), most of the values are much greater than 1 (M_{ratio}=1085.0000), so median non-home dwell times across all CBGs were mostly greater in February than in April. Since all values of "0" were changed to "0.1" during data pre-processing to avoid dividing by 0, a median non-home dwell time value of 0.10 minutes in April was interpreted to mean that a large majority of CBGs experienced essentially no time spent away from home. A majority of April non-home dwell times close to 0 led to larger ratio values for median non-home dwell time compared to the other dependent variables. Lastly, for percentage time at home (Figure 2d), most of the ratio values lie between 0.5 and 1 (M_{ratio}=0.7750), therefore the median percentages of time spent at home across all CBGs were mostly greater in April than in February.

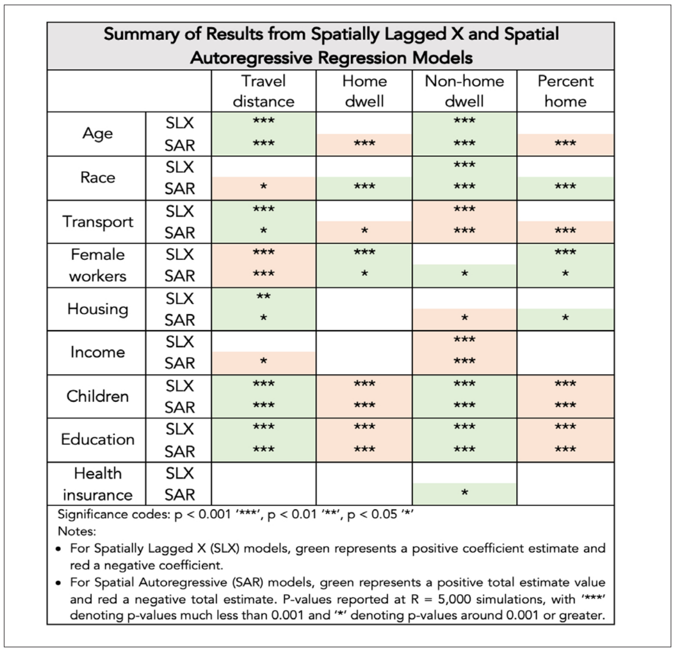
The maps of mobility ratio values for each CBG illustrate spatial variability across NYC. Category cut-off values were determined by natural breaks. For distance traveled from home (Figure 3a), darker red corresponds to a greater difference in median distance traveled from home between February and April. CBGs in the four largest categories had about a two-fold or greater increase in median travel distance from February to April. For home dwell time (Figure 3b), most CBGs had ratio values less than 1.0, which meant more time spent at home in April than in February. For non-home dwell time (Figure 3c), the enormous range in values for the smallest category suggests most CBGs had populations that spent almost no time away from home in April. Lastly, for percentage time at home (Figure 3d), most values were less than 1.0, thus showing that most CBGs experienced a greater percentage of time at home in April than in February.

RQ1: Effects of Socio-Demographic Factors on Mobility

Our first research question sought to understand which socio-demographic factors had the most effect on the change in population mobility in New York City before and after the implementation of COVID-19-related lockdown measures in March 2020. Nine noncollinear explanatory variables were chosen and four regression models were run with the four different measurements of change in mobility from February to April 2020: change in median distance traveled from home, change in median home dwell time, change in median non-home dwell time, and change in median percentage of time spent at home.

Based on the difference between the observed and expected Moran's I value as well as the significant p-value ($p<0.001$ for $\alpha=0.05$) for each of the Global Moran's Index linear correlation for regression residuals tests, we accepted the alternative hypothesis that there existed spatial autocorrelation in the residuals from all OLS model. Similarly, the significant p-value ($p<0.001$ for $\alpha=0.05$) obtained for all the spatial Hausman tests confirmed there were enough differences in the Standard Error Model (SEM) regression coefficients such that neither OLS nor SEM were appropriate models. Thus, only the Spatially Lagged X (SLX) and Spatial Autoregressive (SAR) models were used to interpret the results. Figure 4 summarizes the findings from these two models.

For the SLX model, a positive coefficient estimate associated with an explanatory variable meant that as the value for that variable within a CBG increased, so did the mobility ratio in that CBG (direct effect) and in neighboring CBGs (indirect effect). Median age had a positive direct and indirect (both $p<0.001$) value associated with distance traveled from home and non-home dwell time. Number of white-only residents had positive direct ($p<0.001$) and indirect ($p<0.05$) values for non-home dwell time. Number of families with children and high school graduates both had positive direct (both $p<0.01$) and indirect ($p<0.001$, $p<0.01$ respectively) values for distance traveled from home, negative direct and indirect (all $p<0.001$) values for home dwell time, positive direct (both $p<0.001$) and indirect ($p<0.05$, $p<0.001$ respectively) values for non-home dwell time, and negative direct and indirect (all $p<0.001$) values for percent time at home. Number of public transit users had positive direct ($p<0.01$) and indirect ($p<0.05$) values for distance traveled from home and negative direct ($p<0.001$) and indirect ($p<0.01$) values for non-home dwell time. Number of female workers had negative direct and indirect (both $p<0.001$) values for distance traveled from home and positive direct ($p<0.01$) and indirect ($p<0.001$) values for both home dwell time and percent time at home. Lastly, number of occupied renter units had positive direct and indirect (both $p<0.05$) values for distance traveled from home, while household income had negative direct and indirect (both $p<0.001$) values for non-home dwell time.



Summary of Results from Spatially Lagged X and Spatial Autoregressive Regression Models

		Travel distance	Home dwell	Non-home dwell	Percent home
Age	SLX	🟩 ***		🟩 ***	
	SAR	🟩 ***	🟥 ***	🟩 ***	🟥 ***
Race	SLX			🟩 ***	
	SAR	🟥 *	🟩 ***	🟩 ***	🟩 ***
Transport	SLX	🟩 ***		🟥 ***	
	SAR	🟩 *	🟥 *	🟥 ***	🟥 ***
Female workers	SLX	🟥 ***	🟩 ***		🟩 ***
	SAR	🟥 ***	🟩 *	🟩 *	🟩 *
Housing	SLX	🟩 **			
	SAR	🟩 *		🟥 *	🟩 *
Income	SLX			🟥 ***	
	SAR	🟥 *		🟥 ***	
Children	SLX	🟩 ***	🟥 ***	🟩 ***	🟥 ***
	SAR	🟩 ***	🟥 ***	🟩 ***	🟥 ***
Education	SLX	🟩 ***	🟥 ***	🟩 ***	🟥 ***
	SAR	🟩 ***	🟥 ***	🟩 ***	🟥 ***
Health insurance	SLX				
	SAR			🟩 *	

Significance codes: $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*'

Notes:

- For Spatially Lagged X (SLX) models, green represents a positive coefficient estimate and red a negative coefficient.
- For Spatial Autoregressive (SAR) models, green represents a positive total estimate value and red a negative total estimate. P-values reported at R = 5,000 simulations, with '***' denoting p-values much less than 0.001 and '*' denoting p-values around 0.001 or greater.

Figure 4: Summary of Results from the Spatially Lagged X and Spatial Autoregressive Models

Interpretation of the SAR model relies on the impact measures' p-values and the direction of the direct impact value. Median age had positive direct impact values and consistently significant simulated p-values ($p<0.001$ for every run) for distance traveled from home and non-home dwell time, and negative values for home dwell time and percent time at home. Number of white-only residents had positive values for home dwell time, non-home dwell time, and percent time at home. Number of families with children and number of high school graduates had positive values for distance traveled from home and non-home dwell time and negative values for home dwell time and percent time at home. Number of public transit users had negative values for non-home dwell time and percent time at home. Lastly, number of female workers and household income both had negative values for distance traveled from home and non-home dwell time.

A caveat for the strength of the findings is that the SLX multiple R^2 values, while larger than the OLS multiple R^2 values for each dependent variable, were still quite low despite including nine explanatory variables (R^2=0.048 for distance traveled from home, R^2=0.077 for home dwell time, R^2=0.162 for non-home dwell time, and R^2=0.081 for percentage time spent at home). These low R^2 values indicate that the proportions of the variance in the dependent variables predictable from the explanatory variables were quite low. Solutions for increasing the R^2 value in future research include using other data sources and adding more explanatory variables. Importantly, since the R^2 value is not an indicator of whether the independent variables cause changes in the dependent variable, the interpretations of which explanatory variables affect mobility remain valid.

RQ2: Mobility Variables as a Proxy for Physical Distancing Adherence

This second research question asked which of the variables measuring population mobility served most robustly as a proxy for physical distancing adherence. To propose possible answers, the results from the four regression models were examined in the context of the nine explanatory variables for each of the mobility measures and found that median non-home dwell time yielded the greatest number of significant correlations with the explanatory variables from the SLX and SAR models. Furthermore, the SLX model's R^2 value (0.162) with this dependent factor was the highest of all mobility measures.

Discussion

Implications of Findings

Several of the results from the SLX and SAR models have interesting implications. Since non-home dwell time was the most accurate proxy for adherence to physical distancing, interpretations for some of the socio-demographic effects on this mobility measurement are presented here. For example, the SLX and SAR models found that median age of a CBG correlated positively with the change in median non-home dwell time (both $p<0.001$). This result indicated that the older the median age of a CBG, the less likely its residents were to spend time away from home in April. That CBGs with older populations saw less time spent away from homes suggests that older people were particularly careful about staying at home due to a combination of retirement, fewer reasons to leave the home, and knowledge that the elderly were affected more severely by the disease compared to younger populations[38-40]. Additionally, the scale of the SLX and SAR direct, indirect, and total impacts for age with non-home dwell times as the dependent variable were much higher than for any of the other explanatory variables (SLX: 14.52, 11.11, and 25.63 for age (respectively) compared to impact measures > -2.58 and < 2.12 for all other variables; SAR: 15.08, 6.50, 21.58 for age (respectively) compared to impact measures > -2.07 and < 1.31 for all other variables). These results indicate that a higher median age within a CBG had a greater effect on mobility defined as home and non-home dwell time than the other explanatory variables within that CBG (direct impact), in the CBG's immediate neighbors (indirect impact), and in all CBGs in the data (total impact)[41].

Another strong finding from the SLX and SAR models concerned the estimated number of families with children under age 18, which correlated positively with the change in median non-home dwell time (both $p<0.001$). These results indicated that CBGs with a greater number of families with children experienced less time spent away from home in April. This finding makes sense given that once schools closed, many parents stayed home to take care of young children while juggling full-time jobs. School closures and uncertainty about childcare left parents, particularly working mothers, with home school responsibilities that prompted some mothers to leave their jobs entirely[42]. Research by the U.S. Census Bureau and Federal Reserve found that of the adults not working, women ages 25-44 were almost three times as likely as men (32.1% compared to 12.1%) to not be working due to childcare demands[43]. While these results are based on national data, this phenomenon likely extended to NYC families as well. Furthermore, the U.S. Census study also found that working mothers in states with early stay-at-home orders and school closures were 68.8% more likely to leave their jobs than working mothers in states with later closures[44]. Given that NY state was one of the first states to implement stay-at-home measures, it seems likely that NYC working mothers fit into the category of being more likely to leave their jobs.

Lastly, both the SLX and SAR models found that the estimated number of people with only a high school diploma correlated positively with the change in median non-home dwell time (both $p<0.001$), showing that CBGs with a greater number of high school graduates spent more time at home in April compared to in February. One explanation for this finding was the 15% seasonally adjusted unemployment rate in April and that those more likely to face unemployment due to COVID-19 in NYC were workers with lower educational attainment (i.e., without a bachelor's degree)[44]. As confirmation, 61% NYC adults without a bachelor's degree experienced a loss in income since 13 March 2020 compared with 45% of adults with more than a bachelor's degree[44]. Without a job to go to, this demographic traveled shorter distances and stayed at home for longer periods of time.

Research Limitations

Similar to prior literature using aggregated cellphone mobility data[3, 8, 12, 22-24], the unknown representativeness of SafeGraph's data made it challenging to draw definitive conclusions from regression models. The dataset is certainly one of the largest available, as it came from 500,000 devices in almost every NYC CBG and accounted for one-ninth of the NYC population, which is a staggeringly large sample size compared to early mobility research that relied on participants to self-report data. However, this limitation is still worth noting because any conclusions drawn from these findings must acknowledge that they illustrate general population mobility trends from aggregated data. Running the four regression models using different mobility datasets and comparing the results could also strengthen these findings.

A second limitation to this work was the potential for additional factors, besides stay-at-home restrictions, to influence mobility patterns. For example, warmer weather in April could have contributed to greater time spent away from home for some demographics. To account for this seasonal change, an alternative baseline could have been April 2019, assuming that weather patterns were similar at that time to those observed in April 2020.

Future Directions

There are several ways to build upon the findings in this paper. The first is to extend the methodology to data from other cities, both in the U.S. and internationally. A between-city comparison might provide greater insight into how stay-at-home policies affected regions differently based on socio-demographic patterns, public transit infrastructure, or population density. An international comparison of cities could yield insight into the extent to which government stay-at-home orders reduced population mobility compared with other cross-cultural factors. In addition to comparing cities, other explanatory factors could be added to the regression models, such as the number of households who own second homes or citizenship status. Instead of mobility variables, one could also use points of interest (POI) data. For example, to better determine the large-scale impact of age, it would be useful to understand where younger people were going. National data indicated that younger workers were more likely to face unemployment due to COVID-19, and a survey of NYC metro adults found that 56% had lost income during the pandemic[44]. Therefore, if younger workers were more likely to experience unemployment and 37% of NYC frontline workers are over 50 years old[45], where were the younger age groups going? In addition to POI data, this question could be answered by using age-bracketed data to determine which age group left home the most. Lastly, several types of datasets that could be used to cross-reference our findings and evaluate how other non-pharmaceutical interventions affected mobility. For example, the Delphi Group at Carnegie Mellon University provides a variety of real-time COVID-19 indicators at the U.S. County and state level. Comparing their data on vaccine acceptance or the proportion of mask-wearers with mobility trends at the county level could help illuminate other aspects of disease spread patterns. Exploring other human behavior indicators and non-pharmaceutical interventions has particularly important implications, as researchers found that mobility and infection rates did not positively correlate as strongly after April 2020[46]. Their findings suggest that other non-pharmaceutical interventions like mask-wearing or hand washing played a significant role in mitigating the spread of COVID-19 early in the pandemic, therefore future research should consider these factors in their models when exploring the relationship between mobility and case positivity.

Conclusion

This paper's intent was to provide fine-grained analysis on the varying effects of lockdown measures and to inform future strategies for infection mitigation and safe re-opening. Our findings that there exist significant differences in mobility based on socio-demographic factors, particularly age, education level, and whether families have children, reinforce the need for physical distancing policies that acknowledge the demographic diversity present not only within, but also between cities. Providing resources for populations less able to stay at home (e.g., healthcare workers, service workers) to safely continue working is as important as providing support for populations who end up needing to stay at home (e.g., parents of young children, elderly populations) to minimize the effects of increased childcare demands and isolation. Future research may examine both these findings and the implications of reduced mobility on the spread of COVID-19 compared with other non-pharmaceutical interventions. By providing a detailed analysis of the various socio-demographic effects on different measurements of mobility, this paper emphasizes that there are several ways to measure mobility patterns within a city and that stay-at-home policies introduce unevenly distributed effects to different groups.

Acknowledgments

This research was conducted as part of the author's honours work towards a Bachelor of Arts (Honours) in Urban Systems Geography in the Department of Geography. The author wishes to thank Dr. Grant McKenzie (McGill University) for his guidance throughout the research process as the thesis supervisor, Dr. Clio Andris (Georgia Institute of Technology) for her insight and edits as the thesis reader, and Dr. Sarah Turner (McGill University) for her comments on the work as the professor for GEOG381: Geographic Thought and Practice and GEOG491: Honours Research.

Statement on Open Science

The author is committed to contributing towards the replicability and reproducibility of scientific research. A repository with all aspects of the data collection, cleaning, and analysis processes exists at https://github.com/emilyemchen/covid19mobility. The repository's README.md file provides a broad overview of the file structure and contents.

References

1. Aleta, A. et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour* **4**, 964–971 (2020).

2. Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* **395**, 931–934 (2020).

3. Gao, S. et al. Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. *JAMA Network Open* **3**, e2020485 (2020).

4. Graff Zivin, J. & Sanders, N. The spread of COVID-19 shows the importance of policy coordination. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 32842–32844 (2020).

5. Holtz, D. et al. Interdependence and the cost of uncoordinated responses to COVID-19. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 19837–19843 (2020).

6. Pei, S., Kandula, S. & Shaman, J. Differential effects of intervention timing on COVID-19 spread in the United States. *Science Advances* **6**, eabd6370 (2020).

7. Badr, H. S. et al. Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases* **3099**, 1–8 (2020).

8. Chang, S. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2020).

9. Chen, Y., Jiao, J., Bai, S. & Lindquist, J. Modeling the spatial factors of COVID-19 in New York City. *SSRN Electronic Journal*. https://ssrn.com/abstract=3606719 (2020).

10. Cronin, C. J. & Evans, W. N. Private precaution and public restrictions: What drives social distancing and industry foot traffic in the COVID-19 era? *NBER Working Paper Series*, 1–39 (2020).

11. Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).

12. Lamb, M. R., Kandula, S. & Shaman, J. Differential COVID-19 case positivity in New York City neighborhoods: Socioeconomic factors and mobility. *Influenza and Other Respiratory Viruses* **15**, 209–217 (2021).

13. Chande, A. et al. Real-time, interactive website for US-county-level COVID-19 event risk assessment. *Nature Human Behaviour* **4**, 1313–1319 (2020).

14. Dave, D., Friedson, A., McNichols, D. & Sabia, J. The contagion externality of a superspreading event: The Sturgis Motorcycle Rally and COVID-19. *Southern Economic Journal* **87**, 1–39 (2020).

15. Mangrum, D. & Niekamp, P. JUE Insight: College student travel contributed to local COVID-19 spread. *Journal of Urban Economics* **127**, 103311 (2022).

16. Budd, J. et al. Digital technologies in the public-health response to COVID-19. *Nature Medicine* **26**, 1183–1192 (2020).

17. Smith, C. M. et al. Spatial methods for infectious disease outbreak investigations: Systematic literature review. *Eurosurveillance* **20**, 1–21 (2015).

18. Buckee, C. O. et al. Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).

19. Saraswathi, S., Mukhopadhyay, A., Shah, H. & Ranganath, T. S. Social Network Analysis of COVID-19 Transmission in Karnataka, India. *Epidemiology and Infection* **148**, 1–10 (2020).

20. Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F. & Billa, L. Spatial analysis and GIS in the study of COVID-19. A review. *Science of the Total Environment* **739**, 140033 (2020).

21. Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: A population-based study. *The Lancet Digital Health* **2**, e638–e649 (2020).

22. Bian, B., Li, J., Xu, T. & Foutz, N. Individualism During Crises. *The Review of Economics and Statistics* **104**, 1–18 (2022).

23. Brzezinski, A., Kecht, V. & Van Dijcke, D. The cost of staying open: Voluntary social distancing and lockdowns in the US. *SSRN Electronic Journal*, 1–34. https://ssrn.com/abstract=3614494 (2020).

24. Weill, J. A., Stigler, M., Deschenes, O. & Springborn, M. R. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 19658–19660 (2020).

25. Thompson, C. N. et al. COVID-19 Outbreak — New York City, February 29 – June 1, 2020. *Morbidity and Mortality Weekly Report* **69**, 1725–1729 (2020).

26. Vasquez, J., Shea, T., Rajamani, M., Price, B. & Intarasuwan, K. Timeline: Tracking the Spread of COVID-19 in Tri-State. https://www.nbcnewyork.com/news/local/timeline-tracking-the-spread-of-covid-19-in-tri-state/2313123/ (2020).

27. City of New York. Mayor de Blasio Issues New Guidance to New Yorkers. https://www1.nyc.gov/office-of-the-mayor/news/173-20/mayor-de-blasio-issues-new-guidance-new-yorkers (2020).

28. SafeGraph. *Social Distancing Metrics* https://docs.safegraph.com/docs/social-distancing-metrics (2021).

29. United States Census Bureau. *American Community Survey Information Guide* tech. rep. (2017), 1–15. https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf.

30. Taylor, D. B. A Timeline of the Coronavirus Pandemic. *The New York Times*. https://www.nytimes.com/article/coronavirus-timeline.html (2021).

31. Qin, A. & Hernández, J. C. China Reports First Death From New Virus. *The New York Times*. https://www.nytimes.com/2020/01/10/world/asia/china-virus-wuhan-death.html (2020).

32. New York State Department of Labor. *Minimum Wage Lookup* https://webapps.labor.ny.gov/dolweb/minimum-wage-lookup/ (2021).

33. Medina, J. & Solymosi, R. *Crime Mapping in R* chap. Spatial Regression Models. https://maczokni.github.io/crimemapping_textbook_bookdown/spatial-regression-models.html (2019)

34. Pace, R. K. & LeSage, J. P. A spatial Hausman test. *Economics Letters* **101**, 282–284 (2008).

35. R Core Team. *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2019). https://www.r-project.org/.

36. RStudio Team. *RStudio: Integrated Development Environment for R* (RStudio, Inc., Boston, MA, 2020). http://www.rstudio.com/.

37. Environmental Systems Research Institute. *ArcMap: Version 10.7.1* (Redlands, CA, 2019).

38. Mueller, A. L., McNamara, M. S. & Sinclair, D. A. Why does COVID-19 disproportionately affect older people? *Aging* **12**, 9959–9981 (2020).

39. Santesmasses, D. et al. COVID-19 is an emergent disease of aging. *Aging Cell* **19** (2020).

40. Williamson, E. J. et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436 (2020).

41. Golgher, A. B. & Voss, P. R. How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. *Spatial Demography* **4**, 175–205 (2016).

42. Bateman, N. & Ross, M. *Why has COVID-19 been especially harmful for working women?* tech. rep. (Brookings, 2020). https://www.brookings.edu/essay/why-has-covid-19-been-especially-harmful-for-working-women/.

43. Heggeness, M. L. & Fields, J. M. Working Moms Bear Brunt of Home Schooling While Working During COVID-19. https://www.census.gov/library/stories/2020/08/parents-juggle-work-and-child-care-during-pandemic.html (2020).

44. Nischan, U. *Uneven Impact: What Job Loss During COVID-19 Means for New Yorkers Now and into the Future* tech. rep. (New York City Department of Consumer and Worker Protection, New York City, 2020), 1–19.https://www1.nyc.gov/assets/dca/downloads/pdf/partners/Uneven_Impact.pdf.

45. Stringer, S. M. *New York City's Frontline Workers* tech. rep. (Office of the New York City Comptroller, New York City, 2020), 1–15. https://comptroller.nyc.gov/wp-content/uploads/documents/Frontline_Workers_032020.pdf.

46. Badr, H. S. & Gardner, L. M. Limitations of using mobile phone data to model COVID-19 transmission in the USA. *The Lancet Infectious Diseases* **3099**, 30861 (2020).

Submitted: 01/15/2022
Accepted: 03/12/2022

Research Article

Elena Frie[1], Saskia Gilmer[2], Bryan Buraga[3], and Kevin Franceschini[3]

Quantifying the Albedo of the Montreal Island and its Potential for Increase

[1]Department of Earth and Planetary Sciences, McGill University, Montreal, QC, Canada

[2]Department of Computer Science, McGill University, Montreal, QC, Canada

[3]Department of Arts and Sciences, McGill University, Montreal, QC, Canada

Keywords

Urban albedo, increasing surface albedo (ISA), urban land use

Email Correspondence

elena.frie@mail.mcgill.ca
saskia.gilmer@mail.mcgill.ca

Abstract

Urbanization has changed the Earth's surface, resulting in the urban heat island effect. There has been a recent focus on increasing urban albedo as a strategy to mitigate this phenomenon. Studies on Montreal's albedo have primarily looked at the impact of albedo manipulations upon the urban heat island effect. However, the current albedo of the island, broken down by land use type, has yet to be quantified. Therefore, previous studies often rely on generalized urban albedo and land use estimates that have not been proven to be generalizable to Montreal. This study attempted to quantify the current albedo of the Island of Montreal through urban land use categorization. The findings were then used to estimate albedo increase under different roof replacement scenarios. Data sets for building footprints, vegetation, and roadways were incomplete in Montreal, requiring the combination of several sources to obtain representative data for analysis. This study found the albedo of Montreal island to be 0.19 ± 0.057. Further, the hypothetical roof change scenarios then aligned with a 0.1 albedo increase , which is the albedo change used in current urban heat island effect mitigation literature. Using the albedo increase potential that resulted from the three scenarios tested here, future research should explore further estimation of the associated surface and air temperature decrease.

Introduction

Currently, 55% of the global population resides in urban areas, and the United Nations projects this proportion to increase to 68% by 2050[1]. There are many environmental benefits of increased urban population density. Sustainably planned cities can decrease energy consumption and greenhouse gas emissions per person through shared transport, living spaces, and food access[2]. However, there are clear threats that urban spaces pose to biodiversity, water and air quality, and the surrounding climate. Land use change due to urbanization produces the well-studied phenomena of the urban heat island (UHI) effect, where the temperature in an urban area will be on average 1-3°C warmer than the surrounding rural temperature[3]. The materials that make up city roadways and building roofs have a very low albedo and so absorb far more incident shortwave radiation than natural vegetation does. Further, these materials often have high heat capacities that cause the release of absorbed daytime thermal energy at night[4].

There has been focus on increasing urban albedo as a strategy for combatting the UHI effect. Albedo is the probability that a photon of solar irradiance is reflected on a surface, and can be assessed through the simple ratio of outgoing over incoming radiation[5]. Studies from across all urban regions relate albedo increases to reduced energy consumption, reduced electricity demand, improved air quality, reduced risk of heat-related discomfort and mortality, and changes to precipitation patterns (in non-snow cover regions/seasons)[6].

Generalized estimations find that rooftops and roads occupy 60% of urban spaces, with a 1.5 roof to road ratio[7]. The low albedo of these surfaces provides a potential for alteration of urban rooftops to produce changes to a city's climate. Greening of roofs is one strategy to increase albedo and offset heat through evapotranspiration. Whitening of urban surfaces has been studied as well[6]. Previous studies divide between focus on quantified albedo of urban materials, quantified albedo of total urban regions, and impacts of changes to urban albedo on the UHI effect.

Methods for quantifying albedo values vary. Prado and Ferreira[5] used an experimental spectrophotometer to find the albedos of rooftop types found in Sao Paulo, Brazil. Even with consideration for roughness and aging of materials, white surfaces consistently reported albedo values > 0.5. The Berkeley Lab Heat Island Group[8] found similar values using remote sensing techniques on satellite imagery of Californian cities.

Many studies explore the impacts of albedo increases on the UHI effect. To do so, they almost exclusively use the Weather Research and Forecasting-Urban Canopy model (WRF-UCM), a mesoscale numerical weather prediction model that simulates the urban area[6,9,10]. This WRF model uses a default albedo parameter of 0.2 that is applied to all urban area surfaces. In studies on Montreal and Toronto respectively, Jandaghian and Akbari[11] and Jandaghian and Berari[12] used this default albedo of 0.2, and tested the effects of increased albedo values of 0.65, 0.6, and 0.45. The 0.65 albedo value resulted in an average air temperature decrease of 0.6°C in Montreal and 1°C in Toronto. Another study on Montreal found an average air temperature decrease of 0.25°C due to a similar 0.45 increase in albedo from the baseline of 0.2[7].

Improvement to these models can be realized through more detailed estimations of the albedo in the urban areas of interest. Bretz, Akbari, and Rosenfeld[13] quantified the albedo increase potential of Sacramento through investigating the composition of the city and detailing albedo improvements for each surface type. They found a potential for an albedo improvement of 0.18 through maximizing the reflectivity of urban surfaces.

Although studies have been performed to manipulate the albedo of the Island of Montreal, there does not yet exist a similarly detailed estimate of its actual urban albedo. Cold cities have been largely ignored in urban albedo research due to the annual and prolonged period of snow cover. However, the average number of snow cover days in Montreal has decreased from 103 to 73 in the period since 1985[14]. Further, during intense summer heatwaves, Montreal experiences severe UHI effect. Over 400 deaths have been attributed to these heatwaves over the past 30 years[15].

Better policy must be formulated in order to combat the climate change driven increase in the aforementioned consequences of the UHI effect. Policies and pilot projects have already been implemented across the globe using the strategy of increased albedo to offset urban heat. Notably, New York City has painted 9.2 million square feet of roofs white since 2009[16], and Los Angeles is painting a portion of its roads white as a pilot mitigation effort[17]. A 2020 Government of Canada report on reducing UHIs strongly recommended policy that incentives green and cool roofs across Canadian cities[18].

The objective of this project is twofold. First, this research quantifies the current albedo of the Island of Montreal through consideration for the specific proportions of the land use categories of roofs, roads, vegetation, and impervious other. Second, using this baseline current day albedo, this study estimates the albedo increase possible through three roof replacement scenarios. The scenarios are designed to capture the Canadian trends in policy for UHI effect mitigation.

Methodology

Land Use Modification

Numerous data sources were combined to construct a land use map of Montreal*. Land use is divided into four types: [1] roofs; [2] roadways, not including roadside parking or shoulders; [3] vegetation; and [4] impervious other. Impervious other is defined as all land uses which do not fall into the first three categories. Although some porous surfaces, such as bare soil at construction areas, fall into this category, the majority of these miscellaneous surfaces are man-made features such as sidewalks and surface level parking lots, hence the characterization as impervious[13]. Since some urban areas of Montreal fall outside of the city's legal limits, the terrestrial limits of Montreal were used to delineate the city boundaries. All data sources were masked using a terrestrial limit vector file provided by the City of Montreal.

Roofs

We assumed that using building footprints was an adequate analogue for building roofs since walls are generally constructed at a 90% angle from the ground. Roofs were therefore classified using Microsoft Building Footprints, a dataset of computer-generated polygons. The polygon generator is highly accurate but has a recall of only 72.3%, indicating that not all building footprints are present. To address this systemic under-sampling, we assume that the area classified as rooftops by this dataset represents 72.3% of the total surface area of rooftops in Montreal. In order to assess the uncertainty associated with this estimate of roof area, we compared it with building data from OpenStreetMap (OSM), a large Volunteered Geographic Information dataset[19,20].

Roadways

Vector files from the Government of Canada's CanVec series were used for the identification of roadways. While the quality of these official data were uniform across the area of Montreal, certain features such as alleyways were noticeably absent. Therefore, the roadway data were supplemented with vector files from OpenStreetMap (OSM). In both the CanVec and OSM datasets, roadways are primarily represented in one dimension, whereas two dimensional representations are necessary to calculate surface area. Roadway vectors were converted to two dimensions using the attribute representing their number of lanes. 98.11% of roadways had a lane number associated with them.

In the case when the number of lanes was not specified, a value of one was assigned. Next, a standard lane width was derived by measuring 35 roadways across Montreal using the Google Maps "Measure Distance" tool. The average of these lane widths was applied as a buffer to the roadway lines, forming a collection of polygons. Next, the intersections between polygons were dissolved. The surface area of the resulting roadway shapes was then calculated.

Vegetation

Vector files from the Government of Canada's CanVec series were used in conjunction with vector files from OpenStreetMap (OSM). To account for roadways within parks, any vegetation area which intersected with roadways was removed from our vegetation dataset. In order to validate the categorization of vegetation, the vegetation polygons were overlaid on basemaps of Montreal. Three Planet basemaps, which are derived from satellite imagery, were used. To choose three dates for examination, historical climate data from Environment and Climate Change Canada[21] was used to find the maximum daily temperature from 2011-2021. Next, filters were imposed to ensure satellite images with less than 1% cloud cover and more than 98% of the Montreal Island was captured.

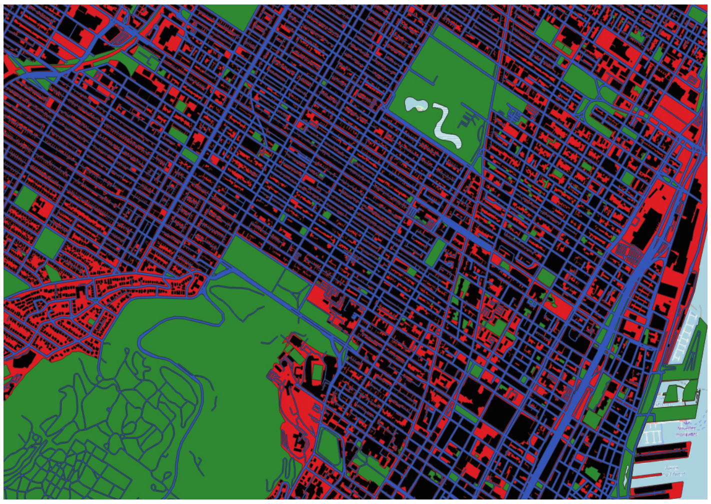


Figure 1. Land use map of Montreal's Plateau neighbourhood. The island area is shown in red. Roadways are overlaid in blue, roofs in black, and vegetation in green.

Upon visual inspection, it was clear that a significant number of vegetated areas shown on the basemaps were not present in the vegetation vector files. Automatic classification was used to address this gap. First, all pixels on the Montreal basemaps that were classified as roof, roadway, or vegetation were discarded. Within the remaining area, values were sampled from points in various vegetated areas across all three basemaps.

Image filters to detect vegetation were constructed using three different pixel value range sizes. The ranges of 50, 60, and 70 pixels were determined so that they captured a visual overestimation or underestimation of vegetation. This resulted in three filters: [1] band 1 values between 30-80, band 2 values between 40-90, and band 3 values between 40-90, [2] band 1 30-90, band 2 40-100, and band 3 40-100, and [3] band 1 30-100, band 2 40-110, and band 3 40-110. These filters were applied to each of the three selected basemaps in order to detect vegetation. The surface area of detected vegetation was calculated by averaging the surface area of the nine outputs. Next, the surface area of detected vegetation was added to the official vegetation surface area obtained from the CanVec and OSM datasets.

Shade Classification

Although roadways in Montreal tend to have dark surfaces, the albedo of roofing materials and other man-made surfaces can vary significantly[22]. To account for this diversity, we further divide the roof and impervious other land uses into white and dark categories.

As with our classification of vegetation, we used a simple threshold-based automatic classification method. The basemaps were first transformed into one-band satellite images of roofs. Then, a 3 by 3 numbered grid was overlaid on the basemaps and the brightest white roof from each grid cell was

---

*The code used to construct this map, as well as further technical details, is made available at https://github.com/sasgilmer/LandUseMap selected. This random sampling technique ensured that all portions of the study area are represented to account for local variances. The darkest value among the samples was used as a threshold value. All pixels whiter than this threshold were categorized as white roofs, and the remaining roofs were categorized as dark. The surface area of the two categories was then calculated and corrected for under-sampling in the same manner as the Building Footprints dataset. The same threshold value was used to divide the impervious other area into white and dark categories.

Table 1: Literature derived surface albedo values for each land use category.

		Albedo Value
Roofs	White	0.5700 ± 0.1075
	Dark	0.1200 ± 0.0436
	White Paint	0.7100 ± 0.0584
	Toughkote	0.8500
	Simple Grass Green Roof	0.2560 ± 0.0247
Roads		0.1400 ± 0.0300
Vegetation		0.2600 ± 0.0133
Impervious Other	White	0.3375 ± 0.0520
	Dark	0.1680 ± 0.0585

Albedo Value Calculation

Standard measurements of the albedo of each category were sourced from numerous materialbased studies[4,5,8,13,22-34]. An average albedo value across studies was used to account for differences in methodology and date of data collection. Despite the range of land uses included in impervious other, average albedo values represent this category with relatively high certainty, as evidenced by Table 1.

The albedo of the island is calculated as follows:

$$A_{\mathrm{MTL}} = \frac{1}{SA_{\mathrm{MTL}}} \sum_{x \in \mathrm{Land\ Uses}} A_x * SA_x \tag{1}$$

A	Albedo
SA	Surface Area
MTL	Montreal (terrestrial limits)
Land Uses	The set of all land uses within Montreal

The current albedo of Montreal was then calculated by setting Land Uses= {Vegetation, Roadways, Dark Roofs, White Roofs, Dark Impervious Other, White Impervious Other}.

We also calculated albedo values for Montreal under different scenarios for land use change. First, we considered whitening roofs using an average white paint. This is reflected in Equation 1 by adding White Paint to the set of land uses while decreasing $SA_{\text{White Roof}}$ and/or $SA_{\text{Dark Roof}}$. However, the albedo of white paint can vary considerably depending on its material composition and the thickness of application[22]. Therefore, we wished to consider a scenario which represents only the highest albedo white paints on the market. To this end, we selected ToughKote as a representative of high albedo white paints since its albedo has been reported as some of the highest in two albedo datasets[22,29]. Similarly, we calculated the albedo of Montreal with increased green roofs.

Results

Current Day Albedo Value

Our analysis of surface albedo numbers in the literature revealed that the attribution of a standard albedo value for urban materials represents a significant source of uncertainty. Table 1 shows the literature derived albedo value for each of our land use categories, as well as the ToughKote and Simple Green Grass Roof albedos used for mitigation scenarios. Vegetation and White Roof values were relatively consistent across publications, while the values for White Paint and White Impervious Other vary. The land use categories, before correction for under-sampling, can be seen in Fig. 1. Note the under representation of roofs and vegetation.

Correcting for under-sampling and augmenting our data using threshold-based classification resulted in a more even breakdown (See Fig. 2). Vegetation occupies the largest proportion of the island with an area of 174.27 ± 34.16 km^2, impervious other has an area of 122.06 ± 52.91 km^2, roofs have an area of 105.88 ± 11.05 km^2. Of the four primary land use types, roadways represent the smallest category with a surface area of 96.43 ± 13.73 km^2.

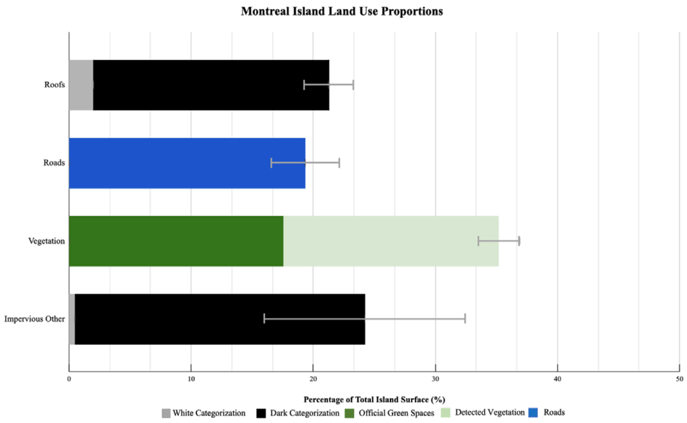


Figure 2. Montreal island land use proportions are reported as percentages of the total Montreal island area.

The roofs category includes 9.92 ± 1.04 km^2 of white roofs and 95.96 ± 10.02 km^2 of dark roofs. The impervious other category has a much smaller proportion of currently white surfaces, occupying 2.31 ± 1.00 km^2 of the island as compared to the 119.75 ± 51.92 km^2 of dark surfaces. The vegetation category is approximately equally made up of official green spaces (87.43 ± 17.08 km^2) and detected vegetation from satellite imagery (88.10 ± 17.08 km^2). The final current day albedo calculated is 0.19 ± 0.057, as indicated by Fig. 3.

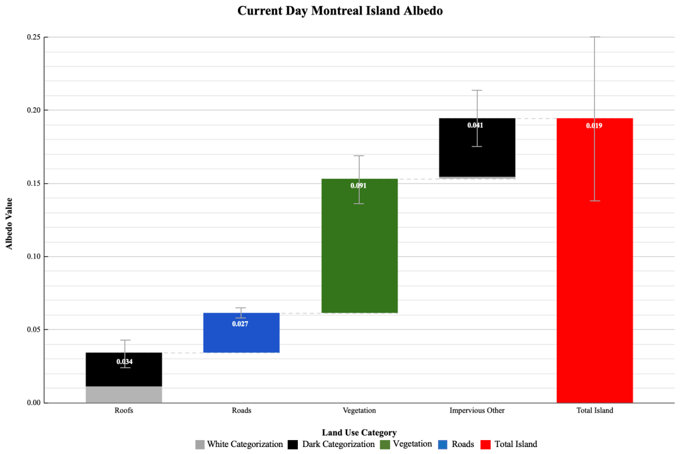


Figure 3. Breakdown of the final current day albedo of the Montreal island. Each land use category's albedo attribution is indicated by the white text.

As demonstrated by Table 2, the land use breakdown of Montreal includes nearly all other cited studies within its uncertainty bounds. However, vegetation occupies a greater proportion of the total area than other studies. This is likely due to the extent of the city incorporated in studies. The terrestrial limits of Montreal that we used as a mask include rural areas which fall outside of official city limits and are primarily agricultural land.

Table 2. Land use proportions of total city area reported across different cities [13, 35-39]

Study	City	Roofs		Pavement			Vegetation	
		White	Dark	Roads	Impervious Other		Official Green Spaces	Detected Vegetation
					White	Dark		
This Paper	Montreal	1.99 ± 0.21	19.24 ± 2.00	19.34 ± 2.75	0.46 ± 0.20	24.02 ± 10.41	17.53 ± 3.42	17.67 ± 3.42
Bretz et al. (1998)	Sacramento	8	20	16	4	10	42	
Akbari et al. (2008)	Global Average	20-25		30-35			40	
Rose et al. (1999)	Sacramento	19.7		44.5	15.4		20.3	
Rose et al. (2003)	Houston	21.3		29.2	12.4		37.1	
Rose et al. (2001b)	Chicago	24.8		37.1	11.4		26.7	
Rose et al. (2001a)	Salt Lake City	21		36.4	8.5		33.3	

Albedo Increase Potential

The potential effects on albedo of various UHI mitigation strategies are outlined in Figure 4. This figure illustrates scenarios where 100% of the current roofs are replaced by one of either average white paint, Toughkote, or simple grass. As the roof land use category has a white and dark component, consideration was later made to replace the roof category both proportionally and in parts.

As seen in Fig. 4, it may be more efficient to isolate certain shades of roofs in policy decisions. Due to the fact that white roofs are only 9.37% of the roof category (and 1.99% of the total island), painting only this proportion does not significantly change the current day albedo in any of the scenarios. In fact, it reduces the albedo in scenario 3 because grass has a lower albedo than the white surfaces. Painting only the dark roofs however nearly achieves the same final albedo increase as 100% of roofs in each scenario. Again, in scenario 3, because of the lower albedo of grass, replacing only dark roofs achieves a higher final albedo than replacing 100% of roofs.

Discussion

By confirming a value for the urban albedo of the Montreal island, we aimed to address the larger issue of the general use of 0.2 albedo for all urban surfaces within the literature. Here, we find that the albedo of the Montreal island is 0.19 +/- 0.057. This albedo value includes the 0.2 urban estimate within its range of uncertainty. Thus we satisfy our first hypothesis. This result implies that the urban albedo of 0.2 does generalize to Montreal. Through examining three roof replacement possibilities, we found a potential albedo increase that surpassed the desired 0.1 in 2 of the 3 scenarios. These increases fell short of the 0.18 potential increase observed by Bretz, Akbari, and Rosenfeld[13] in Sacramento, and thus satisfied our second hypothesis. Further, these results align with the 0.1 albedo increase frequently used in UHI effect literature. Therefore, we are able to assess the significance of the albedo increase scenarios by a comparison with previous studies, even within the Montreal context[35].

Our Montreal island urban fabric breakdown closely aligns to other studies. All previous studies captured in Table 2 demonstrate land use breakdowns near to or within our uncertainties, where differences are attributable to city-specific histories. This result suggests that the use of general urban land use breakdowns is justified in the case of Montreal. Therefore, the majority of the uncertainty in the current day albedo value derives from uncertainty in the albedo values applied to each land use category. Land surface albedo is reported in two categories: intrinsic values, specific to a material, and apparent values, observed to temporally change with solar radiation angle[40]. The smallest apparent albedo occurs at noon, and since intrinsic albedo is a measurement often taken in peak sunlight, there exists systematic misrepresentation in the literature[40]. Further, here we apply our intrinsic albedo values to surfaces based on colour, but surface roughness should also play a role.

Still, the potential for improvement of the Montreal albedo value resulted in values that align with other research[7,41] and allow for comparison between different replacement scenarios. The simple grass green roof scenario resulted in an extremely small albedo change of 0.02, while the two whitening scenarios resulted in values both causing an ultimate increase of albedo to 0.3+. To assess whether these increases are significant to the UHI effect, they must be related to temperature. The relationship between albedo, evapotranspiration and meteorological dynamics in urban spaces is still not well confined, but the abundance of literature allows for general conclusions[6].

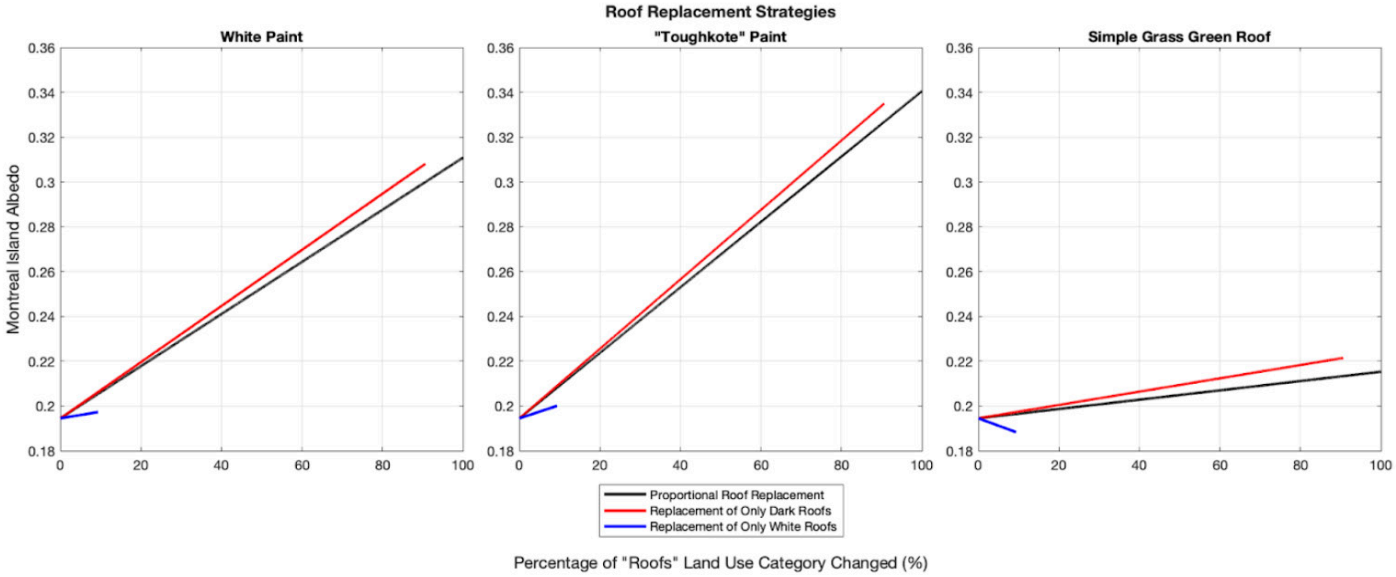


Figure 4. Resulting albedo increase due to each roof replacement scenario. Further, varied strategies for scenario implementation are displayed. Blue indicates replacement of only the white subcategory of the roofs currently present in Montreal. Red indicates replacement of only the dark subcategory of roofs currently present in Montreal. Black indicates 100% roof replacement.

First, we address the negligible effect to albedo caused by the simple grass roof implementation. Green roofs have gained popularity in recent years and covering roofs in short grasses has been widely shown to have a cooling effect[42]. However, these surfaces do not primarily mitigate the UHI effect through albedo. Instead, they primarily mitigate temperature by way of evapotranspiration. A study in Japan found that despite evapotranspiration, white roofs were cooler than simple grass green roofs[43]. This was corroborated by Mackey, Lee, and Smith[44], who found that grass roofs had a negligible effect on surface temperature because vegetation must be dense and varied to cause cooling.

The whitening scenarios may be related to the UHI effect by way of surface temperature, air temperature, and energy offsets. Mackey, Lee, and Smith[44] found a linear relationship between albedo increase and surface temperature change in Chicago by sampling satellite images over a 15-year span. Using their empirical linear relationship, our 0.1 increase to albedo would result in a 1°C reduction in daily average surface temperature. Yang, Wang, and Kaloush[6] find numerous studies that relate surface temperature to albedo changes. This can be a difficult relationship to compare across differing urban land use proportions. For example, a study on Sacramento by Taha[45] found an increase of roof albedo by far less than 0.1 to cause as much as a 10°C reduction in average monthly summer surface temperatures.

Using a Montreal specific model, Touchaei[7] found that an increase of 0.2 in albedo (within 2 standard deviations of both whitening scenarios) resulted in a 2.4°C decrease in surface temperature, and a 0.4°C decrease in air temperature (a function of the surface temperature and sensible heat flux). Relating air temperature to surface temperature is difficult due to turbulent mixing in the lower atmosphere causing an inconsistent relationship[6]. However, many studies still do so, such as Sailor, Kalkstein, and Wong[46] who found a 0.1 albedo increase in Philadelphia to produce a 0.3-0.5°C decrease in day time air temperatures.

Finally, global studies also argue for the significant impact that a 0.1 albedo increase may have. Use of the University of Victoria Earth System Climate Model[47] by Touchaei[7] finds a 0.01-0.07°C decrease through a global increase of 0.1 to only urban areas. Another global model found that a 0.1 albedo increase upon all latitudes within 45 degrees reduces long term global temperature by 2°C[41]. They produced a mathematical relationship where for each 0.01 increase of albedo to 1 m^2 of surface area, there results a long term global temperature decrease of approximately 3×10^{-5} K. A global study by Akbari, Menon, and Rosenfeld[35] related this albedo increase to the offset of 44 Gt of CO_2 by way of reduction of global cooling energy use.

Thus the literature shows that there is a linear relationship between temperature (both surface and air) and change in albedo, within which 0.1 albedo increase is significant both locally, within Montreal, and globally[41,44]. Our study shows that there are multiple routes to achieve this baseline increase of 0.1, by way of implementation of average white surfaces proportionally to all roofs, or by use of the maximal product Toughkote on only the roofs that are not yet light in colour.

However, there are limitations to this work. Although our work aimed at finding a well resolved albedo value for the island of Montreal, this value is not heterogeneous because it does not consider the geographical angles and solar interactions of the Montreal island. The scope of our work is limited to the summer months when the UHI effect is prevalent. Further work is needed to characterize the year-round albedo of Montreal, which involves considering the role of snow cover. As well, characterization of land use categories into vegetation and white/dark were completed through use of visible wavelengths and not the entire radiative spectrum. Improved practices seen in recent work by Mackey, Lee, and Smith[44] use remote sensing techniques to study interaction of all wavelengths with the Earth's surface, and this method would certainly improve the accuracy of the values found in this study. Large uncertainties on all results take these limitations into consideration.

Conclusion

The albedo value calculated for the Island of Montreal in this research substantiates the WRF-UCM model's default albedo value of 0.2. Therefore, general conclusions made in previous literature on the relationship between urban albedo and the UHI effect should continue to be understood as accurately representative. However, as proven in our study, this value will marginally vary between cities depending on their land use category proportions. Further, the uncertainty on the albedo value found here can be better confined through more accurate methods for identifying land use proportions. Thus, for cities considering rooftop changes to mitigate the effect of UHIs, calculating an initial albedo value will produce a more accurate understanding of the potential temperature decrease. Our research can provide a framework for this calculation for cities across the globe.

Secondly, the rooftop whitening scenarios by way of both average white paint and Toughkote paint align with the widely used 0.1 albedo increase present in urban albedo literature. While the simple grass green roof scenario showed a negligible albedo increase, there is opportunity for further Montreal specific research on the potential for green roofs to cool internal building temperatures as an additional UHI mitigation strategy. This research should focus instead on evapotranspiration potential of vegetation presence as opposed to albedo. Furthermore, using our data on the albedo increase potential of white roofs, future study should identify the associated surface and air temperature decrease. Focus should be placed both on Montreal regional temperatures and internal building temperatures as both are relevant to mitigating the UHI effect. These changes translate to improved energy use practices through summer months that present a number of other climate change related benefits. Finally, using a similar model to Jandaghian and Akbari[11], empirical linkages can also be made between changes in albedo and heat related mortalities in Montreal. Such research, made possible by building upon our study, will allow policymakers to weigh the economic, social, and environmental benefits of white roofs as an UHI mitigation strategy for the island of Montreal.

References

1. United Nations. 68% of the World Population Projected to Live in Urban Areas by 2050, Says UN May 2018. https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanizationprospects.html.

2. Kacyira, A. K. Addressing the Sustainable Urbanization Challenge. United Nations. https://www.un.org/en/chronicle/article/addressing-sustainable-urbanization-challenge (2011).

3. Health Canada. Climate Change and Health 2009. https://www.canada.ca/en/health-canada/services/environmental- workplace- health/reports-publications/climate- change- health/climate- changehealth- adaptation-bulletin- number- 1- november- 2009- revised- december- 2010- health-canada2009.html.

4. Wang, Y. & Akbari, H. Analysis of urban heat island phenomenon and mitigation solutions evaluation for Montreal. Sustain. Cities Soc. 26, 438–446 (2016).

5. Prado, R. T. A. & Ferreira, F. L. Measurement of albedo and analysis of its influence the surface temperature of building roof materials. Energy Build. 37, 295–300 (2005).

6. Yang, J., Wang, Z.-H. & Kaloush, K. E. Environmental impacts of reflective materials: Is high albedo a "silver bullet" for mitigating urban heat island? Renew. Sustain. Energy Rev. 47, 830–843 (2015).

7. Touchaei, A. Characterizing the Effect of Increasing Albedo on Urban Meteorology and Air Quality in Cold Climates, a Case Study for Montreal PhD thesis (2015).

8. Berkeley Lab Heat Island Group. California Rooftop Albedo 2014. https://albedomap.lbl.gov/.

9. Tewari, M., Chen, F., Kusaka, H. & Miao, S. Coupled WRF/Unified

Noah/Urban-Canopy Modeling System https://ral.ucar.edu/sites/default/files/public/product-tool/WRF-LSM-Urban.pdf (2007).

10. Vahmani, P. & Ban-Weiss, G. A. Impact of remotely sensed albedo and vegetation fraction on simulation of urban climate in WRF-urban canopy model: A case study of the urban heat island in Los Angeles. J. Geophys. Res. Atmos. 121, 1511–1531 (2016).

11. Jandaghian, Z. & Akbari, H. The Effects of Increasing Surface Reflectivity on Heat-Related Mortality in Greater Montreal Area, Canada. Urban Clim. 25, 135–151 (2018).

12. Jandaghian, Z. & Berardi, U. Effects of increasing urban albedo in the Greater Toronto Area. IOP Conf. Ser. Mater. Sci. Eng. 609, 072002 (2019).

13. Bretz, S., Akbari, H. & Rosenfeld, A. Practical issues for using solar-reflective materials to mitigate urban heat islands. Atmos. Environ. 32. Conference on the Benefits of the Urban Forest, 95–101 (1998).

14. Communauté Métropolitaine de Montréal. Climate Change Adaption Plan for the Montreal Urban Agglomeration 2015-2020 (2017).

15. Suh, C. The Urban Heat Island Effect in Montréal Effective Policy through an Interdisciplinary Perspective https://www.socialconnectedness.org/wp-content/uploads/2019/10/The-Urban-Heat-Island-Effectin-Montre%5C%CC%5C%81al.pdf (2019).

16. Kotecki, P. New York City has painted over 9.2 million square feet of rooftops white — and it could be a brilliant heat-fighting plan Aug. 2018. https://www.businessinsider.com/new-york-city-painted-6-million-square-feet-of-rooftop-white-2018-8.

17. McPhate, M. California Today: A Plan to Cool Down L.A. The New York Times. https://www.nytimes.com/2017/07/07/us/california-today-cool-pavements-la.html (2017).

18. Service Canada. Reducing urban heat islands to protect health in Canada May 2020. https://www.canada.ca/en/services/health/publications/healthy- living /reducing- urban- heat- islands- protect- healthcanada.html.

19. OpenStreetMap contributors.Planet dump retrieved from https://planet.osm.org https://www.openstreetmap.org. 2017.
20. Mocnik, F.-B., Mobasheri, A. & Zipf, A. Open source data mining infrastructure for exploring and analysing OpenStreetMap. Open Geospatial Data, Software and Standards 3, 1–15 (2018).

21. Environment and Climate Change Canada. Historical Climate Data Nov. 2021. https://climate.weather.gc.ca.

22. Berdahl, P. Cool roofing materials database 1998. https://heatisland.lbl.gov/resources/cool-roofingmaterials-database.

23. Gaffin, S. et al. Energy Balance Modeling Applied to a Comparison of White and Green Roof Cooling Efficiency. Center for Climate Systems Research, Columbia University (2005).

24. Page, J. The role of solar radiation climatology in the design of Photovoltaic Systems. Practical Handbook of Photovoltaics, 5–66 (2003).

25. Taha, H., Sailor, D. J. & Akbari, H. High-Albedo Materials for Reducing Building Cooling Energy Use (1992).

26. Goodman, D. S. J. NASA/GHCC Project Atlanta Aug. 1999. https://weather.msfc.nasa.gov/urban/urban_heat_island.html.

27. Kotak, Y., Gul, M., Muneer, T. & Ivanova, S. Investigating the impact of ground albedo on the performance of PV systems Apr. 2015.

28. Campra, P., Garcia, M., Canton, Y. & Palacios-Orueta, A. Surface temperature cooling trends and negative radiative forcing due to land use change toward greenhouse farming in southeastern Spain. J. Geophys. Res. Atmos. 113 (2008).

29. Santamouris, M. Energy and Climate in the Urban Built Environment (Routledge, June 2013).

30. Andrews, R. W. & Pearce, J. M. The effect of spectral albedo on amorphous silicon and crystalline silicon solar photovoltaic device performance. Sol. Energy 91, 233–241 (2013).

31. Gul, M., Kotak, Y., Muneer, T. & Ivanova, S. Enhancement of Albedo for Solar Energy Gain with Particular Emphasis on Overcast Skies. Energies 11, 2881 (2018).

32. Li, D., Bou-Zeid, E. & Oppenheimer, M. The effectiveness of cool and green roofs as urban heat island mitigation strategies. Environmental Research Letters 9, 055002 (2014).

33. Qin, Y., Liang, J., Luo, Z., Tan, K. & Zhu, Z. Increasing the southern side-slope albedo remedies thermal asymmetry of cold-region roadway embankments. Cold Reg. Sci. Technol. 123, 115–120 (2016).

34. Roesch, A., Wild, M., Ohmura, A. & Gilgen, H. Assessment of GCM simulated snow albedo using direct observations. Clim. Dyn. 15, 405–418 (1999).

35. Akbari, H., Menon, S. & Rosenfeld, A. Global cooling: increasing world-wide urban albedos to offset CO2. Climatic Change 94, 275–286. https://www.energy.ca.gov/2008publications/CEC-999-2008-020/CEC-999-2008-020.PDF (2008).

36. Rose, L. S., Akbari, H. & Taha, H. Characterizing the Fabric of the Urban Environment: A Case Study of Sacramento, California (1999).

37. Rose, L. S., Akbari, H. & Taha, H. Characterizing the Fabric of the Urban Environment: A Case Study of Greater Houston, Texas (2003).

38. Rose, L. S. & Akbari, H. Characterizing the Fabric of the Urban Environment: A Case Study of Chicago, Illinois (2001).

39. Rose, L. S. & Akbari, H. Characterizing the Fabric of the Urban Environment: A Case Study of Salt Lake City, Utah (2001).

40. Wang, D. et al. Estimating daily mean land surface albedo from MODIS data. J. Geophys. Res. Atmos. 120, 4825–4841 (2015).

41. Akbari, H., Damon Matthews, H. & Seto, D. The long-term effect of increasing the albedo of urban areas. Environ. Res. Lett. 7, 024004 (2012).

42. Jamei, E., Chau, H. W., Seyedmahmoudian, M. & Stojcevski, A. Review on the cooling potential of green roofs in different climates. Sci. Total Environ. 791, 148407 (2021).

43. Takebayashi, H. & Moriyama, M. Surface heat budget on green roof and high reflection roof for mitigation of urban heat island. Build. Environ. 42, 2971–2979 (2007).

44. Mackey, C. W., Lee, X. & Smith, R. B. Remotely sensing the cooling effects of city scale efforts to reduce urban heat island. Build. Environ. 49, 348–358 (2012).

45. Taha, H. Meso-urban meteorological and photochemical modeling of heat island mitigation. Atmos.
Environ. 42, 8795–8809 (2008).

46. Sailor, D. J., Kalkstein, L. S. & Wong, E. The Potential of Urban Heat Island Mitigation to Alleviate Heat-
Related Mortality: Methodological Overview and Preliminary Modeling Results for Philadelphia (2002).

47. Weaver, A. J. et al. The UVic earth system climate model: Model description, climatology, and applications
to past, present and future climates. Atmos. Ocean 39, 361–428 (2001).

Submitted: 01/24/2022
Accepted: 03/19/2022

Research Article

Maxine Forder[1], Maureen McKeague[1,2]

Tuning Aptamer-Switching for Biosensing Malarial Proteins

[1]Department of Chemistry, McGill University, Montreal, QC, Canada

[2]Department of Pharmacology, McGill University, Montreal, QC, Canada

Keywords

Aptasensor, diagnosis, Malaria, fluorescence

Email Correspondence

maxine.forder2@mail.mcgill.ca

Abstract

Malaria is a life-threatening disease caused by a protozoan transmitted through mosquito bites. Early diagnosis is essential to start treatment and prevent further transmission. However, current diagnostic methods are expensive, time consuming, and lack the portability required for efficient testing. Emerging methods that are faster and more portable include specialized biosensors called aptasensors. This study compared a range of complementary probes against malaria-binding aptamers to develop an aptamer-switch based platform. One candidate showed a promising increase in fluorescence upon incubating with a malaria target protein. This work has the potential to be incorporated into an aptasensor for rapid detection of malaria infections.

Introduction

Malaria is a potentially lethal parasitic infection present throughout the world. It is mainly concentrated in Africa but is also present in South-East Asia, the Mediterranean, and the Americas to varying degrees[1]. Despite advances in therapeutics and diagnostic methods[2], over 200 million cases and 400,000 deaths were reported globally in 2019[1]. Malaria is caused by five species of the protozoan *Plasmodium*, including *P. falciparum, P. vivax, P. malariae, P. ovale,* and *P. knowlesi*. The *P. falciparum* species is most prevalent and lethal, while *P. vivax* is the widest spread[3]. These protozoans are transmitted between humans through female *Anopheles* mosquito bites[4]. Early during infection, fever, chills, vomiting, and dizziness occur, which, if left untreated, may progress to severe anemia, acute renal failure, pulmonary edema, and even death[5,6].

1.1 Current diagnostic tools for malaria detection

Early diagnosis of malaria and treatment is essential. The gold standard of diagnosis is conventional microscopy of blood smears, which can identify malaria species, parasite stages, and quantify red blood cell counts[7]. However, this process takes an hour at minimum, requires trained staff, and is expensive[8,9]. Rapid diagnostic tests are a common alternative that rely on an antibody-based detection. Although these tests require less training and are faster, developed tests currently only identify two species and are prone to false negatives[7,8]. Polymerase Chain Reaction (PCR)-based tests form a third major diagnostic test that can be used elsewhere for high sensitivity detection to identify species and monitor responses to treatment[9]. However, PCR-based tests are not used in malaria-endemic regions due to high cost, complexity, and personnel requirements[10]. Evidently, there is a need for faster, cheaper, and more portable diagnostic methods.

Recent research has focused on developing new biosensors. Biosensors are devices that detect chemical or biological substances of interest by coupling a biological recognition element with a transducer that converts any recognition into an electrical signal[11]. Often, antibodies are chosen as the biological recognition element, although alternatives such as enzymes and DNA probes may also be used[11]. Biosensors can test for parasites in blood and are being tested for reliability in both saliva and urine as they are non-invasive[12,13]. Malaria biosensors already exist and have higher sensitivity than conventional methods[9,14,15]. However, low antibody stability and short half-life is problematic in the field, in which the uncontrolled environment may cause antibodies to denature irreversibly[16].

1.2 Aptamer-based biosensors

Aptasensors, biosensors that rely on aptamers instead of antibodies[17], provide a promising alternative to antibody-based biosensors. Aptamers are nucleic acid molecules that bind molecular targets with high selectivity, high affinity, and most importantly, improved stability than antibodies; if denatured, they can simply refold once conditions return to conditions in which they were optimized[18]. This stability makes them better candidates for testing in the field than antibodies[16]. Aptamers can be developed into biosensors by incorporating a variety of transducer systems including gold nanoparticles, fluorophores, electrochemical probes, and more[19,20,21].

One of the many ways that fluorescence aptasensors can indicate the presence of protein is by sending a fluorescence signal to the electronics of a smartphone, for instance[22]. For this type of aptasensor to function, the aptamer must change conformation upon target binding to obtain a measurable change in fluorescence[23]. Typical aptamers often must be re-designed, potentially making use of complementary quencher probes, to achieve a significant conformational “switch” (Figure 1).

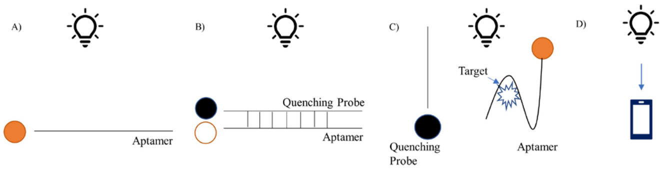


Figure 1. An example of an aptamer-switch system implemented in the development of an aptasensor. In the switch system, the aptamer tagged with a fluorophore emits light (A) and is quenched when a complementary sequence (probe with a fluorescence quencher) is present (B). In the presence of the target protein, the aptamer preferentially binds to the target, releasing the quenching probe. As such, the tagged aptamer emits light again (C). Electronics can convert the fluorescent light signal into a recognizable signal measurement for a smartphone (D).

1.3 Aptamers developed for malaria detection

Several aptamers that have potential for detecting malaria exist or are currently being developed. The majority of aptamers target plasmodium lactate dehydrogenase (PfLDH) because this was the first important diagnostic target identified, although there is now substantial research on glutamate dehydrogenase[24]. Two aptamers towards PfLDH have been developed into aptasensors that can distinguish between *P. falciparum* and *P. vivax*[25,26], providing an excellent tool to identify the plasmodium species for treatment purposes. However, since the PfLDH aptamer cannot detect *P. knowlesi* infection, a strain that requires immediate and aggressive treatment[27], these aptasensors must be used in a multi-panel array with other tests before a negative result can be confirmed[16]. More recently, the glutamate dehydrogenase aptamer has been developed into an aptasensor that detects *P. falciparum*[28]. Although the system may detect *P. vivax*[29], it remains limited to this strain until research in other species is undertaken. Looking to the future, it would be useful to develop a wide array of malaria-detecting aptamers with varying strain specificity that can be made into aptasensors. For example, an aptamer that broadly detects all malaria strains would enable a rapid preliminary diagnostic method. However, there are no currently reported aptamers with this capability[16,24,30]. Furthermore, additional aptamers specific to each species are needed following an initial diagnosis to detect each type of malaria strain so that doctors can adjust treatment accordingly to each specific strain. Substantial aptamer discovery work has identified additional aptamers that can bind and interact with malaria, but require further development into aptasensors[30].

One potential target for aptamer development is Heat shock protein 90 (Hsp90), an ATP-dependent molecular chaperone that is located on the cell surface and used by *P. falciparum* to regulate its development[31,32]. Although Hsp90 is conserved across many species, PfHsp90 is proposed to be distinct enough from human Hsp90 to allow for selective targeting[31]. Thus, this paper tests PfHsp90 as the target for the aptamer-switch system. Researchers can develop potential aptamers to this target using cell-SELEX (Systematic evolution of ligands by exponential enrichment) or protein-SELEX, an in vitro evolution method[33]. Developing an aptamer-switch system using Hsp90 is our first step towards developing an aptasensor for malaria detection. This paper will explore quenching probes to achieve "aptamer switching" which can be then incorporated into an aptasensor connected to a smartphone. If successful, these aptasensors will provide a cheap, portable method of malaria detection.

Methods

2.1 Materials

Buffers were prepared starting with 1X Phosphate Buffered Saline (PBS) at pH 7.4 from Gibco (Amarillo, TX). The Triton-X100/RPMI Buffer used Triton-X100 from BioBasic (Markham, ON). The Recombinant Hsp90 (*P. falciparum*) partial Protein SPR-122A was ordered from StressMarq Biosciences (Victoria, BC). Two aptamer sequences, Aptamer.1 and Aptamer.2, were selected by collaborators by employing the cell-SELEX method with malaria-infected cells as previously described[30]. All oligonucleotides, including the aptamers, were purchased from Integrated DNA Technologies (Coralville, IA) and are available from the McKeague lab upon request. Sequences are in Table 1. 'FAM' is the fluorophore and 'IABkFQ' is the Iowa Black® fluorescence quencher.

Table 1. Oligonucleotide Names and Sequences

Oligos	Sequence
Quencher.1C8	TTCTCCACCC- 3'Q
Quencher.1C10	CCCTCCACCC- 3'Q
Quencher.1C12	ACCCCTCCACCC- 3'Q
Quencher.2C10	CGTAATAGAC- 3'Q
Quencher.2C12	CCCGTAATAGAC- 3'Q
Quencher.2C14	CCCCCGTAATAGAC- 3'Q
Quencher.1C25mis	AGGCGATTTTTTTTTTTTTCCACCC - 3'Q
5'Q Quencher.1C8	5'Q - CAAACTTCTA
5'Q Quencher.1C10	5'Q - CAAACTTCCT
5'Q Quencher.1C12	5'Q - CAAACTTCCTAA
5'Q Quencher.2C10	5'Q - CAGATAATGC
5'Q Quencher.2C12	5'Q - CAGATAATGCCC
5'Q Quencher.2C14	5'Q - CAGATAATGCCCCC

[1] The 3' Quencher (3'Q) is 3IABkFQ; The 5'Quencher (5'Q) is 5IABkFQ

2.2 Preparing Regeants

Binding Buffer was prepared by adding 0.5 mM $CaCl_2$, 0.5 mM $MgCl_2$, and 1 mM KCl to PBS pH 7.4. The Binding Buffer for Fluorescence Anisotropy experiments included 0.05% Triton-X100.

The oligonucleotides were diluted with MilliQ water to form 100 µmol/L stock solutions. For the fluorescence-quenching switching experiments, aptamers were made to 600 nM or 1200 nM using Binding Buffer. Quencher probes were made to 1200 nM using Binding Buffer. For Fluorescence Anisotropy experiments, aptamers were made to 10 nM using Binding Buffer with Triton-X.

The target, Heat shock protein 90 (Hsp90), was diluted using the Binding Buffer to 2000 nM or 5000 nM as needed. Hsp90 concentrations were prepared using two-fold serial dilutions and ranged from 5000 nM to 1.56 nM when starting with 5000nM stock solution, or from 2000 nM to 0.1 nM, when starting with 2000 nM stock solution.

2.3 Testing aptamer switching with fluorescence-quenching measurements

Aptamers, buffer, and target Hsp90 protein were added to 96-well plates and incubated for 30 minutes at room temperature to allow Hsp90 binding. Controls included removing one of each of these components and replacing with buffer. After 5 minutes of incubation at room temperature, the fluorescence intensity was measured with the BioTek Cytation 5 Cell Imaging Multi-Mode Reader at excitation and emission wavelengths of 495 nm and 520 nm +/-10 nm, respectively, after 10 seconds of plate shaking. Conditions were prepared in triplicate whenever possible, as indicated. In some cases, duplicates were necessary to reduce material costs and to meet space constraints of a 96 well plate.

2.4 Fluorescence Anisotropy

Binding of the aptamer to the heat shock protein was tested using fluorescence anisotropy. Each aptamer was labelled with a 5'-FAM. Triton-X100 detergent was added to the binding buffer to 0.05% as previously described[34]. Aptamers had a constant concentration of 10 nM. Hsp90 concentrations were prepared using two-fold serial dilutions and varied from 5000 nM to 1.56 nM. The negative control was a 10% glycerol solution in binding buffer that was prepared similarly to Hsp90, using two-fold dilutions to generate solutions ranging from 5000 nM to 1.56 nM. After shaking the plate for 15 minutes at room temperature, the BioTek Cytation 5 Cell Imaging Multi-Mode Reader was used to take polarization measurements using the same procedure as in fluorescence-quenching measurements, but with an extra step to shake the plate for one minute once inside the plate reader. GraphPad Prism 6 was used to plot the difference between polarization values in the presence and absence of Hsp90 against the total concentration of Hsp90 and to generate a binding isotherm using a one-site binding (hyperbola) fit. The dissociation constant, K_D, was reported as the mean across duplicates. This value indicates the strength of interaction between two molecules, with a low K_D indicating strong affinity between target molecules.

Results

3.1 Confirming binding between the aptamers and Hsp90

We first wanted to confirm that Hsp90 bound sufficiently to the aptamers by measuring the dissociation constant, K_D. Several methods are possible to measure interactions between aptamers and their targets[35]; however, since we are interested in developing fluorescence-based aptasensors, measuring aptamer binding with the fluorophore tag is useful to ensure it does not interfere with target binding. As such, we made use of a fluorescence anisotropy assay to measure the aptamer affinity in solution.

The K_D values between Aptamer.1 and Hsp90 and between Aptamer.2 and Hsp90 were 154 nM and 116 nM, respectively (Figure 2). These values represent a reasonably good affinity between Aptamer/Hsp90 binding. Notably, the 10% glycerol control showed no change in polarization. This indicates that binding was between the aptamer and Hsp90 instead of the aptamer and glycerol it came stored in.

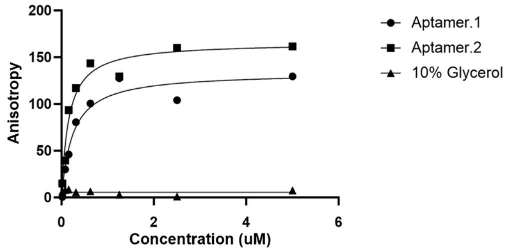


Figure 2. K_D Results of HSP90 Binding to Aptamer.1 or Aptamer.2. The K_D of Aptamer.1 and HSP90 is 116 nM and the K_D of Aptamer.2 and HSP90 is 154 nM. The glycerol control condition showed no change in polarization, suggesting that binding is indeed due to the aptamer/HSP90 interaction.

3.2 Screening for aptamer probe displacement as an "aptamer switch"

We first tested both aptamers with a panel of complementary quencher probes ranging in sizes from 10 to 14 nucleotides (Figure 3A). These quencher probes contained a quencher and were complementary to the 5' end of each aptamer such that the fluorescein present on the 5' end would be quenched. If an aptamer "switch" was to occur, the quencher probe would be released from the aptamer once the target was present and cause an increase in fluorescence signal, as shown in Figure 1. Aptamers (600 nM), quencher probe (1200 nM), and Hsp90 (2000 nM) were all added at the same time, incubated for 20 minutes at room temperature, and fluorescence was measured.

Results yielded excellent quenching but little switching upon addition of Hsp90 (Figure 3B and 3C). Both aptamers showed significant quenching upon addition of the probes. However, there were no significant changes in fluorescence between the Aptamer/Quencher Probe conditions and the Aptamer/Quencher Probe/Hsp90 conditions. This lack of switching means that Hsp90 was not detected.

3.2 Testing quencher probes that bind to the 3' end of the aptamer

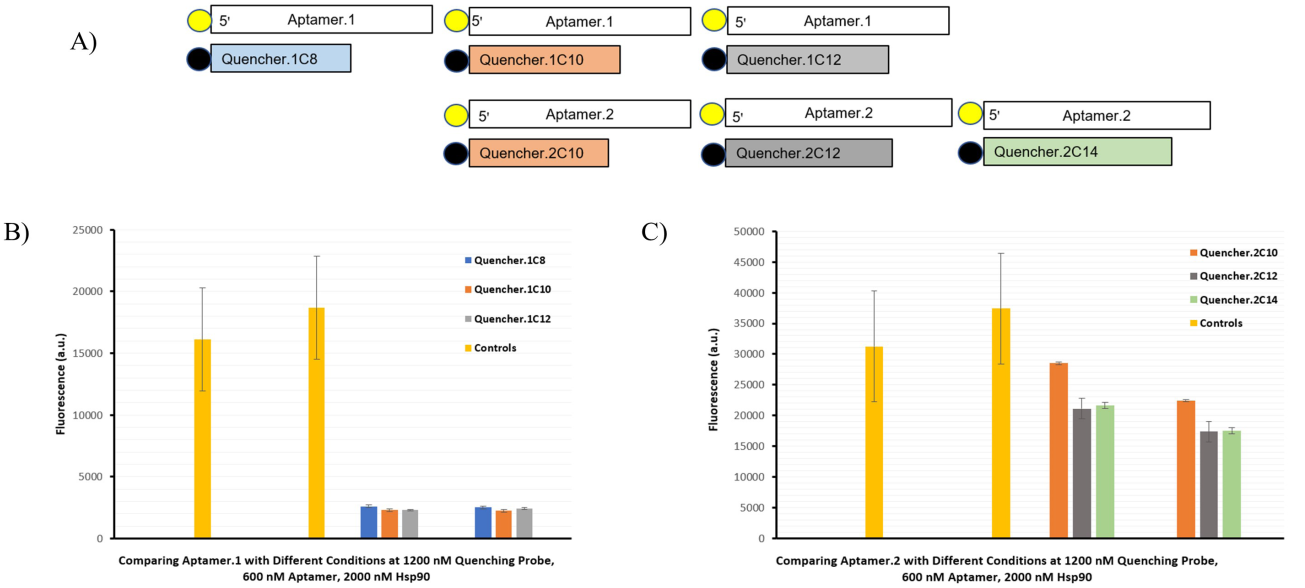
Given that there is no information about the structure of the aptamers and the positions where binding is expected, we hypothesized that the binding site was far away from the 5' end of the aptamer. As such, we designed new quencher probes that interact with the 3' end of the aptamer. Accordingly, we flipped the aptamer design so that the fluorophore was also at the 3' end. We tested 3 designs for each aptamer that ranged from 10 to 12 nucleotides in length (Figure 4A). We specifically tested Aptamer.1 and not Aptamer.2 because we wanted to focus on improving one aptamer at a time in more detail.

Unfortunately, there was still significant quenching but no significant switching (Figure 4B). This suggests that the binding site of the target is not in this region.

3.3 New Quencher Probe Design to Bind the Center of Aptamer.1

Next, we tested our hypothesis that Hsp90 binds to the center of the aptamer sequence by designing a new quencher probe that bound the center of the aptamer. As seen in Figure 5A, the Quencher.1C25mis quencher probe (a 25 nucleotide long strand with a mismatched, non-complementary, region) has two fully complementary sequences to the aptamer, at the 5' end of the aptamer and the center of the aptamer, connected by a non-complementary thymine repeat segment that decreases the otherwise high affinity that would result from a long complementary strand.

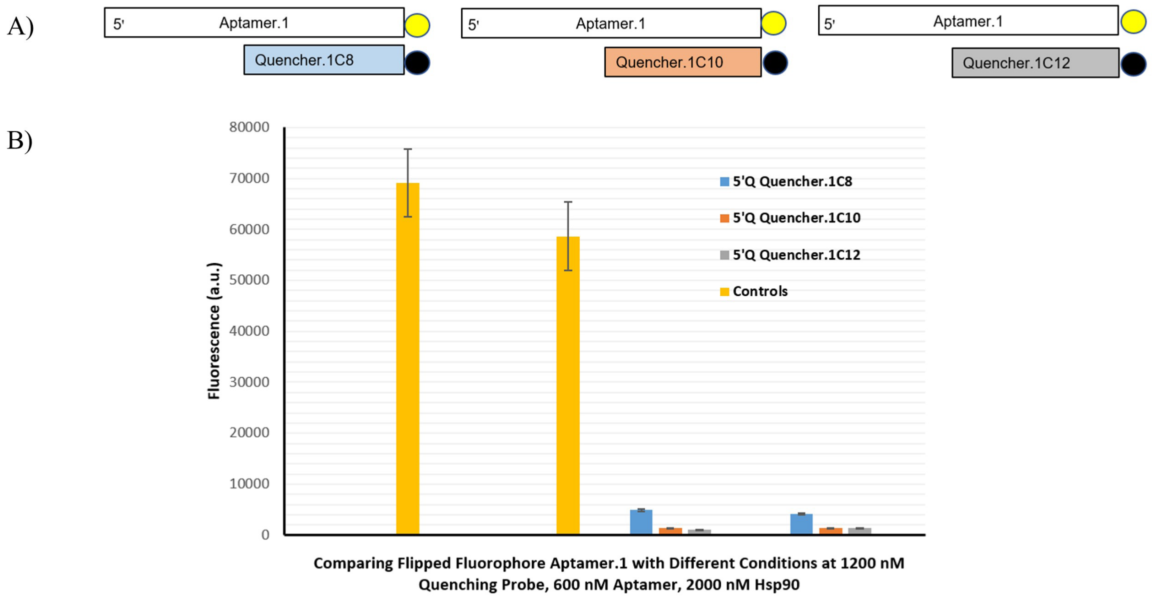
Results from this design showed significant quenching and switching as defined by an increase in fluorescence larger than error bars (Figure 5B). Moreover, a 5-fold increase in switching was seen when Hsp90 binding was favored through an increase in Hsp90 concentration and decrease in quencher probe concentration (Figure 5C), as compared to conditions in Figure 5B. More specifically, we increased the concentration of Hsp90 from 2000 nM to 5000 nM and quencher probe decreased from 1200 nM to 600 nM. This provides evidence that the Hsp90 is indeed resulting in the switch behavior. Overall, these results suggest that Hsp90 does indeed bind to the center of Aptamer.1; when the quencher probe binds to the center of the aptamer, switching occurs because the Hsp90 binds to the same area to displace the quencher probe.



Aptamer.1	X	X	X	X
Hsp90		X		X
Quencher Probe			X	X

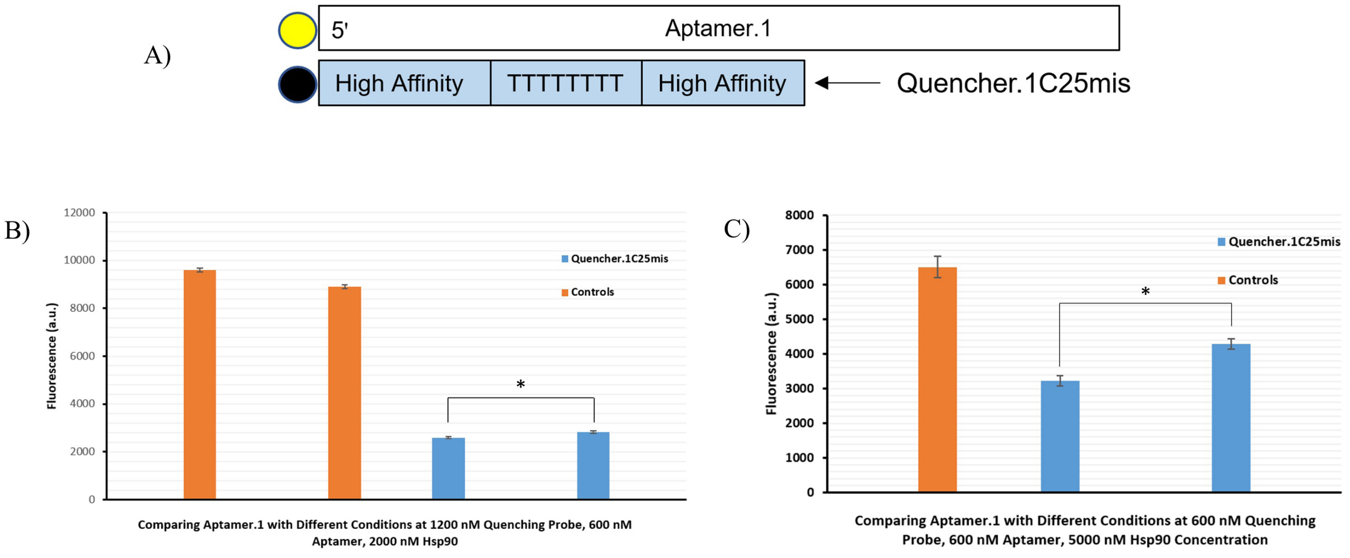
Aptamer.2	X	X	X	X
Hsp90		X		X
Quencher Probe			X	X

Figure 3. Fluorescent changes representing aptamer switching with probes that interact with the 5' end of the aptamers. A) Aptamer.1 and Aptamer.2 were each tested with 3 complementary quenching probes of different lengths. The yellow circle is the fluorescein, the black circle is the quencher. B) The Aptamer.1 aptamer-switch system showed strong quenching upon addition of all quenching probes but no significant change in fluorescence upon Hsp90 introduction. C) The Aptamer.2 aptamer-switch system showed some fluorescence quenching upon addition of the quenching probes but no change in fluorescence upon addition of the target. Error bars represent standard deviation across duplicates.



Flipped Aptamer.1	**X**	**X**	**X**	**X**
Hsp90		**X**		**X**
Quencher Probe			**X**	**X**

Figure 4. Fluorescent changes representing aptamer switching with probes that interact with the 3' end of the aptamers. A) Aptamer.1 was tested with 3 complementary probes of different lengths. B) The Aptamer.1 aptamer-switch system showed strong quenching upon addition of all probes but no significant change in fluorescence upon Hsp90 introduction. Error bars represent standard deviation across duplicates.



Aptamer.1	**X**	**X**	**X**	**X**
Hsp90		**X**		**X**
Quencher Probe			**X**	**X**

Aptamer.1	**X**	**X**	**X**
Hsp90			**X**
Quencher Probe		**X**	**X**

Figure 5. Fluorescent changes representing aptamer switching with probe that binds the center of Aptamer.1. A) Design of the new probe. There are two complementary sequences, one at the 5' end of the aptamer and one at the center. These are joined by a low-affinity segment. B) Results of the aptamer-switch system shows significant quenching and switching (p=0.02) in conditions of 2000 nM Hsp90, 1200 nM Aptamer, and 600 nM Quenching Probe. C) Results of aptamer-switch system in a new condition of raised Hsp90 to 5000 nM and decreased quenching probe to 600 nM also shows significant quenching and switching (p=0.02). Error bars represent standard deviation across triplicates. P-values were calculated using a t-test. * indicates P < .05.

Discussion

In this study we tested two aptamers for their potential to be incorporated into an aptasensor. Both aptamers displayed high affinity binding to the target Hsp90. With this information in hand, we sought to design complementary quencher probes to the aptamers that would generate a significant conformational "switch" in the presence of Hsp90. Such quencher probes must bind the same location on the aptamer as the target, at a lower affinity than the target, to ensure that the target can successfully displace the quencher probe. This displacement separates the fluorophore on the aptamer from the quencher on the quencher probe, generating an increase in fluorescence indicative of target binding.

Since the chosen aptamers have no structural or binding information, we had to test three novel sets of quencher probe designs to achieve this switching. Our initial designs included quencher probes that interacted with the 5' end of the aptamer only. These showed no change in fluorescence signal in the presence of the target, suggesting that the binding region of the aptamer was not in the 5' end region. We next tested quencher probes that interacted with the 3' end of the aptamer and again observed no binding. This suggested that the majority of the target interaction occurred in the center of the aptamer. Indeed, designing quencher probes that interact with the center of an aptamer but that function as fluorescence-quencher switches is challenging.

By making use of a polyT linker we designed a final quencher probe that bound the center of the aptamer while preserving aptamer "switching". This quencher probe was fully complementary to the aptamer in two short segments, one at the 5' end of the aptamer to maintain proximity between the fluorescein and quencher, and the other at the center of the aptamer, where the target likely binds. The polyT linker reduced the high binding affinity that would have resulted from a long complementary sequence to ensure the target could still displace the long quencher probe.

The significant switching demonstrated with Aptamer.1 and the poly-thymine containing quencher probe provides an exciting design. Before this can be translated into an aptasensor, however, it can be improved. First, the 25-nucleotide quencher probe should be shortened to reduce costs. A structural analysis approach could be used to determine the exact sequence of Hsp90 binding. More specifically, the K_D of a series of alterations to the aptamer could be made and compared to the K_D values found in this experiment. Segments that increase the K_D will be involved in Hsp90 binding, while those that do not change the K_D can be removed. Following structural analysis, the quencher location must be determined. If the quencher must be directly adjacent to the fluorophore for functionality, the fluorophore can be switched from the 5' to the 3' end of the aptamer depending on which location shortens the linker sequence connecting the quencher probe to the quencher. Alternatively, a quencher could be added directly to the short segment and tested to see if it can still act on the fluorophore despite being further away. This structural analysis approach will also provide more overall knowledge about the sequence of these aptamers that can be used towards other applications. In future avenues of research, Aptamer.2 switch systems could be tuned in more detail and compared to the current Aptamer.1 switch-system.

Furthermore, this aptamer-switch system needs to be tested in whichever bodily fluid it will be used in. This is of particular importance in the blood, as aptamers are known to denature easily in blood[18]. It must also be examined for specificity by testing in cell culture or with similar proteins such as other chaperones. Moreover, this aptamer switch system should be tested with all components (Aptamer, Quencher Probe, and Hsp90) added at the same time, instead of with aptamer and HSP90 incubated first for 30 minutes. If the system still works, this will accelerate testing. Last, the system should be tested with live malaria. If all goes well, it can be sent to a biotechnology company to be made into an aptasensor. This aptasensor can be tested in a variety of live malaria strains and compared to other diagnostic tests to assess its use and value.

Conclusion

A novel aptamer-switch system that detects Hsp90, an important protein in malaria, was tuned. Testing has optimized the quencher for the aptamer-switch system and indicated that Hsp90 likely binds to the center of the aptamer sequence. Although further studies are needed to optimize the system and test for stability and specificity, this system shows potential for development into an aptasensor. Once developed, a smartphone-compatible aptasensor could be compared to other aptasensors to determine how the novel target compares to existing targets. If superior, this aptasensor will be used in the field to detect malaria in a cheap, portable method to help combat the spread of malaria.

Acknowledgements

We would like to thank Dr. Eugene Oteng and Najeeb Khalid for useful discussions. We thank the entire McKeague Lab for support and feedback. We thank Serge Hirka for helping to troubleshoot fluorescence polarization measurements. We thank Canada Research Chairs Program and McGill University for funding.

References

1. World Health Organization. World malaria report 2019. (2019)

2. Zheng, Z. & Cheng, Z. Chapter Four – Advances in Molecular Diagnosis of Malaria. Adv. in Clin. Chem. **80**, 155-192 (2017).

3. Snow, R. Global malaria eradication and the importance of Plasmodium falciparum epidemiology in Africa. BMC Med. **13**, (2015).

4. Venugopal, K., Hentzschel, F., Valkiūnas, G. & Marti, M. Plasmodium asexual growth and sexual development in the haematopoietic niche of the host. Nat. Rev. Microbiol. **18**, 177–189 (2020).

5. Bria, Y., Yeh, C. & Bedingfield, S. Significant symptoms and nonsymptom-related factors for malaria diagnosis in endemic regions of Indonesia. Int. J. of Infect. Diseases. **103**, 194-200 (2021).

6. Trampuz, A., Jereb, M., Muzlovic, I. & Prabhu, R. M. (2003). Clinical review: Severe malaria. Critical Care. **7**, (2003).

7. Bronzan, R.N., McMorrow, M.L. & Patrick Kachur, S. (2008) Diagnosis of Malaria. Mol. Diag. Ther. **12**, 299–306 (2008).

8. Mbanefo, A. & Kumar, N. Evaluation of Malaria Diagnostic Methods as a Key for Successful Control and Elimination Programs. Trop. Med. and Infect. Disease. **5**, 102 (2020).

9. Talapko, J., Škrlec, I., Alebić, T., Jukić, M. & Včev, A. Malaria: The Past and the Present. Microorg. **7**, (2019).

10. Berzosa, P. et al. Comparison of three diagnostic methods (microscopy, RDT, and PCR) for the detection of malaria parasites in representative samples from Equatorial Guinea. Malaria J. **17**, (2018).

11. Koyun A., Ahlatcıoğlu E. & İpek Y.K. Biosensors and Their Principles. A Roadmap of Biomedical Engineers and Milestones. InTech. 117–142 (2012).

12. Mharakurwa, S., Simoloka, C., Thuma, P.E., Shiff, C. & Sullivan, D. J. PCR detection of Plasmodium falciparum in human urine and saliva samples. Malaria J. **5**, (2006).

13. Nwakanma, D. et al. Quantitative Detection of Plasmodium falciparum DNA in Saliva, Blood, and Urine. The J. of Infect. Diseases. **199**, 1567–1574 (2009).

14. Ittarat, W. et al. Biosensor as a molecular malaria differential diagnosis.

Clinical Chimica. Acta. **419**, 47-51 (2013).

15. Kumar, B., Bhalla, V., Pratap, R. & Suri, C. Label-free electrochemical detection of malaria-infected red blood cells. RSC Adv. **6**, (2016).

16. Krampa, F. D., Aniweh, Y., Kanyong, P. & Awandare, G.A. Recent Advances in the Development of Biosensors for Malaria Diagnosis. Sensors. **20**, 799 (2021).

17. Sullivan, C. Aptasensors – the future of biosensing? Anal. and Bioanal. Chem. **372**, 44-48 (2001).

18. McKeague, M. & DeRosa, M. Challenges and Opportunities for Small Molecule Aptamer Development. J. of Nucleic Acids (2012).

19. Alsaafin, A. & McKeague, M. Functional nucleic acids as in vivo metabolite and ion biosensors. Biosens. and Bioelect. **94**, 94-106 (2017).

20. Pfeiffer, F. & Mayer, G. Selection and Biosensor Application of Aptamers for Small Molecules. Front. in Chem. **4**, (2016).

21. Zhou, W., Huang, P., Ding, J. & Liu, J. Aptamer-based biosensors for biomedical diagnostics. Analyst. **11**, (2014).

22. Giannetti, A. & Tombelli, S. Aptamer optical switches: From biosensing to intracellular sensing. Sensors and Actuators Rep. **3**, (2021).

23. Feagin, T., Maganzini, N. & Soh, H. T. Strategies for Creating Structure-Switching Aptamers. ACS Sens. **3**, 1611-1615 (2018).

24. Ospina-Villa, J. D., Cisneros-Sarabia, A., Sánchez-Jiménez, M. M. & Marchat, L. A. Current Advances in the Development of Diagnostic Tests based on Aptamers in Parasitology: A Systematic Review. Pharmaceutic. **12**, 1046 (2020).

25. Lee, S., Song, K., Jeon, W., Jo, H., Shim, Y. & Ban C. A highly sensitive aptasensor towards Plasmodium lactate dehydrogenase for the diagnosis of malaria. Biosens. Bioelectron. **35**, 291-296 (2012).

26. Jain, P., Das, S., Chakma, B. & Goswami, P. Aptamer-graphene oxide for highly sensitive dual electrochemical detection of Plasmodium lactate dehydrogenase. Anal. Biochem. **514**, 32-27 (2016).

27. McCutchan, T. F., Piper, R. C. & Makler, M.T. Use of malaria rapid diagnostic test to identify Plasmodium knowlesi infection. Emerging infectious diseases **14**, 1750–1752 (2008).

28. Sanjay, M., Singh, N., Ngashangva, L. & Goswami, P. A smartphone-based fiber-optic aptasensor for label-free detection of Plasmodium falciparum glutamate dehydrogenase. Anal. Methods **12**, 1333-1341 (2020).

29. Kori, L., Valecha, N. & Anvikar, A. Glutamate dehydrogenase: a novel candidate to diagnose Plasmodium falciparum through rapid diagnostic test in blood specimen from fever patients. Sci. Rep. **10**, 6307 (2020).

30. Oteng, E., Gu, Wm. & McKeague, M. High-efficiency enrichment enables identification of aptamers to circulating Plasmodium falciparum-infected erythrocytes. Sci. Rep. **10**, (2020).

31. Pesce, E-R., Cockburn, I. L., Goble, J. L., Stephens, L.L. & Blatch, G. Gl. Malaria heat shock proteins: drug targets that chaperone other drug targets. Infect. Disord. Drug Targets. **10**, 147-57 (2010).

32. Bzowska, M., Nogiec, A., Bania, K., Zygmunt, M., Zarebski, M., Dobrucki, J. & Guzik, K. Involvement of cell surface 90 kDa heat shock protein (HSP90) in pattern recognition by human monocyte-derived macrophages. J. Leukoc. Biol. **102**, 763-774 (2017).

33. Sefah, K., Shangguan, D., Xiong, X., O'Donoghue, M. & Tan, W. Development of DNA aptamers using Cell-SELEX. Nature Protocols. **5**. (2010).

34. Hirka, S. & McKeague, M. Quantification of aptamer-protein binding with fluorescence anisotropy. Aptamers. (2021).

35. McKeague, M., et al. Comprehensive analytical comparison of strategies used for aptamer evaluation. Analyt. Chem. **87**, 8608–8612 (2015).

Submitted: 01/24/2022
Accepted: 04/05/2022

Research Article

Emmanuelle Melis[1,*], Emma Armitage[1,*], Yuxin Ma[1], Amelia Weiss[1], and Brian Robinson[1,**]

[1] Department of Geography, McGill University, Montreal, QC, Canada

Spatial distribution and socioeconomic differences between urban farms' production and distribution points in Chicago, IL

Keywords

Chicago, choropleth, urban agriculture, food security, socioeconomic, spatial distribution

Email Correspondence

emmanuelle.melis@mail.mcgill.ca
emma.armitage@mail.mcgill.ca

Abstract

Urban farming remains popular as a potential sustainable replacement or supplement for traditional agricultural models, but little comprehensive research has been done regarding the socioeconomic characteristics of this food production and distribution model. Using the City of Chicago as a case study, this research seeks to understand where urban commercial farms are located and whether there exists a significant disparity between the social demographics of the neighbourhoods where these farms produce crops compared to where their goods are sold. The distribution of urban farm and sale points was determined by geocoding the location of all production (farms) and distribution points (sale points) for commercial urban farming companies in Chicago, then calculating spatial statistics and calculating the mean centers, standard distance, and standard deviational ellipses (SDE) for each. These were then overlaid onto choropleth maps containing socioeconomic indicator data derived from the US 2016 census. These socioeconomic indicators — median annual household income, mean home value, and percent racialized minority population — were analysed to determine if a correlation exists between each socioeconomic indicator and the location of farm and sale points. Findings reveal statistically-significant differences in the socioeconomic indicators of census tracts of farm versus sale point locations, showing a skewness in distribution of farm locations towards areas of lower socioeconomic status versus a skewness in distribution of sale point locations towards areas of higher socioeconomic status. The results suggest that, while farms are more likely to be located in marginalized neighbourhoods in Chicago, most produce grown by these farms is sold in more privileged areas.

Introduction

Food inequality occurs at every scale of production and distribution. At international and local scales, food tends to be produced in lower-cost areas and then redistributed to higher-value areas for sale. In North America, food is largely produced in poorer rural areas, and then sold in comparatively wealthier cities, but the food distribution and access within these cities can vary wildly[3]. This variation in food access is referred to as food insecurity, which is defined as "a lack of consistent access to enough food for an active, healthy life"[4] and is a major problem in many North American cities, including Chicago, where 600,000 residents are food-insecure[5]. Food deserts, areas of high food insecurity where residents have little or no access to stores and restaurants that provide fresh, healthy, and inexpensive foods[6], are particularly prevalent in marginalized communities[7,8] within cities. In an effort to make food systems and cities more sustainable, many have looked to urban farming, believing that bringing food production closer to the people it serves may reduce disparity. Urban agriculture, the practice of cultivating, processing, and distributing food in or around urban areas, can be divided into two main categories: private gardens and farms[9]. Although they use the same intensive cultivation methods as other commercial agriculture, commercial urban farms are different in that they produce food on a local scale in a metropolitan setting for widespread distribution to retailers within the city.

Some have suggested that urban farms could serve to address food insecurity in impoverished neighbourhoods. Thus, urban farms have been a popular model of social and environmental sustainability amongst city planners. However, it remains unclear whether residents living proximally to such farms benefit from the farms' presence and outputs. Little is understood about where and how urban farmers distribute their food, and, most importantly, to whom[10]. Since food insecurity is a matter of physical as well as social and financial distance from healthy produce, there is potential for urban farms to address food deserts[11], but there are also concerns these farms may do more harm than good. Studies show that many city greening projects, such as urban agriculture, result in eco-gentrification - when property values rise surrounding urban greening projects, displacing local businesses and residents[12]. The presence of farms in urban areas may serve to take up space while transporting food to wealthier neighbourhoods or providing food that is financially inaccessible.

In order to examine the potential socioeconomic inequalities underlying distribution of commercial urban farms and sale points, Chicago, Illinois, one of the most segregated cities in the United States[13], was selected as the study area for this study. Concurrently, through advocacy, policies, and projects, Chicago has made efforts to build a local, sustainable urban food system[14]. The city has proved fertile ground for the study of urban agriculture, with over 890 farms, gardens, and other initiatives[15], including a number of commercial farms that serve as the basis for our research. Therefore, Chicago lends itself well to examining if there are social or economic differences between the location of urban commercial farms and their sale points.

When considering the social implications of urban agriculture, there is a bias in the research towards smaller-scale efforts like community gardens[10], which leaves a wide knowledge gap regarding commercial urban farms and their social impacts. Therefore, our research aimed to examine any socioeconomic differences between urban commercial farm locations and their sale points in the city of Chicago.

The purpose of our research was two-fold:

- Firstly, this research aimed to understand where urban farms and their sale points are located and if there is any pattern to their distribution throughout the city of Chicago.

- Secondly, if there is a pattern, this research investigated if any disparities exist between the socioeconomic indicators of areas where urban farms are located, versus the areas where their sale points are located.

---

* Contributed equally and are co-first authors.
** Principal Investigator.

This research would contribute to understanding if there is a tendency for urban farms to be built in socially vulnerable areas while serving more privileged communities.

Methodology

2.1 Spatial Data

First, a spreadsheet of commercial urban agricultural production points (farms) and distribution points (sale points) was obtained. Farms were classified as places where produce is grown, and sale points as where the produce is sold. All urban agriculture projects in Chicago were located using the CUAMP database[16] and selected from locations classified as "Urban Farms". From there, each location was manually reviewed and filtered so that all selected results all met the following criteria:

- Farms that grow produce for sale;

- Farms that are not community gardens;

- Farms that have sufficient available information (location, publically accessible website, published distribution/sale points).

These criteria generated a list of 27 farms within Chicago's municipal boundaries. To determine the farms' sale points, data about sale points was collected from each farm's website to see where the produce was sold. Sale points were used as a proxy for access to produce grown by urban commercial farms. In total, there were 112 sale points. Each farm and sale points' address, name, larger parent company, and ZIP code were entered into a spreadsheet.

For spatial analysis, the production and distribution points were geocoded using GIS software via ArcGIS World Geocoding Service. Geocoding is the process of assigning an XY coordinate pair to the description of a place[17].

2.2 Spatial Statistical Methods

To measure differences between production and distribution points, three spatial statistical measures were calculated: mean center, standard distance, and standard deviational ellipse (SDE). Mean center, a widely-used measure of central tendency in point features, used all farm and sale points, respectively, to determine the mean location of each. Standard distance and SDE are measures of spatial dispersion. To determine the amplitude of spatial spread of data points, the standard distance, which is the radius of one standard deviation from the mean center, was calculated from the mean center of farm and sale points. Because standard distance does not capture directional bias, the SDE was calculated for urban commercial farm locations and sale point locations using the mean center and dispersions along a major and a minor axis in order to define the direction of maximum spread of the distribution[18]. An asymmetrical and elliptical SDE indicates skewness and bias in spatial distribution and allows for visual interpretation of the skewness of spatial points.

To visually determine the farm and sale points' respective distributions in relation to the socioeconomic make-up of different Chicago neighbourhoods, we created a choropleth map using census data published from the 2016 US Census[19]. Indicators of socioeconomic status were median annual household income, mean home value, and percent racialized minority population, a measure that represents the non-white population percentage. The standard distance and SDE for the comparison of farm and sale points' distributions to one another as well as to their census tract information. Next, statistical analyses were performed to provide further evidence in support of the visual relationships.

2.3 Statistical Analyses

A spatial join was performed in GIS to relate census tract demographic data with farm and sale point ZIP code locations. Then, "social difference" was analyzed to determine the difference between the socioeconomic indicators of farm and sale point locations. First, the aggregate mean values for each indicator were calculated for all the farms and sale points and then compared on a 1:1 basis. An identical operation was performed for median values.

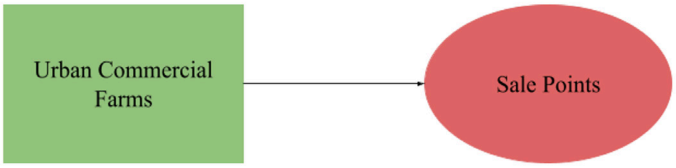


Figure 1. The green box on the left represents the aggregate mean or median data for all 27 farms, and the red oval on the right represents the aggregate mean or median data for all 54 sale points. By comparing the data on a 1:1 basis, differences between socioeconomic indicators were numerically determined.

For more granularity, social difference was analyzed in a 1:many relationship, in which each of the 27 farm locations was considered individually. Here, farm location was assumed to distribute produce equally to all associated sale points. To compare the socioeconomic indicators between individual farm locations and their associated sale points, the mean value of each indicator was calculated for all sale points (many) associated with each individual farm (one) and repeated across each of the 27 farms.

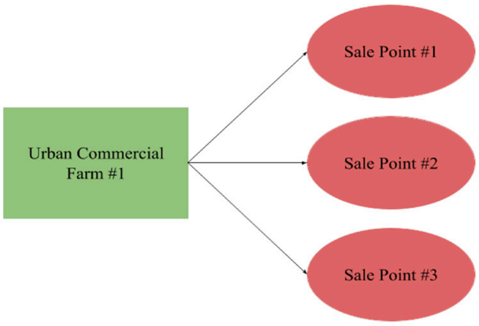


Figure 2. The green box represents a single farm location and its corresponding indicator data. The red ovals represent associated sale point locations and their corresponding indicator data. Sale point indices were averaged and compared to singular urban farm indices. Comparing data on a 1:many basis demonstrates socioeconomic indicator differences for each farm location.

The general premise of the social difference calculation is to subtract the mean or median value for each socioeconomic indicator of farm locations from the same socioeconomic indicator of sale point locations. The result demonstrates the indicator's direction of change:

$$\textit{Social Difference}_{(general)} = \textit{Sale Point Socioecon. Indicator} - \textit{Farm Socioecon. Indicator}$$

To ensure the direction of change was not influenced by the calculation method, calculations were performed differently for each conceptual relationship.

For the 1:1 relationship, the mean or median value of the socioeconomic indicator for all farms was subtracted from the mean or median value of the same indicator for all sale points.

$$\textit{Social Difference}_{(1:1\ mean)} = \textit{Mean(Socioecon. Indicator for Sale Points)} - \textit{Mean(Socioecon. Indicator for Farms)}$$

$$\textit{Social Difference}_{(1:1\ median)} = \textit{Median(Socioecon. Indicator for Sale Points)} - \textit{Median(Socioecon. Indicator for Farms)}$$

For the 1:many relationship, farms' socioeconomic indicator value was subtracted from the associated sale point mean indicator value.

$$\textit{Social Difference}_{(1:many)} = \textit{Mean(Socioecon. Indicator for Associated Sale Points)} - \textit{Farm Socioecon. Indicator}$$

A table (Table 1) was created consisting of 27 rows representing each farm, and columns representing the change in each socioeconomic indicator moving from farm to sale point. Mean and median social difference values for each socioeconomic indicator were calculated from this table.

2.4 Test for Statistical Significance

The Kruskal-Wallis H-test was used to determine the statistical significance of socioeconomic indicators' difference. The Kruskal-Wallis test is a rank-based nonparametric test used to determine if there are statistically significant differences between the medians of two or more groups[18]. Given that the data in this analysis were very skewed, we opted to conduct the Kruskal-Wallis H-test, which makes no assumptions about normality[18]. The Kruskal-Wallis H-test assumes the observations in each group come from populations with the same distribution shape, that the samples are independent, and that the dependent variable is continuous[18], assumptions with which our data aligns.

Within the 1:1 framework, the Kruskal-Wallis test was conducted to assess if the difference between medians is statistically significant. Assessing inferential statistics of the 1:many relationship is outside the scope of this research question. Rather, we seek to determine if there is a difference in aggregate between socioeconomic indicators of farms and sale points.

2.5 Calculation Method

Procedurally, the Kruskal-Wallis H-test involves pooling observations (k) from the samples into one combined sample, then ranking them in ascending order from 1 to N where $N=n_1+n_2+n_3+\ldots+n_k$[20]. When ranking, the sample from which each observation originated from is tracked[20]. Next, the test statistic, H, is calculated:

$$H = \left(\frac{12}{N(N+1)}\sum_{j=1}^{k}\frac{R_j^2}{n_j}\right) - 3(N+1)$$

Where N=the total sample size, n_j=sample size of the jth group, and R_j=the sum of the ranks in the jth group[20]. The Kruskal-Wallis H-test was conducted using R, at p-values of 0.10, 0.05, and 0.01, testing for different levels of statistical significance.

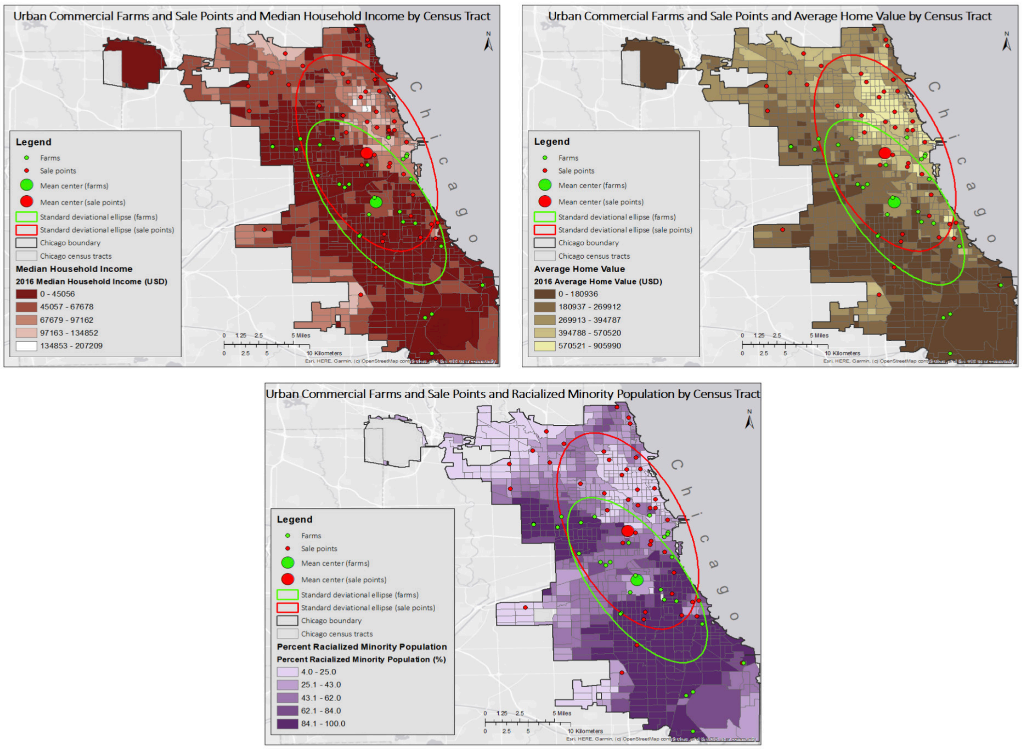
Results

3.1 Mean Center and Standard Distance

As can be seen in Figures 1-3, the distribution of farms and of sale points differ visually, which is shown by the different mean center locations and different standard distance radii. The sale points' standard distance from the mean center is larger than for the urban commercial farm locations, indicating that sale points are more dispersed. Conversely, urban commercial farm locations are more densely clustered.

3.2 Standard Deviational Ellipse

Additionally, Figures 1-3 depict the results of the SDEs for farm and sale points. Chicago's municipal borders and lake shoreline contribute to the city's northwest-southeast leaning shape, which affect the SDE narrowness and diagonal skewness. The differences between the SDEs can be visually judged, however, by the size of their major and minor axes and by the ellipses' location around their mean centers. The major and minor axes of the sale points' ellipse are greater in size than those of the farms' ellipse, indicating a large north-south (X) and east-west (Y) spread. Conversely, the major and minor axes of the farm points are smaller, indicating more spatial clustering in a smaller north-south (X) and east-west (Y) range. When comparing the skewness visually, it is obvious that sale point locations slant toward the north of Chicago, and census tracts with higher socioeconomic indicators. Conversely, urban commercial farm locations are located in the centre and south parts of the city, skewing toward census tracts with lower socioeconomic indicators. Statistical analysis in the following sections will numerically support this relationship.



Urban Commercial Farms and Sale Points and Median Household Income by Census Tract

Urban Commercial Farms and Sale Points and Average Home Value by Census Tract

Urban Commercial Farms and Sale Points and Racialized Minority Population by Census Tract

Figure 3. (Clockwise). All figures show the spatial distribution of urban commercial farms and sale point locations in Chicago overlaid on census tracts showing various socioeconomic indicators. The maps show the mean center and the standard deviational ellipse for farms and sale points respectively. A) Median Household Income by Census Tract. B) Average Home Value by Census Tract. C) Percent Racialized Minority by Census Tract.

3.3 Maps and Spatial Distribution of Socioeconomic Indicators

Median Annual Household Income by Census Tract

Figure 3A shows the median annual household income by census tract. Breaks in median income were determined using US Federal tax brackets[21]. The mean center of the production points (farms) is located in a lower-income neighbourhood, compared to that of the distribution points (sale points). Additionally, the SDE of sale points skews toward higher median income census tracts. Conversely, the SDE of farm points is narrower, primarily encompassing census tracts that fall within the two lowest income tax brackets.

Mean Home Value by Census Tract

Figure 3B shows the mean home value by census tract. The mean centers are in census tracts with similar home values: $269,913-$394,787 for commercial urban farms and $394,788-$570,250 for sale points. The SDE of sale points skews toward census tracts with higher mean home values, compared to the SDE of farm locations.

Percent Racialized Minority Population by Census Tract

Finally, Figure 3C shows the percent racialized minority population by census tract. The racialized minority populations are similar for the census tracts underlying each mean center. However, similar to median income and mean home value, the sale points SDE skews toward census tracts with smaller racialized minority populations, whereas farms' the SDE skews towards census tracts with higher percent racialized minority populations.

3.4 Results from Relationship Frameworks and Kruskal-Wallis H-Test

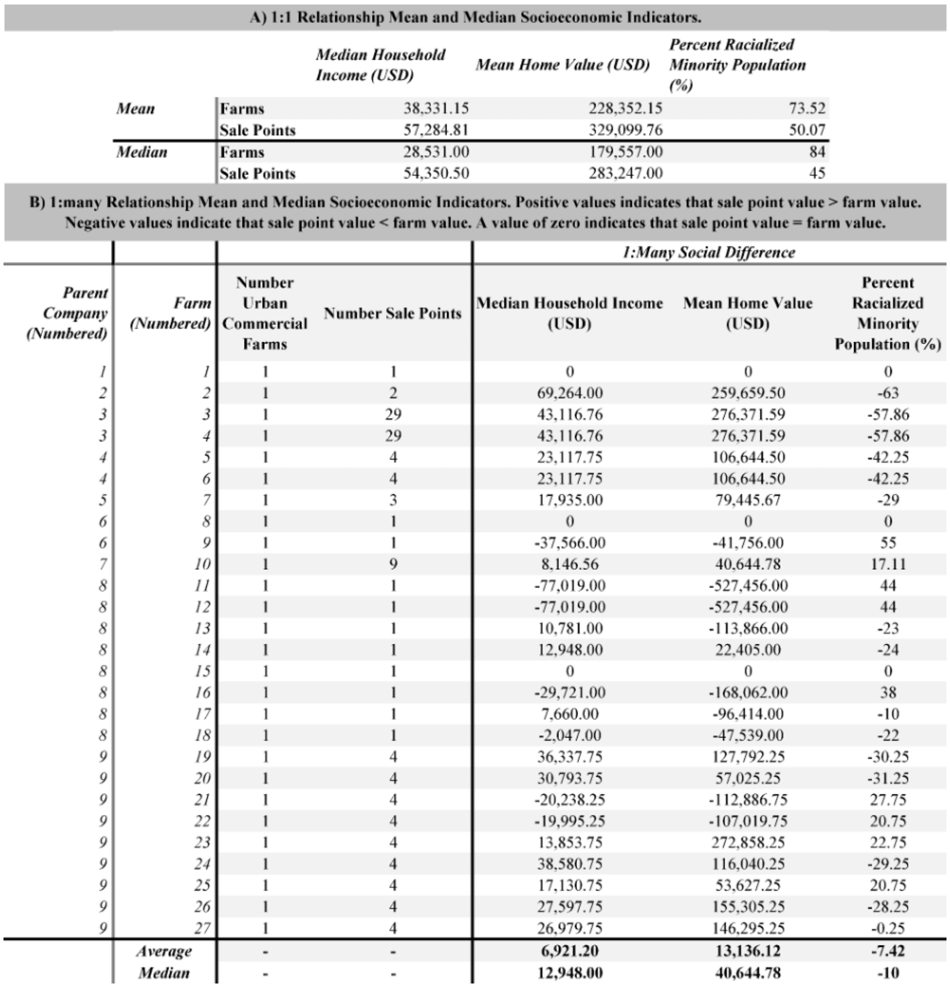
Descriptive Statistics – 1:1 Relationship

Table 1A below shows that, across mean and median values, sale points are consistently located in areas with higher median household incomes, higher mean home value, and lower percent racialized minority population. The exact opposite trend is observed for the socioeconomic indicators of farm locations.

Descriptive Statistics – 1:many Relationship

As indicated by Table 1B below, there is much variation in social difference when examining the socioeconomic indicators on a farm-by-farm basis, compared to those of its sale points. Social difference results of the 1:many relationship include negative, positive or zero values. A positive value indicates that sale points' socioeconomic indicator is higher than farms' socioeconomic indicator. Conversely, a negative value indicates that sale points' socioeconomic indicator is lower than for farms. A zero value indicates no change. More research should be conducted to understand the reasoning behind the variation. However, despite the variation, the mean and median values of the 1:many social differences support the 1:1 relationship findings above. Median household income, mean home value, and percent white population are higher on average where sale points are located, indicated by the positive values in Table 1B. Conversely, percent racialized minority population is lower in areas with sale points than it is in areas with urban commercial farms.

Table 1. (Top to bottom) A) Mean and Median Values of Socioeconomic Indicators as calculated in the 1:1 Relationship to compare farms and sale points on aggregate; B) 1:many Relationship mean and median values of socioeconomic indicators. A positive value indicates that sale points' socioeconomic indicator is higher than farms' socioeconomic indicator. Conversely, a negative value indicates that sale points' socioeconomic indicator is lower than for farms. A zero value would indicate no change.



A) 1:1 Relationship Mean and Median Socioeconomic Indicators.

		Median Household Income (USD)	*Mean Home Value (USD)*	*Percent Racialized Minority Population (%)*
Mean	**Farms**	38,331.15	228,352.15	73.52
	Sale Points	57,284.81	329,099.76	50.07
Median	**Farms**	28,531.00	179,557.00	84
	Sale Points	54,350.50	283,247.00	45

B) 1:many Relationship Mean and Median Socioeconomic Indicators. Positive values indicates that sale point value > farm value. Negative values indicate that sale point value < farm value. A value of zero indicates that sale point value = farm value.

				1:Many Social Difference		
Parent Company (Numbered)	***Farm (Numbered)***	**Number Urban Commercial Farms**	**Number Sale Points**	**Median Household Income (USD)**	**Mean Home Value (USD)**	**Percent Racialized Minority Population (%)**
1	*1*	1	1	0	0	0
2	*2*	1	2	69,264.00	259,659.50	-63
3	*3*	1	29	43,116.76	276,371.59	-57.86
3	*4*	1	29	43,116.76	276,371.59	-57.86
4	*5*	1	4	23,117.75	106,644.50	-42.25
4	*6*	1	4	23,117.75	106,644.50	-42.25
5	*7*	1	3	17,935.00	79,445.67	-29
6	*8*	1	1	0	0	0
6	*9*	1	1	-37,566.00	-41,756.00	55
7	*10*	1	9	8,146.56	40,644.78	17.11
8	*11*	1	1	-77,019.00	-527,456.00	44
8	*12*	1	1	-77,019.00	-527,456.00	44
8	*13*	1	1	10,781.00	-113,866.00	-23
8	*14*	1	1	12,948.00	22,405.00	-24
8	*15*	1	1	0	0	0
8	*16*	1	1	-29,721.00	-168,062.00	38
8	*17*	1	1	7,660.00	-96,414.00	-10
8	*18*	1	1	-2,047.00	-47,539.00	-22
9	*19*	1	4	36,337.75	127,792.25	-30.25
9	*20*	1	4	30,793.75	57,025.25	-31.25
9	*21*	1	4	-20,238.25	-112,886.75	27.75
9	*22*	1	4	-19,995.25	-107,019.75	20.75
9	*23*	1	4	13,853.75	272,858.25	22.75
9	*24*	1	4	38,580.75	116,040.25	-29.25
9	*25*	1	4	17,130.75	53,627.25	20.75
9	*26*	1	4	27,597.75	155,305.25	-28.25
9	*27*	1	4	26,979.75	146,295.25	-0.25
	Average	-	-	**6,921.20**	**13,136.12**	**-7.42**
	Median	-	-	**12,948.00**	**40,644.78**	**-10**

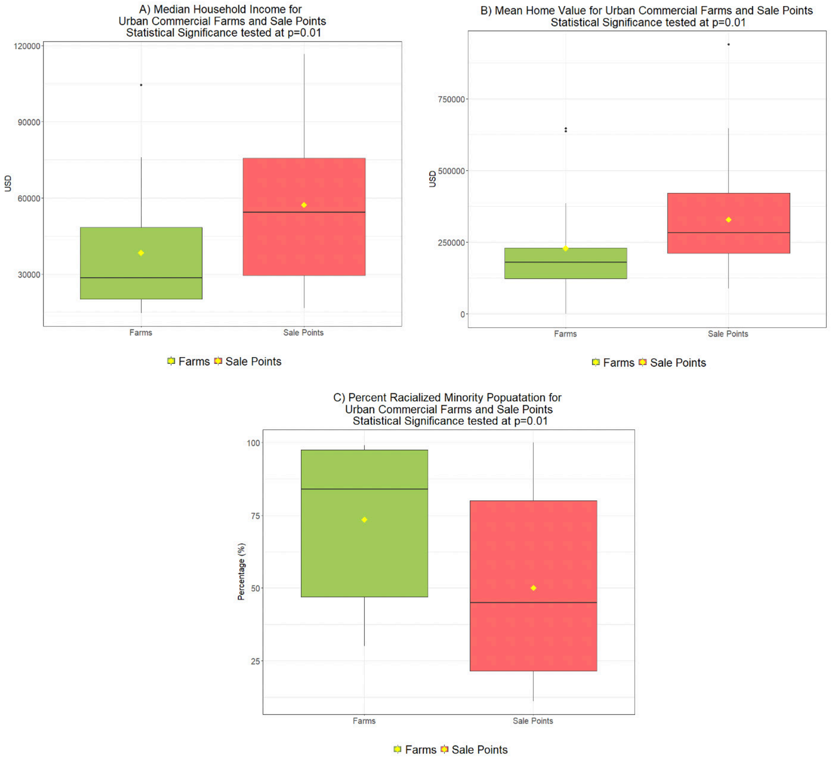


Figure 4. (Clockwise). Boxplots showing the mean and median values for each socioeconomic indicator as calculated by the 1:1 relationship. Here, all farms were compared to all sale points on aggregate and tested for statistical significance at p=0.01. A) Boxplot showing Median Household Income. B) Boxplot showing Mean Home Value. C) Boxplot showing Percent Racialized Minority.

Inferential Statistics

Additionally, to quantify the visual relationships that the maps indicate, the Kruskal-Wallis H-test was conducted for the 1:1 relationship. The boxplots in Figure 4 show the mean and median results for each socioeconomic indicator. By the Kruskal-Wallis H-test, the null hypothesis that the farm and sale point median values are equal is rejected for every socioeconomic indicator at 99% confidence. As such, sale point locations are significantly correlated (p=0.01) with census tracts of higher median household income, greater mean home value, and lower percent racialized minority population. Farm locations are significantly (p=0.01) correlated with the opposite trend.

3.5 Social Difference Results

Across different calculation methodologies, the same trend is repeated. Higher indicators of socioeconomic status are found across both mean and median 1:1 and 1:many calculations, and are found to be statistically significant. As such, our findings (Table 2) are not a result of a specific calculation methodology. Therefore, it can be concluded that sale point locations are significantly correlated with areas of higher socioeconomic status whereas farm locations are significantly correlated with areas of lower socioeconomic status, and thus, that there is a "social difference" between areas of production and of distribution.

Table 2. Mean and median social difference values for 1:1 and 1:many relationships. Meaning of positive values for socioeconomic indicators: sale point value > farm value. Meaning of negative values for socioeconomic indicators: sale point value < farm value.

Summary of Social Differences

	Median Household Income (USD)	Mean Home Value (USD)	Percent Racialized Minority Population (%)
1:1 Mean	18953.67	100747.61	-23.44
1:1 Median	25819.50	103690.00	-39.00
1:many Mean	6921.20	13136.12	-7.42
1:many Median	12948.00	40644.78	-10.00

Discussion

4.1 Interpretation and Implications

Hoping to narrow food access inequalities by bringing production into cities and increasing the supply of fresh produce in food-insecure areas, urban farms have been touted as a model of environmental and social sustainability. However, comparatively little research has examined distribution patterns of food grown in urban locales. As such, this paper examines if there are underlying socioeconomic differences between where food is grown versus where food is distributed in Chicago, Illinois.

There are two key findings, both of which have broader implications for the study of urban farming in major metropolitan cities such as Chicago. First, the majority of urban farms in the city appear to be located in areas of lower socioeconomic status. This holds true for all socioeconomic indicators, including median annual household income, mean home value, and percent racialized minority and white population. Second, while farms may grow most of their produce in lower-income areas, it is largely being sold in areas of higher socioeconomic status. The majority of sale points associated with the farms analyzed in this study were located in regions with a higher mean home value, higher median annual household income, and lower percent visible minority.

It appears that whiter, wealthier neighbourhoods might have more access to fresh produce being grown within the bounds of the City of Chicago, despite that food being grown in marginalized communities at risk of food insecurity. The benefit produced by these farms, in the form of healthy food, is being transferred out of the neighbourhoods that provide these farms with infrastructure use and space, and into more privileged areas.

It is clear that living proximally to an urban farm does not secure access to fresh, healthy food which challenges the notion that urban commercial farms help alleviate food insecurity. One of the many hopes of urban agriculture is that producing food within cities could help supplement the nutritiously poor and limited diets of residents in urban food deserts[11,22,23]. Yet, even when farms are directly located in areas associated with higher food insecurity, they do not seem to reduce the physical barriers of access that plague residents of food deserts within Chicago.

Our findings suggest that higher-income neighbourhoods may benefit the most from urban farms, thereby challenging the idealized image of urban farming, often seen in environmental and public spaces as a panacea to food insecurity and conventional farming.

However, the findings of this study suggest that urban commercial farms are not serving the communities where they are located, at least not with the food they are growing. The socioeconomic difference between farm location and sale point location means that there is an externalisation of costs to poorer neighbourhoods and a net transfer of benefits to more privileged populations. Our research shows that commercial urban agriculture, as implemented in Chicago, selects for and reinforces the same conditions found in conventional agricultural systems — production in low-cost areas and distribution to higher-income areas. We propose that this urban agricultural model in Chicago has not actually served the purpose of improving equity of food access but has instead scaled down and localized this systemic failure to the city of Chicago. If we are to endorse the current urban agriculture model as a solution to feeding future cities, then we must take a sober and critical approach to the existing systems.

This research was conducted in Chicago, Illinois, which is located in a developed country in North America, inevitably restricting any implications for questions of urban agriculture in developing nations. Also, Chicago is one of the most segregated cities in the world[13], potentially amplifying any inequalities demonstrated by this case study regarding urban farms. While our work provides a foundation for potential future research, one must be cautious in generalizing our findings to other urban areas. Additionally, this is a cross-sectional study that looks at the state of urban farming in 2016. Inferences regarding changes in possible socioeconomic indicators over time potentially due to the creation of urban commercial farms are outside the scope of this study.

This research does not account for any outputs that urban commercial farms may produce beyond the sale of produce. Many of these companies have stated social missions, including community-building, agricultural education, and employment, implying that their benefit might extend beyond physical products[25-31]. The impact these additional social programs might have on residents of neighbourhoods adjacent to these farms is outside of the scope of this study.

Finally, this study only considers physical access to produce, not financial access. As mentioned before, food deserts can be caused by a lack of healthy food within a physically accessible vicinity, or they can occur if produce is physically available but financially inaccessible[32].

There are a number of avenues for potential future research. An exploration of the financial accessibility of urban commercial farm produce will provide further insights into these companies' capacity to address food insecurity and whether urban-produced food is more expensive than traditionally produced food. Furthermore, with the understanding provided in this paper that urban farms may be more likely to reside in low-income neighbourhoods, future research could explore possible eco-gentrification as a result of massive farms purchasing real-estate in these neighbourhoods. Finally, additional research is needed to understand the full social implications of urban farms, specifically with regards to their purported social missions[10].

Conclusion

Using Chicago as a case study, this research sought to understand where urban commercial farms are located and whether there exists a significant disparity in the social demographics of neighbourhoods where urban commercial farms produce crops, compared to where their goods are sold. Our results proved the existence of this disparity, showing that the majority of Chicago urban farms appear to be located in areas with lower socioeconomic status while their produce is largely being sold in areas of higher socioeconomic status. Living proximally to an urban farm does not secure access to fresh and healthy produce, because the benefits of the commercial urban farms are being transferred to more privileged communities. Our findings challenge the mainstream perception that the urban farm model is a sustainable solution to food inequality and suggest that it simply condenses and reinforces the pre-existing inequities inherent in conventional food systems.

While acknowledging our limitations, we hope that our findings will open up future research avenues, contributing to the current understanding of the socio-economic impacts of commercial urban farms and providing a knowledge base to create a more socially-just urban agriculture model.

Acknowledgements

Thank you to Brian Robinson for guiding our research and to Tim Elrick and Kareem Hammami for ArcGIS support.

References

1. FAO, IFAD, UNICEF, WFP & WHO. The State of Food Security and Nutrition in the World 2021. (Rome, Italy, 2021).

2. Deaton, A. Policy Implications Of The Gradient Of Health And Wealth. *Health Affairs* **21**, 13-30, doi:10.1377/hlthaff.21.2.13 (2002).

3. Ruel, M. (Center for Strategic and International Studies, 2020).

4. USDA. *Definitions of Food Security*, <https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/definitions-of-food-security.aspx> (2021).

5. Chicago Sun-Times. in *Chicago Sun-Times* (2021).

6. Cummins, S. & Macintyre, S. "Food deserts"—evidence and assumption in health policy making. *BMJ* **325**, 436-438, doi:10.1136/bmj.325.7361.436 (2002).

7. Patterson, J. G., Russomanno, J., Teferra, A. A. & Jabson Tree, J. M. Disparities in food insecurity at the intersection of race and sexual orientation: A population-based study of adult women in the United States. *SSM - Population Health* **12**, 100655, doi:https://doi.org/10.1016/j.ssmph.2020.100655 (2020).

8. Jha, M. K. Food security in perspective: the significance of social action. *Community Development Journal* **44**, 351-366 (2009).

9. Bailkey, M. & Nasr, J. From brownfields to greenfields: Producing food in North American cities. *Community Food Security News* (2000).

10. Santo, R., Palmer, A., & Kim, B. Vacant Lots to Vibrant Plots: A Review of the Benefits and Limitations of Urban Agriculture., (Johns Hopkins Center for a Livable Future, 2016).

11. Meenar, M. R. & Hoover, B. M. Community Food Security via Urban Agriculture: Understanding People, Place, Economy, and Accessibility from a Food Justice Perspective. *Journal of Agriculture, Food Systems, and Community Development* **3**, 143-160, doi:https://doi.org/10.5304/jafscd.2012.031.013 (2012).

12. Haffner, J. in *The Guardian. Guardian News and Media* (2015).

13. Menendian, S., Gambhir, S., & Gailes, Arthur. *The Roots of Structural Racism Project.*, <https://belonging.berkeley.edu/roots-structural-racism> (2021).

14. Rosing, H. & Block, D. R. Farming Chicago: Prospects for Higher Education Support of Sustainable Urban Food Systems in the U.S. Heartland. *Metropolitan Universities* **28**, 27-46, doi:https://doi.org/10.18060/21464 (2017).

15. Advocates For Urban Agriculture. *Chicago Urban Agriculture Mapping Project.*, <https://www.auachicago.org/chicago-urban-agricul-

ture-mapping-project/> (2018).

16 CUAMP. Chicago Urban Agriculture Mapping Project. DataMade (2022).

17 Zandbergen, P. A. A comparison of address point, parcel and street geocoding techniques. Computers, Environment and Urban Systems 32, 214-232, doi:https://doi.org/10.1016/j.compenvurbsys.2007.11.006 (2008).

18 Ayhan, I. & Mert Cubukcu, K. Explaining historical urban development using the locations of mosques: A GIS/spatial statistics-based approach. Applied Geography 30, 229-238, doi:https://doi.org/10.1016/j.apgeog.2009.05.002 (2010).

19 Vaillancourt, C. Chicago_Tracts. ESRI Demographics (2016).

20 Ostertagová, E., Ostertag, O. & Kováč, J. Methodology and Application of the Kruskal-Wallis Test. Applied Mechanics and Materials 611, 115-120, doi:10.4028/www.scientific.net/AMM.611.115 (2014).

21 LaMorte, W. W. Nonparametric tests - Tests with More than Two Independent Samples. , <https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/bs704_nonparametric7.html> (2017).

22 Orem, T. 2021-2022 federal income tax brackets & tax rates., <https://www.nerdwallet.com/article/taxes/federal-income-tax-brackets> (2021).

23 Siegner, A., Sowerwine, J. & Acey, C. Does Urban Agriculture Improve Food Security? Examining the Nexus of Food Access and Distribution of Urban Produced Foods in the United States: A Systematic Review. Sustainability 10, 2988 (2018).

24 Hagey, A., Rice, S., & Flournoy, R. Growing Urban Agriculture: Equitable Strategies and Policies for Improving Access to Healthy Food and Revitalizing Communities., (2012).

25 Chicago Lights. Building Brighter Futures. Chicago Lights (2022). Available at: https://chicagolights.org/. (Accessed: 30th March 2022)

26 Chicago Patchwork Farms. Chicago Patchwork Farms (n.d.). Available at: http://chicagopatchworkfarms.com/. (Accessed: 30th March 2022)

27 City Farm Chicago. Sustainable Urban Agriculture. City Farm Chicago (n.d.). Available at: https://cityfarmchicago.org/. (Accessed: 30th March 2022)

28 Growing Home. Our impact. Growing Home (2022). Available at: https://www.growinghomeinc.org/our-impact/. (Accessed: 30th March 2022)

29 Heartland Alliance. Chicago FarmWorks. Heartland Alliance (n.d.). Available at: https://www.heartlandalliance.org/program/chicago-farmworks/. (Accessed: 30th March 2022)

30 Urban Growers Collective. Food Justice & Policy. Urban Growers Collective (2021). Available at: https://urbangrowerscollective.org/food-justice/. (Accessed: 30th March 2022)

31 Windy City Harvest. Windy City Harvest classes. Chicago (n.d.). Available at: https://www.chicagobotanic.org/adult_education/windy_city_harvest. (Accessed: 30th March 2022)

32 Larsen, K. & Gilliland, J. Mapping the evolution of 'food deserts' in a Canadian city: Supermarket accessibility in London, Ontario, 1961–2005. International Journal of Health Geographics 7, 16, doi:10.1186/1476-072X-7-16 (2008).

Submitted: 12/08/2021
Accepted: 03/13/2022

Research Article

Sam Aucoin[1], Alex Briand[1], Béatrice Duval[1], and Zoya Qudsi[1]

[1]Department of Earth and Planetary Sciences, McGill University, Montreal, QC, Canada

Micro-Urban Heat Islands in the City of Montreal

Keywords

Geo-spatial heat distribution, satellite image analysis, urban heat island, micro urban heat island

Email Correspondence

zoya.qudsi@mail.mcgill.ca

Abstract

Heat within a city is not evenly distributed, giving rise to regions of relatively warm and cold temperatures. Regions of very high heat are referred to as micro-urban heat islands (MUHIs) and can be severe enough to harm human health. Despite MUHIs being an important factor in urban health, they are extremely under-researched. In this study we mapped the locations of MUHIs on the island of Montréal and compared them with the locations of vegetation on three clear, sunny days: August 10th, 2021; July 6th, 2020; and June 20th, 2020 using Landsat 8 thermal images with 30 m resolution. We compared two criteria for MUHIs and quantified their composition based on unsupervised classification done on ENVI 5.6.1, and Normalized Difference Vegetation Index (NDVI) calculations. Our results show that MUHIs are mainly associated with the presence of asphalt and concrete, and the absence of dense vegetation. The presence of these materials is not, however, a strong predictor of the formation of MUHIs in themselves. Though variability in unsupervised classifications between images introduces uncertainty in MUHI composition, these results suggest that increasing dense vegetation coverage in Montréal could prevent MUHI development during the summer.

Introduction

The micro-urban heat island (MUHI), or surface heat island, is a microclimatic phenomenon where isolated urban locations have high surface temperatures compared to surrounding areas[1]. There is no standard definition for what constitutes a MUHI; Aniello et al.[1] define a MUHI as any area whose surface temperature is higher than the maximum tree canopy surface temperature. Although MUHIs pose a serious risk to human health[2-4], it is an extremely understudied phenomenon. This contrasts with general urban heat island (UHI) effect which is typically studied at a larger scale, such as an entire city having warmer surface and air temperatures than surrounding rural areas.

MUHIs are one of four main types of urban heat islands. While MUHIs are detected on micro-scales using remotely sensed land surface temperature (LST) data, an UHI may also be defined using air temperature and on local and meso-scales. Surface temperatures are often higher, and air temperatures are more sensitive to vegetation density. The two measures are, however, often closely linked[5]. Canopy and boundary layer heat islands are both defined based on atmospheric temperature differences between urban and rural areas; sub-surface heat islands are defined based on temperature differences between urban and rural subterranean ground. Despite the differences between the four types of UHIs, all arise from differences in the energy budgets between areas[6].

Surface UHIs are associated with materials with a combination of [1] low albedo, which means less solar radiation is reflected and more is absorbed; [2] high heat capacity, which allows more energy to be stored; and [3] low emissivity, which is the the effectiveness of a material in emitting energy, causing the surface temperature to be higher for a given amount of absorbed radiation[7]. Urban areas also have anthropogenic heat input from cars, electricity generation, and industrial processes which may contribute to the UHI effect[8-10]. The impact of anthropogenic heat on MUHIs has not been investigated. Additionally, heat distribution is profoundly affected by latent heat flux, which is the heat that is removed or added to a system via phase change processes like evaporation[11,12]. Vegetated areas lose significant amounts of latent heat via evapotranspiration. In the case of large green areas such as parks, this cooling effect extends outside of the area of the park[13]. For example, Jáuregui (1990)[14] found that the cooling effect of the Chapultepec Park, Mexico extends to a radius of 2 km around the park, which approximately corresponds to the width of the park. At smaller scales, vegetation can have a significant cooling effect through shading[15,16]. Besides the contributions of different materials to the urban landscape, the configuration of these materials, or surface form, is also important. To become surface heat islands, materials with the characteristics listed must also be dry and be oriented to receive direct solar radiation[17]. The 3D structure of cities contributes to UHI formation by increasing the active surface area for energy exchange compared to flatter rural areas. Canyons formed by buildings and roads trap radiation through multiple reflection of incoming solar radiation between canyon walls and reduction of radiation loss from surfaces by shielding them from the cold sky. Canyons also shield urban surfaces from wind, which decreases sensible heat loss[6]. Though remote sensing can provide continuous LST data for an entire city, its images only provide a 2D overhead view. Therefore, it misses the contribution of vertical surfaces and structures which are critical to UHI phenomena[18].

Previous work on MUHIs is mostly confined to the mapping of this phenomenon in different cities in the USA, Greece, and India[1,19,20], and one study investigating their effect on heat-related mortality in Montréal[2]. MUHIs have not yet been mapped in Montréal nor have their mechanisms been studied despite this being an important aspect of the urban environment and health. Based on previous work, we expect that MUHIs in Montréal are located on urban surfaces such as asphalt and concrete, and that vegetation has a negative effect on MUHI formation.

The purpose of defining MUHIs is to understand the spatial distribution of urban heat. Materials within a city are distributed heterogeneously, meaning that some areas, such as parks and gardens, may radiate less heat than other areas such as parking lots or concrete buildings. Though MUHIs are defined based on surface temperature and not atmospheric temperature, which is what is experienced by pedestrians, hot urban surfaces still contribute to atmospheric heat islands and negatively impact human health. Understanding the spatial distribution will allow us to pinpoint urban locations that pose a larger risk to human health on very hot days and to identify which features exacerbate urban heating. Moreover, learning how MUHIs form can help predict future MUHI development and advise mitigation efforts.

Methods

Data Acquisition

We used Landsat 8 Collection 2 Level-2 images. Level-2 products, provided by the United States Geological Survey (USGS), are corrected for easy use and are freely available to the public. Landsat 8 images are collected at 30 meter resolution and at near-nadir angles (±7.5°)[21]. The images provided are already corrected by USGS for sensor degradation and changes, solar elevation, bandwidth, Earth-sun distance and effects of the atmosphere to obtain surface reflectances[22,23]. These were then converted by USGS to land surface brightness temperatures using the Planck function[24]. Brightness temperature represents the temperature that a blackbody would have if it were emitting the same amount of radiation. Therefore, to calculate actual temperature, information on the emissivity of the ground must be known. The Level-2 products combine the Global Emissivity Database (GED), a measure global emissivity calculated through the Temperature–Emissivity Separation (TES) algorithm, and presented at 100-meter[25] resolution with brighness temperature[26]. The measurements for the GED are collected by a sensor on the Terra satellite, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and values are accurate to ±0.015 [27], but can vary more in a city due to urban development and heterogeneity. The result is a land surface temperature product that is ready to use.

From Landsat's collection of measured frequency bands, we used bands 1 to 5 and 10. Band 1 corresponds to ultra blue (443 nm), bands 2 to 4 to blue, green, and red (482 nm, 561.5 nm and 654.5 nm) respectively, band 5 corresponds to near infrared (NIR) (865 nm), and band 10 corresponds to thermal infrared (TIRS) (10895 nm)[28]. Combining bands 2 to 4 results in the visible true color image, and bands 4 and 5 are used to determine the Normalized Difference Vegetation Index (NDVI). We used bands 1-5 to classify surface cover types, and band 10 is used to obtain land surface temperatures.

In order to select comparable images, we used the Sentinel-Hub EO-Browser to visualise thermal images from Landsat 8. Three images corresponding to warm summer days with a clear sky within the last two years were qualitatively selected to have a comparable set of images in terms of meteorological conditions, urban development and overall surface temperature. These three images are from August 10th 2021, July 6th 2020, and June 20th 2020. All images are taken at a similar time of the day: from 10:30 to 11:30 AM EST. We downloaded the images from the USGS Earth Explorer platform.

Material Classification

ENVI Classification We performed a classification on all three selected images using ENVI version 5.6.1 to identify materials. The classifications were unsupervised, and classes were identified in post-processing. We used an Isodata (also known as Iso-cluster) unsupervised classification algorithm, consisting of a calculation of class means distributed across the image followed by an iterative clustering of remaining pixels using minimum distance. At each iteration, the mean is recalculated, followed by a reclassification of pixels according to this new mean[29]. We selected 2 combinations of bands for our classifications: bands 1 to 4 and bands 2 to 5. All images were layer-stacked as a combination of these selected bands and cropped in a quadrilateral over the region of Montréal. We entered a condition of 5 to 7 classes within the algorithm, with a total of 100 iterations for each classification.

Two classifications were performed on the August 10th image with bands 1 to 4 and 2 to 5 combinations. The classification for bands 1 to 4 has a total of 6 material classes, with 2 different classes corresponding to urban materials: one for asphalt and gravel, and a second one for concrete and rooftops. Vegetation for this classification is spread across 3 material classes: grass, grass and urban vegetation, and forest cover. Water is the final class. For bands 2 to 5, a total of 6 material classes are identified: urban materials (concrete and asphalt), concrete and highly reflective roofs, water, short vegetation (grass and urban trees), and forest cover. While some material classes, such as concrete and reflective roofs, appear on both classifications of the different multispectral images, the spatial distribution and accuracy of differentiating materials from one class to another may vary widely, meaning that despite having similar classes, both classifications are different.

Classifications for all images were completed using bands 1 to 4 because they made a clearer distinction between urban material classes. These classifications, however, proved to be variable between different images and conditions; the 2020 images had one more material class than the 2021 image, and varying degrees of importance between the urban materials. We ultimately conducted the following analyses using the material classification from August 10th 2021, as it was the most precise classification performed in terms of defining and isolating urban materials of different nature. Though ENVI was effective in distinguishing urban materials, we did not judge its vegetation classification as satisfactory.

NDVI Classification We used the NDVI as an alternative method to classify vegetation across Montréal. The NDVI is an indicator of vegetation proportion by measuring the difference in near-infrared (NIR) and red (R) values for surface reflectance captured by satellite sensors[30,31], based off the fact that NIR is scattered by mesophyll leaf structure, while red is strongly absorbed by chlorophyll[32]:

$$NDVI = \frac{NIR\ (band\ 5) - R\ (band\ 4)}{NIR\ (band\ 5) + R\ (band\ 4)},$$

where each band's Surface Reflectance pixel values (PV) are scaled as follows[33]:

$$Surface\ Reflectance = 0.0000275PV - 0.2.$$

We applied a scaling factor because the Landsat Collection 2 Level-2 surface reflectance data product is stored in 16-bit integer format with values ranging from 0 to 65535. The scaling factor simply converts the unsigned 16-bit integer to a float value representing surface reflectance in usable units.

NDVI values run from -1.0 to 1.0. Areas of sparse vegetation, which include grass and areas with isolated shrubs or trees, show moderate NDVI values (approximately 0.2 to 0.5) while high NDVI values (approximately 0.6 to 0.9) correspond to dense vegetation such as forests[34].

Image Manipulation

Identifying MUHIs The Landsat 8 thermal images (band 10) we selected were processed and analysed using the Rasterio package in Python. First, the surface temperature rasters were masked using a shapefile defining the island of Montréal[35]. Similar to the surface reflectance data products, the masked rasters were then converted to surface temperature in Kelvin using the following scaling factor[33]:

$$Surface\ Temperature\ (K) = 0.00341802PV + 149.0,$$

where PV is the Surface Temperature value at a given pixel in a raw Landsat 8 Collection 2 Level-2 surface temperature image. After we obtained surface temperatures in Kelvin, two temperature thresholds were applied. The first is defined as "any urban areas radiating higher temperatures than the warmest temperatures associated with tree canopy," as defined by Aniello et al.[1]. Therefore, any area with a surface temperature greater than the highest dense vegetation temperature, found using NDVI, was considered a MUHI. The second threshold is defined as the top 2% of temperature values which was chosen since it roughly corresponds to two standard deviations above the mean, and can therefore be considered significantly higher than the mean temperature. By identifying which pixel temperatures are greater than the threshold, we generated a mask, or binary image, that shows MUHI locations on the map of Montréal.

Comparing MUHIs and Surface Types Similarly, we generated masks showing the locations of asphalt and concrete on the island by identifying pixels whose RGB values match the colour corresponding to that material in the classification. Masks were also generated to locate regions of dense and sparse vegetation. To compare MUHI locations and surface material types, we generated a third map by "overlapping" the MUHI and material masks. A new raster was created from these two masks by assigning each pixel a value corresponding to one of the following: MUHI, material, MUHI and material, or neither.

Results

The Tables 1, 2 and 3 summarize the main parameters obtained for images from August 10th, 2021, July 6th 2020, and June 20th, 2020, respectively, and compare the two temperature thresholds. In each, "Portion of a Material that Contains a MUHI" refers to the proportion of total area of each material on the island that is classified as a MUHI, and similarly, "Portion of MUHIs that Contain Material", refers to the proportion of total MUHI area that is classified as that material. MUHIs are composed of mostly asphalt and concrete: more than 83%, 88% and 86% of MUHIs using either threshold definition are associated with one of these materials on August 10th 2021, July 6th 2020 and June 10th 2020 respectively. However, most asphalt and concrete are not part of MUHIs: on August 10th 2021, only 13.5% of asphalt-covered areas (unvegetated and vegetated) and only 8.83% of concrete-covered areas are associated with MUHIs defined by the 2% threshold, and even less for the canopy threshold. Results are similar for the 2020 images.

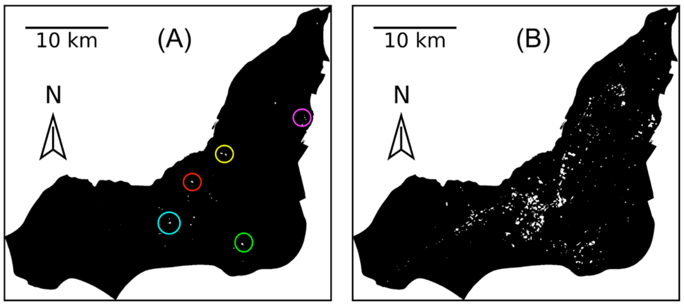


Figure 1. MUHI Distribution on the Island of Montréal on August 10th 2021. (A) Canopy threshold, (B) Top 2% threshold. MUHIs are in white. Notable MUHI locations are circled in (A): Trudeau International Airport (light blue), Bombardier Aerospace Complex (red), STM Centre de transport Legendre (yellow), Canadian Forces Base (pink), and Carrefour Angrignon (green).

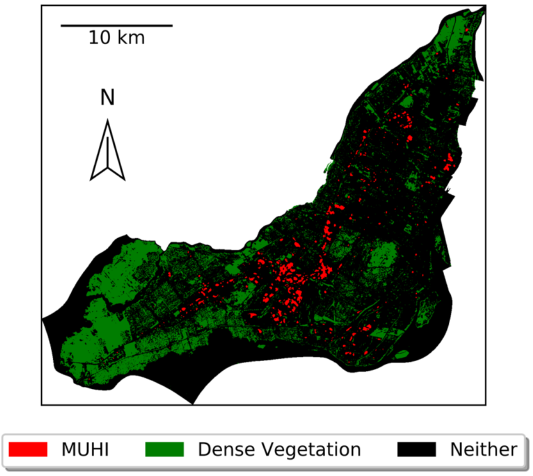


Figure 2. MUHI and Dense Vegetation Cover on the Island of Montréal on August 10th 2021. MUHIs have been identified using the top 2% threshold.

MUHI distributions for the tree canopy and the 2% thresholds were plotted over the island of Montréal for August 10th (see Fig. 1), 2021, July 6th 2020 (results not shown), and June 20th, 2020 (results not shown). MUHIs on the island are mainly located in Dorval and Saint-Laurent. These are where the Montréal Trudeau International Airport, the industrial Technoparc, and many large box stores are located.

To investigate the effects of vegetation on MUHI formation, we examined asphalt-covered areas that overlap with sparse vegetation and compare to asphalt areas with no vegetation. We found that unvegetated asphalt has approximately 5-20 times more MUHI area than sparsely vegetated asphalt (Tables 1, 2 and 3). We also compared MUHI locations and dense vegetation coverage, shown in Fig. 2.

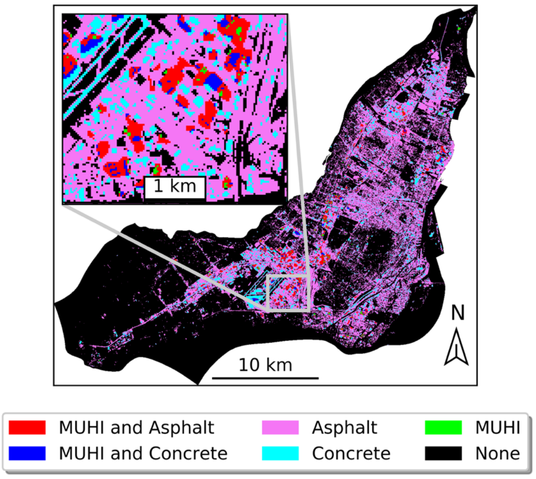


Figure 3. MUHI overlap with asphalt and concrete on August 10th, 2021. MUHIs are identified using the top 2% threshold. The region of the Montréal Trudeau International Airport is zoomed in.

Table 1. MUHI parameters for the two threshold definitions on August 10th, 2021.

August 10, 2021				
	Canopy Threshold		Top 2% Threshold	
Threshold Temperature	323 K		320 K	
Total MUHI Area	0.54270 km^2		12.3237 km^2	
	Portion of a Material That Contains a MUHI	Portion of MUHIs That Contains Material	Portion of a Material That Contains a MUHI	Portion of MUHIs That Contains Material
Concrete	0.74 %	40.43 %	8.83 %	21.13 %
Unvegetated Asphalt	0.40 %	38.94 %	11.21 %	47.14 %
Vegetated Asphalt	0.02 %	3.83 %	2.29 %	21.87 %
Sparse Vegetation	0.03 %	9.82 %	0.55 %	8.34 %
Dense Vegetation	0.00 %	0.00 %	0.02 %	0.18 %

Table 2. MUHI parameters for the two threshold definitions on July 6th, 2020.

July 6, 2020				
	Canopy Threshold		Top 2% Threshold	
Threshold Temperature	324 K		322 K	
Total MUHI Area	3.3246 km^2		12.3534 km^2	
	Portion of a Material That Contains a MUHI	Portion of MUHIs That Contains Material	Portion of a Material That Contains a MUHI	Portion of MUHIs That Contains Material
Concrete	2.91 %	25.85 %	8.41 %	20.07 %
Unvegetated Asphalt	2.51 %	56.06 %	9.62 %	57.77 %
Vegetated Asphalt	0.23 %	6.74 %	1.75 %	13.50 %
Sparse Vegetation	0.10 %	5.14 %	0.38 %	5.41 %
Dense Vegetation	0.00 %	0.00 %	0.01 %	0.09 %

Table 3. MUHI parameters for the two threshold definitions on June 20th, 2020.

	June 20, 2020			
	Canopy Threshold		Top 2% Threshold	
Threshold Temperature	329 K		326 K	
Total MUHI Area	$1.1079\ km^2$		$12.3543\ km^2$	
	Portion of a Material That Contains a MUHI	Portion of MUHIs That Contains Material	Portion of a Material That Contains a MUHI	Portion of MUHIs That Contains Material
Concrete	1.32 %	35.09 %	7.79 %	18.59 %
Unvegetated Asphalt	0.67 %	46.95 %	8.99 %	56.42%
Vegetated Asphalt	0.05 %	4.55 %	1.76 %	13.20 %
Sparse Vegetation	0.03 %	5.85 %	0.55 %	8.39 %
Dense Vegetation	0.00%	0.00 %	0.02 %	0.14 %

Discussion

MUHIs and Surface Types

Urban Materials Areas with high MUHI density are sparsely vegetated and are dominated by asphalt and concrete surface coverage - all of these factors are known to enhance urban heating[36]. However, the presence of asphalt or concrete in itself does not necessarily lead to MUHI formation. In Montréal, most asphalt and concrete areas appear to lack other qualities that would otherwise make them MUHIs, namely receiving direct solar radiation. Because the measured quantity is surface temperature, a full picture of the effect of MUHIs is obscured since air temperature is more akin to what one would "feel". Because of this, the intensity of MUHIs may be overestimated, and the effect of vegetation density maybe be underrepresented. In the absence of high-resolution air temperature measurements, remote sensing is the only way to study this phenomenon, despite it introducing a bias toward warmer temperatures. The images being taken early in the day also affects MUHI composition; urban materials with high heat capacities may be underrepresented as they did not have time to heat up, while materials with low heat capacities are overrepresented. Despite most asphalt and concrete not being part of MUHIs, MUHI composition is still dominated by these materials. This indicates that the surface material type influences MUHI generation, but is not the only factor.

Vegetation Our observation that unvegetated asphalt has several times more MUHI coverage than vegetated asphalt may be partly due to thermal anisotropy in cities, e.g. the satellite only detected the surface temperature of a tree and not the road beneath. Nonetheless, it is well-documented that vegetation can cool urban spaces through evapotranspiration[37-39]. The stark difference in MUHI coverage between vegetated and unvegetated asphalt suggests that vegetation plays some role in preventing MUHI development in Montréal. However, we found that sparse vegetation alone is not always sufficient to prevent MUHI formation; on August 10th, 2021, 9.82% of MUHIs as defined by the canopy threshold contained sparse vegetation.

Sparse vegetation includes grass, which dries out much faster during periods of low precipitation and becomes hotter. Soil moisture generally increases with depth[40], so trees, which have deeper roots, have better access to moisture during dry spells and can continue transpiring to maintain cooler surface temperatures. Surface soil moisture varies diurnally and seasonally: high soil moisture is associated with cold near-surface air temperature and low near-surface wind speed, while dry soil is associated with warm temperatures and high wind speeds[41]. In Montréal, the monthly average precipitation in August 2021 was one of the lowest of that year (36.1 mm), while June and July 2020 had slightly more precipitation (46.4 and 86.8 mm, respectively)[42]. Low soil moisture may be responsible for the greater contribution of sparse vegetation to MUHIs on August 10th, 2021 compared to the other two dates.

In our selected images, densely vegetated surfaces remained cooler overall. In addition to dense vegetation reducing heating on the surface where it is present, it may also reduce heating around the boundary of vegetation. In Fig. 2, we see that there is a MUHI-free buffer zone between the hottest areas (i.e., top 2% of surface temperatures) on the island and large areas of dense vegetation like Mount Royal Park and Morgan Arboretum. This buffering phenomenon has previously been observed for air temperatures[13].

MUHI Locations As shown in Fig. 1, MUHIs as defined by the canopy threshold are sparsely distributed compared to the percent threshold - the former method is much more selective, as the canopy threshold temperature is much higher than the 2% threshold temperature (see Table 1). The highly selective canopy method allows the identification of specific locations and structures that are associated with the absolute hottest MUHIs on the island. Notable MUHI locations include the Montréal Trudeau International Airport (Dorval), the Canadian Forces Base of Montreal, the Bombardier Aerospace complex (Saint-Laurent), the STM Centre de transport Legendre, and the Carrefour Angrignon (LaSalle). In Fig. 3 we can see that these areas are not fully covered by MUHIs, but MUHIs are located towards the center of these areas. This is consistent with the observation in previous work that MUHIs are hottest at the center[1]. We consistently observe that urban locations surrounded by large, uninterrupted areas of urban surface cover, rather than vegetation, are most susceptible to MUHI formation.

Limitations

Material Classification We found inconsistencies between the classifications with bands 1 to 4 of our 3 images, which introduces uncertainty in the MUHI compositions presented in Tables 1, 2 and 3. The unsupervised classification algorithm from ENVI version 5.6.1 identified 7 classes for both images from 2020, but identified 6 classes for August 10th 2021. Since the classification was unsupervised, clusters were identified automatically, which may introduce variability if the images were taken under different conditions. Vegetation cover and biomass, as well as vegetation species distribution can vary slightly in relatively short periods of time, which may introduce variability between classifications of images between 2020 to 2021. Knowing this, it is possible that the unsupervised classification algorithm generated different clusters between images from different periods, resulting in fewer material classes for the image in August 10th[43]. Moreover, the quadrilateral region of the images under which classifications were performed includes several fields and agricultural complexes. This likely introduced additional variability within the clustering done during the unsupervised classification. It is also important to note that the choice of bands introduced flaws within the unsupervised classification. The ultra-blue band (443 nm) easily penetrates water and vegetation, which often results in the classification of some vegetated areas as water and vice-versa[44]. However, classification from the combination of bands 1 to 4 proved to be more reliable to differentiate between urban materials than the combination of bands 2 to 5. It is possible that under different conditions, another combination of bands could have shown better results.

Thermal Anisotropy Another limitation is the effective thermal anisotropy, or uneven surface temperatures, produced by Montréal's 3D geometry and uneven surface heating. This is particularly limiting in densely built areas like Montréal's downtown core. Satellites only capture a 2D view of surface temperature, so some features may not be "seen" by the satellite. For example, a small park may be obstructed by a tall building; the satellite may only see the surface temperature of the building and mischaracterize the area as hotter than it actually is. Though Landsat 8 viewing angles are near-nadir, surface temperatures can still significantly vary between images collected at these slightly different angles. Vertical surfaces are almost entirely absent from the imagery, so we could not assess their contribution to MUHIs in Montréal.

Conclusion

In this study, we identified the location and general causes of two definitions of micro-urban heat islands on the island of Montréal using 3 thermal images from the Landsat 8 Earth-observing satellite. Classifications done during our analysis identified several urban materials on the island of Montréal, but there were important factors that introduced variability within the unsupervised classification algorithm, thus affecting clustering and the number of classes between images. Using a larger set of images spread over a longer period of time could reduce variability. In addition, using a mask to isolate the island before classification could reduce the influence of agricultural activity in surrounding rural areas on the unsupervised classification clustering. Performing a supervised classification, while being more time consuming, could generate a much more reliable set of classified images. Areas of future work include analysing the seasonality and evolution of MUHIs using Landsat images over several seasons and years, and correcting for anisotropy by estimating Montréal's complete urban surface. This would include surfaces normally absent from satellite images, and thus provide a more accurate classification of materials and MUHIs on the island[6].

In conclusion, we find that micro-urban heat islands on the island of Montréal are mainly associated with urban materials, but the presence of urban materials in itself is not a strong predictor for MUHI formation. We also find that almost no MUHIs (defined by the top 2% threshold) exist very close to dense vegetation; these zones act as a buffer to MUHI generation. In contrast to our hypothesis, we find that a significant proportion of MUHIs are associated with sparse vegetation, implying that not all forms of vegetation are equally efficient at reducing MUHI formation. These findings highlight the importance of dense green cover within Montréal to reduce the intensity of urban heating, and can be used to focus heat-related health mitigation efforts on the most vulnerable parts of the city, as well as inform building practices to reduce the generation and intensity of MUHIs.

Acknowledgements

We thank Dr. Eric Galbraith, Dr. Jeff McKenzie, and Tanya Matitia for their support and comments throughout this project.

References

1. Aniello, C., Morgan, K., Busbey, A. & Newland, L. Mapping micro-urban heat islands using LANDSAT TM and a GIS. Computers & Geosciences 21. Environmental Geology, 965–969. iSSN: 0098-3004. https://www.sciencedirect.com/science/article/pii/0098300495000335 (1995).

2. Smargiassi, A. et al. Variation of daily warm season mortality as a function of micro-urban heat islands. Journal of Epidemiology and Community Health 63, 659–664 (2009).

3. Kovats, R. S. & Hajat, S. Heat Stress and Public Health: A Critical Review. Annual Review of Public Health 29, 41–55 (2007).

4. Hajat, S., O'Conner, M. & Kosatsky, T. Health effects of hot weather: from awareness of risk factors to effective health protection. The Lancet 375, 856–863 (2010).

5. Kawashima, S., Ishida, T., Minomura, M. & Miwa, T. Relations between surface temperature and air temperature on a local scale during winter nights. Journal of Applied Meteorology and Climatology 39, 1570–1579 (2000).

6. Oke, T. R., Mills, G., Christen, A. & Voogt, J. A. Urban climates (Cambridge University Press, 2017).

7. Goward, S. N. Thermal Behavior of Urban Landscapes and the Urban Heat Island. Physical Geography 2, 19–33. https://doi.org/10.1080/02723646.1981.10642202 (1981).

8. Haddad, L. & Aouachria, Z. Impact of the transport on the urban heat island. International Journal of Environmental and Ecological Engineering 9, 968–973 (2015).

9. Ryu, Y.-H. & Baik, J.-J. Quantitative analysis of factors contributing to urban heat island intensity. Journal of Applied Meteorology and Climatology 51, 842–854 (2012).

10. Phelan, P. E. et al. Urban heat island: mechanisms, implications, and possible remedies. Annual Review of Environment and Resources 40, 285–307 (2015).

11. Holmer, B. & Eliasson, I. Urban-rural vapour pressure differences and their role in the development of urban heat islands. International Journal of Climatology: A Journal of the Royal Meteorological Society 19, 989–1009 (1999).

12. Liu, K., Li, X., Wang, S. & Li, Y. Investigating the impacts of driving factors on urban heat islands in southern China from 2003 to 2015. Journal of Cleaner Production 254, 120141 (2020).

13. Shashua-Bar, L. & Hoffman, M. E. Vegetation as a climatic component in the design of an urban street: An empirical model for predicting the cooling effect of urban green areas with trees. Energy and buildings 31, 221–235 (2000).

14. Jáuregui, E. Influence of a large urban park on temperature and convective precipitation in a tropical city. Energy and buildings 15, 457–463 (1990).

15. Sharlin, N. & Hoffman, M. The urban complex as a factor in the air-temperature pattern in a Mediterranean coastal region. Energy Build.;(Switzerland) 7 (1984).

16. Oke, T. R. The micrometeorology of the urban forest. Philosophical Transactions of the Royal Society of London. B, Biological Sciences 324, 335–349 (1989).

17. Oke, T. R. The energetic basis of the urban heat island. Quarterly Journal of the Royal Meteorological Society 108, 1–24 (1982).

18. Stewart, I. D. & Mills, G. The Urban Heat Island (Elsevier, 2021).

19. Stathopoulou, M., Cartalis, C. & Keramitsoglou, I. Mapping micro-urban heat islands using NOAA/AVHRR images and CORINE Land Cover: an application to coastal cities of Greece. International Journal of Remote Sensing 25, 2301–2316. https://doi.org/10.1080/01431160310001618725 (2004).

20. Amirtham, L. R., Devadas, M. D. & Perumal, M. Mapping of Micro-Urban Heat Islands and Land Cover Changes: A Case in Chennai City, India. The International Journal of Climate Change: Impacts and Responses 1, 71–84 (2009).

21. Li, J., Roy, D. P. & Zhang, H. Comparison of Sentinel-2A and Landsat-8 Nadir BRDF Adjusted Reflectance (NBAR) over Southern Africa in AGU Fall Meeting Abstracts 2016 (2016), B31B–0472.

22. Thenkabail, P. S. Satellite Sensor Data Normalization Issues: A User Perspective http://www.pancroma.com/downloads/General%20Landsat.pdf. Oct. 2009.

23. United States Geological Survey. Landsat Collection 2 Level-2 Science Products https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-level-2-science-products (2021).

24. Price, J. C. Estimating surface temperatures from satellite thermal infrared data—A simple formulation
for the atmospheric effect. Remote sensing of environment 13, 353–361 (1983).

25. Hulley, G. C. et al. The ASTER Global Emissivity Dataset (ASTER

GED): Mapping Earth's emissivity at 100 meter spatial scale. Geophysical Research Letters 42, 7966–7976 (2015).

26. United States Geological Survey. Landsat Collections https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-atmospheric-auxiliary-data?qt-science_support_page_related_con=1#qt-science_support_page_related_con.

27. Sabol Jr, D. E., Gillespie, A. R., Abbott, E. & Yamada, G. Field validation of the ASTER temperature–emissivity separation algorithm. Remote Sensing of Environment 113, 2328–2344 (2009).

28. Sentinel-Hub. About Landsat 8 OLI-TIRS Collection 2 Level 2 Data https://docs.sentinel- hub.com/api/latest/data/landsat-8-l2/ (2021).

29. Environmental Systems Research Institute. Fonctionnement de l'outil Iso cluster https://desktop.arcgis.com/fr/arcmap/10.3/tools/spatial-analyst-toolbox/how-iso-cluster-works.htm.

30. Malik, M. S., Shukla, J. P. & Mishra, S. Relationship of LST, NDBI and NDVI using landsat-8 data in Kandai-himmat watershed, Hoshangabad, India (2019).

31. United States Geological Survey. Landsat Normalized Difference Vegetation Index https: / /www.usgs.gov/core-science-systems/nli/landsat/landsat-normalized-difference-vegetation-index?qt-science_support_page_related_con=0#qt-science_support_page_related_con (2021).

32. Pettorelli, N. et al. Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in ecology & evolution 20, 503–510 (2005).

33. United States Geological Survey. Landsat Collection 2 Level-2 Science Products https://www.usgs.gov/core-science-systems/nli/landsat/landsat-collection-2-level-2-science-products (2021).

34. United States Geological Survey. NDVI, the Foundation for Remote Sensing Phenology https://www.usgs.gov/core-science-systems/eros/phenology/science/ndvi-foundation-remote-sensing-phenology?qt-science_center_objects=0#qt-science_center_objects (2021).

35. Ville de Montréal. Limite administrative de l'agglomération de Montréal (Arrondissements et Villes liées) 2020. https://donnees.montreal.ca/ville-de-montreal/polygones-arrondissements (2021).

36. Kim, H. H. Urban heat island. International Journal of Remote Sensing 13, 2319–2336 (1992).

37. Qiu, G.-y. et al. Effects of evapotranspiration on mitigation of urban temperature by vegetation and urban agriculture. Journal of Integrative Agriculture 12, 1307–1315 (2013).

38. Qiu, G. Y. et al. Experimental studies on the effects of green space and evapotranspiration on urban heat island in a subtropical megacity in China. Habitat international 68, 30–42 (2017).

39. Small, G., Jimenez, I., Salzl, M. & Shrestha, P. Urban heat island mitigation due to enhanced evapotranspiration in an urban garden in Saint Paul, Minnesota, USA. WIT Transactions on Ecology and the Environment 243, 39–45 (2020).

40. Shi, B., Tang, C.-S., Gao, L., Liu, C. & Wang, B.-J. Observation and analysis of the urban heat island effect on soil in Nanjing, China. Environmental Earth Sciences 67, 215–229 (2012).

41. Husain, S. Z., Bélair, S. & Leroyer, S. Influence of soil moisture on urban microclimate and surface-layer meteorology in Oklahoma City. Journal of Applied Meteorology and Climatology 53, 83–98 (2014).

42. Total precipitation - monthly data for Montréal https://montreal.weatherstats.ca/charts/precipitation-monthly.html.

43. Tigges, J., Lakes, T. & Hostert, P. Urban vegetation classification: Benefits of multitemporal RapidEye satellite data. Remote Sensing of environment 136, 66–75 (2013).

44. Acharya, T. D. & Yang, I. Exploring Landsat 8. International Journal of IT, Engineering and Applied Sciences Research (IJIEASR) 4, 4–10 (2015).

Submitted: 01/21/2022
Accepted: 04/03/2022

Research Article

Liesl Van Wyk[1]

[1]Department of Bioresource Engineering, McGill University, Montreal, QC, Canada

Potential for use of Spent Substrate of Pleurotus Mushrooms Grown on Urban Waste as Feed for Dairy Cattle

Keywords

Pleurotus, spent mushroom substrate (SMS), cattle feed, spent coffee grounds (SCG), urban waste, circular economy, urban waste

Email Correspondence

liesl.vanwyk@mail.mcgill.ca

Abstract

Mushroom wastes are available in high volumes, with 5 million tons of spent mushroom substrate (SMS) being disposed of globally every year. Due to this high availability, various forms of SMS have been researched for their use as alternative animal feeds. Additionally, experimental techniques can be used to grow certain mushroom species, such as oyster mushrooms (*Pleurotus* sp.) on various lignocellulosic waste materials. Therefore, the SMS from *Pleurotus* sp. grown on these waste materials may offer a promising conversion from a waste material to a low-cost, nutritionally sufficient feed. However, little research has been done to determine if feeds from *Pleurotus* SMS specifically grown on urban waste substrates offer the same benefits. Given rising awareness on circularity and urban self-sufficiency, growing mushrooms on urban waste is a promising solution which should be investigated. This paper assesses the feasibility of using SMS from golden oyster mushrooms (*Pleurotus citrinopileatus*) grown on urban waste as dairy cattle feed, comparing substrate ratios to determine which would result in the most desirable protein and fiber contents. SMS from three experimental substrates of cardboard and spent coffee grounds (SCG) were compared to traditional dairy cattle feeds. Treatments 2 and 3 were found to be suitable for use as additives to traditional feeds in small replacement amounts. However, both treatments also had high fiber content, which may affect practicality of use as feeds.

Introduction

The cultivated mushroom industry is growing rapidly, with global production of edible mushrooms having increased 30% since 1978[1]. In Canada, the mushroom market is dominated by *Agaricus* sp. (including button mushrooms – *Agaricus bisporus*), which account for 98% of production; other "specialty" mushrooms such as *Pleurotus* sp. account for the remaining 2%[2]. Valued at 63 billion USD in 2013, the global mushroom market is likely to continue growing, given the rising demand for non-animal proteins[3]. The rise in mushroom consumption is promising from an environmental perspective due to the intense resource use associated with animal protein; however, environmental issues also arise in the form of spent mushroom substrate (SMS), the material left behind after mushroom fruiting bodies have been harvested[4,5,6].

The large quantity of SMS left over, approximately 5 kg for every 1 kg of harvested mushroom in *A. bisporus* production, is seen by mushroom producers as waste, resulting in an astonishing 5 million tons of SMS solid waste being disposed of annually[7,8]. However, this SMS has several documented alternative uses, primarily being used as fertilizer. Despite its effectiveness as a fertilizer, the storage and transportation costs associated with disposal of SMS by field application incurs such high costs that it can be less economically viable than chemical fertilizers[6]. Additionally, with growing awareness on the benefits of circularity, SMS uses which can replace raw inputs by being returned into a cycle, such as animal feed, are more desirable[9].

Typical agricultural products used for cattle feed contain high amounts of nutrients, but are difficult to digest and are therefore inefficient in their conversion of a raw agricultural product to usable energy[7]. There are also issues with importation of more nutritionally valuable feeds; the European Union is aiming to reduce its high import dependency (70%) on soy-based, protein-rich animal feed[9]. Therefore, alternatives are needed for local production of high quality, protein-rich animal feeds which are high in nutrients, easy to digest, and economically viable. In this paper we investigated the possibility of using SMS from mushrooms grown on locally generated urban waste—cardboard and spent coffee grounds (SCG)—as dairy cattle feed.

Literature Review

Traditional ruminant feeds include straw and other agricultural residues; however, these feeds have low available energy, protein, and mineral content because digestion is impeded by high quantities of hard-to-digest cell wall components such as cellulose, hemicellulose, and lignin[10]. Delignification of straw through chemical treatments can increase nutritional value; however, these processes are both economically and environmentally undesirable[11].

A good alternative is delignification through biological processing of raw materials; some fungi are very efficient decomposers of these cell wall components, especially of lignin[12], and therefore can be used for biological delignification. An added benefit of biological delignification using mushrooms is the production of a valuable food source for humans (harvested mushrooms).

A preliminary report by Weiss et al. (1980) discussed the initial results of their ruminant feed study, which incorporated *A. bisporus* mushroom waste in the form of SMS and mushroom stumps[13]. In this study, mushroom wastes ensiled with hay and corn showed increases in crude protein (CP), calcium, and acid detergent fiber (ADF). However, CP content decreased in treatments without corn, implying that the increased protein content could be more attributable to corn than mushrooms. Nonetheless, the analysis showed that mushroom supplemented diets could meet nutrient requirements for a wide range of ruminants, subject to confirmation with a metabolism study. One core issue found by the authors is the low dry matter (DM) content of the *A. Bisporus* SMS, which makes transport of this SMS unnecessarily costly due to high moisture content.

The nutritional consistency of SMS is another issue to consider, as a standard diet must be maintained for cattle; unfortunately, the authors found the primary hurdle in feed development to be the inconsistency of *A. bisporus* mushroom waste-based livestock feed, which makes it difficult to formulate a standard diet. However, this statement is not consistent with general knowledge as *A. bisporus* cultivation is highly standardized and has quite low variability compared to other mushroom species which are less commonly cultivated[14]. Additionally, using the results of metabolism studies, standard diet formulations can be developed which help both the farmer (by lowering feed costs) and the mushroom producers (by aiding in waste disposal).

A paper by Wilson et al.[15] discusses the results of a lamb metabolism trial using ensiled hay, corn, and *A. bisporus* mushroom waste feed. This study tested three diets containing 10% hay, 15% corn, and 75% *A. bisporus* mushroom waste, the last component being varied between trials with either all compost, all stumps, or a half and half mix of the two. The results of the feeding trial showed that lambs experienced a reduced rate of growth when consuming feed with mushroom wastes, compared to a standard diet. The low energy value of SMS made it ineffective in meeting the nutritional demands of young animals, however the authors noted that SMS could be incorporated at 25-33% in diets of mature animals, who have lower nutritional requirements, and could be included at levels less than 15% in the diets of growing animals. The results of this study may discourage the search for a suitable, mushroom-based livestock feed, but expanding beyond *A. bisporus* waste to other mushroom species may provide different results.

Pleurotus sp., commonly known as oyster mushrooms, are the second most cultivated mushroom worldwide[16], accounting for 27% of global mushroom production[17]. Oyster mushrooms are well known because they are easy to grow, highly nutritious, and can be grown on a wide variety of agricultural wastes although with varying yield rates[7]. *Pleurotus* sp. are high in protein (15-35% on a dry weight basis) and vitamins B and C, and can be productively grown on a huge variety of lignocellulosic compounds, including industry waste products such as pulp sludge, coffee residues, agave waste, and soy pulp[12].

A 1998 feeding trial by Adamović et al.[18] studied the use of SMS from *P. ostreatus* grown on wheat straw as a cattle feed. They found that cell-wall components of the straw, especially lignin and cellulose, decreased during incubation due to degradation by *P. ostreatus* enzymes, corresponding with an increase in protein content and digestibility of the substrates. However, despite a theoretical improvement in feed quality, the feeding trial showed that average daily gains were smaller in both groups consuming *P. ostreatus* SMS compared to a control group eating their regular feed. This result can be attributed to low palatability of the SMS feed since during the trial, the cattle rejected SMS unless it was mixed with silage, refusing to consume anything with more than 17% SMS as a portion of total feed dry matter (lowered from the original trial goal of 20%). The group consuming 10% SMS had only 10 g less gain than the control group, compared to 60 g less for the group consuming 17% SMS. It is difficult to determine how much of this reduction is due to reduced feed intake, and how much is due to the quality of the feed itself; if some solution could increase the willingness of cows to eat the SMS feed, P. *ostreatus* SMS could be a valuable feed additive to increase protein intake for livestock.

While the use of agricultural wastes such as *Pleurotus* sp. substrate has been extensively studied, there is far less published academic information regarding the use of urban wastes as substrates. However, one student research paper conducted at McGill University studied the feasibility of growing oyster mushrooms on SCG and either cardboard or coffee filter paper, finding that using SCG as a substrate resulted in satisfactory fruiting results, and also reduced both energy costs and urban generated waste compared to typical commercial substrates[19]. One issue found in this research was the increased risk of contamination by what was referred to as "green mold" when using SCG and cardboard substrates, compared to SCG and filter paper.

The author speculated that this may be a result of introduced contaminants from the cardboard, as the filter paper is covered until use as a substrate. Green mold does not refer to a specific species, so it is unclear what organism the author is referring to. However, common contaminants of *P. ostreatus* mushrooms include competitors such as *Pseudomonas*, *Bacilli*, and coliform bacteria, and undesirable fungi such as *Trichoderma, Penicillium*, and *Aspergillus*, all of which have inhibited growth in more alkaline substrates[20]. Therefore, the acidity of SCG may also have raised the contamination risk through lowering of the substrate pH to the point of increased risk of contamination.

Use of SCG in substrates may increase contamination risk, but it also increases protein content which helps to increase yields; therefore, a balance of SCG content must be found[12]. In a previous study conducted by the authors, this balance was tested by growing grey oyster mushrooms (*Pleurotus ostreatus var. columbinus*) on five substrates with coffee contents of 0%, 25%, 50%, 75%, and 100% by wet weight, with cardboard composing the rest of the substrate. The treatments with 25% and 50% coffee were the best performing, with treatments having higher than 50% coffee failing, and the 0% coffee treatment performing poorly.

The lack of further academic research on *Pleurotus* sp. cultivation on urban wastes is indicative of a gap between academic and general knowledge. Information is widely available online regarding the efficacy of growing oyster mushrooms on cardboard and SCG, as both products are widely available in urban settings; however, little academic research has been done to support these claims. Given growing awareness on the need for increased urban waste redirection through the circular economy, in which resources are recovered and reapplied in different cycles[21], this research gap should be rectified. This paper will contribute by growing oyster mushrooms on urban waste substrates (cardboard and coffee) and assessing the suitability of the resulting SMS for use as cattle feed.

Materials and Methodology

Materials

All substrates used were diverted from the waste streams of local Montreal businesses. Cardboard was collected from a recycling bin behind a grocery store, with only cardboard that was clean and without visible glue or ink being selected. Coffee was collected from Café Névé with help from employees, who placed the SCGs in a closed container after brewing for collection.

Mushroom spawn was purchased in 1 kg quantity from Mycoboutique (Montreal, QC). The strain used was *Pleurotus citrinopileatus* (Yellow Oyster, or Golden Oyster).

A shotgun fruiting chamber (SGFC) was constructed following instructions from FreshCap Mushrooms[22], shown in Figure 1. Once constructed, the SGFC was propped up on cups to ensure it was high enough off the ground for proper airflow to be established, according to recommendations on an online SGFC forum.

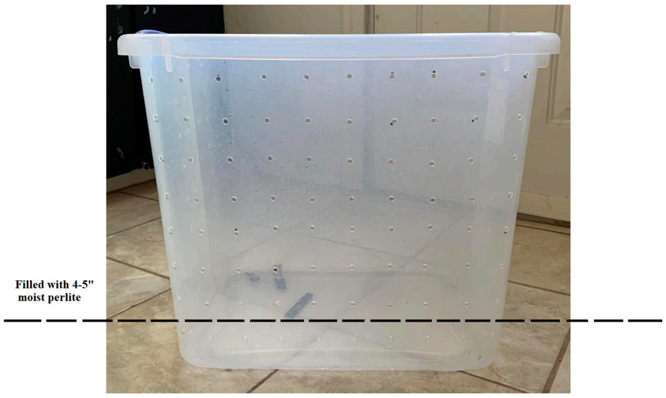


Figure 1. SGFC chamber pictured before being filled with moist perlite and propped up.

The mushrooms were grown in #4T polypropylene bags with 0.2 micron filter patches (Mycoboutique, Montreal, QC). Sterilization of tools and surfaces was done with 70% isopropyl alcohol. A generic kitchen scale with ±1 g accuracy was used to weigh the substrates and spawn.

Methodology

The main goal of this study was to find which substrate ratio of cardboard to SCG produced the best SMS for dairy cattle feed after oyster mushroom cultivation. The protein and fiber content of the SMS were used as proxies for nutritional quality and digestibility, with higher protein and lower fiber being desirable. I hypothesize that:

1. Higher SCG content in substrate will result in lower fiber content.
2. Higher SCG content in substrate will result in higher protein content.

In order to test these hypotheses, substrate samples will be taken both before and after mushroom harvest (section 4). These samples will then be tested to allow for full comparison to recommended dairy cattle diets (section 5), and statistical analysis will be performed, using these results, to test the hypotheses (section 6). Based on the results of previous research, three substrate treatments were devised, all with a 20% spawn rate, equal total substrate weight, and lower than 50% coffee content. Three replicates were prepared for each treatment. The substrate ratios for each of the three treatments is displayed in table 1 below.

Table 1. Summary of substrate composition for each treatment.

Treatment	Cardboard (g)	Coffee (g)	Spawn (g)
1	615	0	123
2	492	123	123
3	369	246	123

1. Preparation

Cardboard was cleaned using hot water pasteurization[23]. Boxes were cut into large pieces, placed in a large sturdy plastic storage container, and soaked in boiling water for two hours. Then, pieces were drained, stacked, and covered. Collected SCG were pasteurized in the brewing process and used within 24 hours of collection; therefore, no sterilization was performed. After substrate preparation and material collection, all surfaces were sterilized using 70% isopropyl alcohol. After this had completely evaporated, inoculation began.

2. Inoculation Procedure

Three bags were prepared for each treatment, for a total of nine inoculated bags. Cardboard was torn into small pieces, then layered in the grow bags with coffee and/or spawn[23]. The bags were then sealed with zip ties, placed in a dark room out of direct light, and left to colonize. After 21 days, the bags were fully colonized, and fruiting was initiated.

3. Fruiting Procedure

Bags were cut open to sample substrate (section 4), then firmly sealed with tape. About halfway down the front of the bag, a 1 in. incision was made to allow fruiting[12]. The bags were then placed in the SGFC and misted 3-6 times a day. Fruiting time varied greatly for each bag; the first fruits were harvested 12 days after being placed in the SGFC, compared to 38 days for the last fruits.

4. Sampling

Pre-fruiting sampling was very conservative, as too much disturbance of the substrate could increase contamination risk[12]. Six samples were taken from various locations in each bag, for a total sampled mass of approximately 10 g per bag. After samples were taken, they were placed in a Ziploc bag, labelled, and frozen.

The second round of samples were taken after one flush of mushrooms had been harvested. This post-fruiting sampling was done by cutting the bag open and mixing up its contents in a large bowl. Then 200 g samples were taken in small, random increments from the bowl, placed into a Ziploc bag, labelled, and frozen.

5. Testing

Samples were sent to Agrianalyse (Sherbrooke, Quebec) for analysis. In order to compare the nutritional content of SMS feed to traditional feeds, the samples were tested for crude protein, moisture, neutral-detergent fibre (NDF), and mineral content. NDF content was chosen as it is a measure of cell wall content and, therefore, can be used to determine fiber content and digestibility[24], as well as being a predictor of voluntary intake of feed[25].

6. Statistical Analysis

Results were imported to Excel and then checked for normality using the Shapiro-Wilke test[26]. One-way ANOVA tests were performed on normal data sets to assess statistical significance between the means of the three treatment groups[27]. An independent two sample t-test was also performed to determine if there was a statistically significant difference between two treatment means[28].

Results

Change in NDF content as a result of mushroom digestion was calculated from the pre- and post-fruiting values, shown below in Figure 2. No statistically significant difference in mean change in NDF was found when comparing treatments using one-way ANOVA. Samples of post-fruiting substrates were much larger than pre-fruiting samples, which may have influenced the comparison of pre- and post-fruiting results.

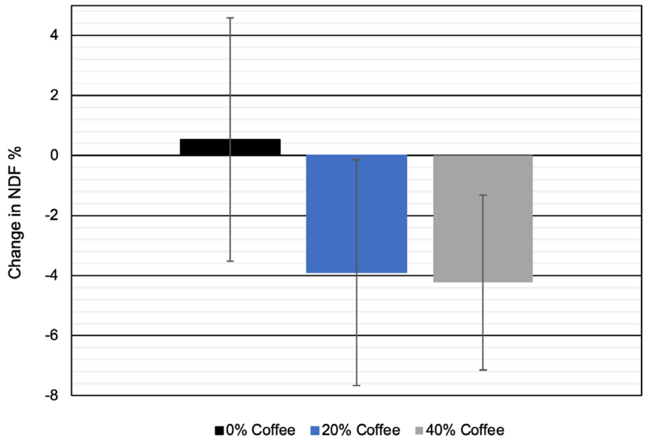


Figure 2. Change in fiber content in SMS as a result of mushroom digestion.

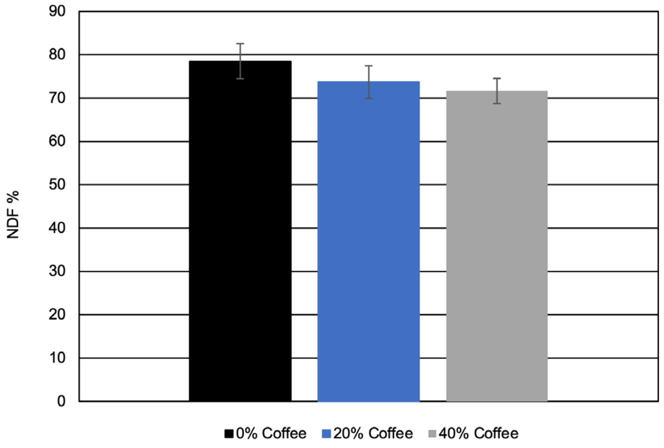


Figure 3. Final (post-fruiting) fiber content of SMS.

The difference in final fiber content between treatments was assessed using both a one-way ANOVA test on all treatments and two-sample t-tests between treatments. High variability in fiber content resulted in a lack of statistical significance from all tests; therefore, the null hypothesis of no difference in mean NDF content between treatments was accepted. Final NDF content for all treatments is shown below in Figure 3.

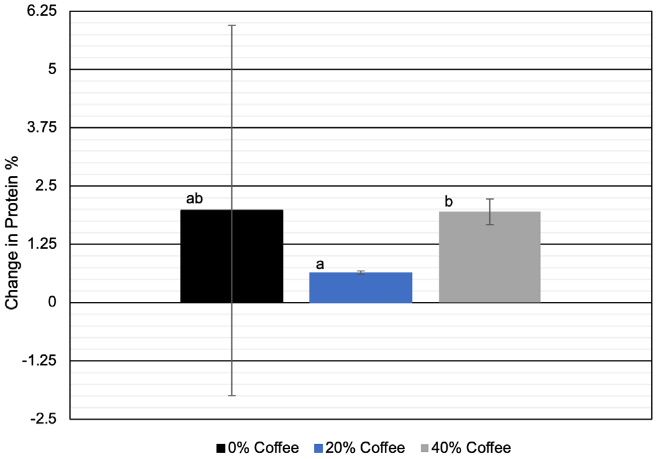


Figure 4. Change in protein content in SMS as a result of mushroom digestion.

Using the pre- and post-fruiting protein values, the change in protein as a result of mushroom digestion was plotted. No statistically significant difference in mean change in protein was found when comparing all treatments using one-way ANOVA; however, the two-sample t-test found a significant difference between treatments 2 and 3 (20% & 40% coffee, respectively), with protein content in treatment 3 increasing by 1.31% more than in treatment 2 (Figure. 4). Samples of post-fruiting substrates were much larger than pre-fruiting samples, which may have influenced the comparison of pre- and post-fruiting results.

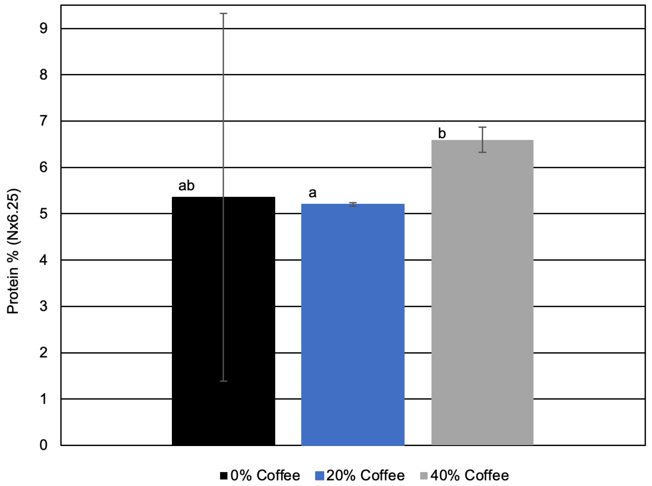


Figure 5. Final (post-fruiting) protein content of SMS.

The difference in final protein content between treatments was plotted, with the one-way ANOVA test on all treatments showing no significant difference between group means, where high variability in treatment 1 was likely a factor. If this variability had resulted from lack of sterile lab space or lack of environmental control, both treatments 2 and 3 would presumably have had comparably high variabilities. However, as both had comparably low variabilities, a more likely cause was that treatment 1 was the only one with 0% SCG, which would affect both the pH and the substrate structure, two major determinants of substrate performance[12]. It is also possible that a random testing error may have contributed to high variability. A two-sample t-test revealed a significant difference between the mean protein content of treatment 2 (5.20% ± 0.0346) and treatment 3 (6.60% ± 0.274), supporting hypothesis 2 (higher SCG content correlates to higher protein content) for these two samples. The results of this analysis are plotted in Figure 5.

Discussion

It is important to note that the sample size of this study was relatively small due to the lack of lab space and high cost of testing. Therefore, the results discussed here should be interpreted with caution before they are confirmed with further studies.

Treatment 1 had the highest variability in protein content, but a slightly higher mean protein content than treatment 2, making it difficult to assess its potential suitability as a feed. The source of variability in measured protein content for treatment 1 was discussed briefly in the previous section but must be further investigated before drawing any definite conclusions regarding the feasibility of using treatment 1 as a feed. Additionally, since this research is focused on the use of urban wastes as cattle feed, and since SCG represent a large, pre-sterilized urban waste feedstock, the treatments containing SCG will be the focus of analysis here. Treatment 3 had the highest final protein content; however, treatment 2 had lower variability, and no significant findings can differentiate NDF content. Therefore both substrate ratios could be considered for use as SMS feed. The results for these treatments are compared to typical dairy cow diets below in Table 2. As predicted, both treatments had much higher protein contents than these typical diets, with over three times the protein required for early lactation cows, demonstrating their potential use as protein supplements. Additionally, both treatments contain around twice the calcium content of a typical diet, so use as a calcium supplement is also possible. It should be noted that the NDF content of both feeds is quite high, more than twice the minimum for dry cows. This is understandable given the high proportion of cardboard present in both treatments, but it unfortunately detracts from the benefits of high protein content due to the inverse correlation between digestibility, voluntary intake, and NDF content.

Table 2. Comparison of typical diet formulations for dairy cows to post-fruiting SMS values for treatments 2 (20% coffee) and 3 (40% coffee).

	Typical diets			SMS values	
	Early	Mid	Late/Dry	20% coffee	40% coffee
Crude Protein (g kg^{-1} DM)	17	14	12	52.00 ± 0.346	65.97 ± 2.74
NDF (%)			33 (min) *	73.70 ± 3.76	71.67 ± 2.91
Calcium (g kg^{-1} DM)	8	6	5	14.6 ± 3.90	12.4 ± 4.40
Phosphorus (g kg^{-1} DM)	4.5	3.5	3.0	0.70 ± 0.140	0.75 ± 0.070
Magnesium (g kg^{-1} DM)	1.8	1.5	1.5	1.03 ± 0.150	1.00 ± 0.170
Sodium (g kg^{-1} DM)	1.8	1.5	1.5	0.50 ± 0.100	0.40 ± 0

All typical diet data from Phillips et al.[10] unless noted

*Erickson & Kalscheur [29]

It is important to consider the application context when discussing use of these treatments as dairy cattle feed. Due to the palatability issue highlighted earlier, the SMS feed should only be used in small quantities (less than 20%), and due to the undesirably high NDF content of the SMS, this issue of palatability may likely be exacerbated. One solution could be use of SMS in a compound feed, in which several ingredients are mixed to supplement nutritional intake of ruminants whose diet consists mainly of forage intake[30]. Compound feeds are typically pelleted; however, another option is pelleting the SMS as a stand-alone supplement without other additions. Pelleted feeds are easier to handle and distribute because they have a reduced dry matter content compared to non-pelleted feeds. During the pelleting process, the moisture content of the feed is reduced, and the feed compressed, resulting in increased bulk density and a corresponding reduction in transportation costs[31]. Pelleted feeds also have their energy content increased compared to the raw input material due to the addition of oil during the pelleting process and the common use of sugarcane molasses as binding agents[10]. Additional advantages of pelleting include enhancement with additives for a number of reasons, such as increased nutritional value and increased palatability[31,32]. On-site pelleting could make transportation of SMS easier and more economical; however, the overhead costs of pelleting must be considered. Pelleting costs vary depending on operation size, but assuming a rough production cost of €101 per tonne of DM for straw pelleting, with major costs being raw materials (66%), various plant operations (21%), and labour (9%)[33], an estimated cost for SMS pelleting can be calculated. Considering SMS as a waste, the only raw material cost will be transportation of SMS, which will conservatively be estimated at 30% of raw material costs, resulting in a conservative total production cost of €50.3 per tonne of DM ($69.6 CAD). If pelleting occurs at the site of SMS production, transportation costs are negated, and total production cost is reduced to €30.3 per tonne of DM ($41.9 CAD). Considering current feed costs (in $CAD / tonne DM) of $312 for hay, $80 for corn silage, or $687 for performance supplements[34], pelleting SMS for feed is a viable option which should be researched further.

Conclusions

In this study, golden oyster mushrooms (*Pleurotus citrinopileatus*) were grown on three treatment substrates of cardboard and SCG, with the goal of determining which substrate ratio would produce the best SMS for use as dairy cattle feed. The hypothesis was that higher substrate SCG content would result in better SMS for feed, specifically stating that higher SCG content would result in lower fiber content and higher protein content. Hypothesis 1 was rejected due to high variability in treatment results; no statistically significant difference in fiber content between treatments could be proven. Hypothesis 2 was accepted. High variability in treatment 1 was an issue which must be further investigated; however, the t-test showed a statistically significant difference between protein content in treatments 2 and 3, with treatment 3 having 13.97% higher protein content. When comparing treatment results to typical dairy cattle diets, treatments 2 and 3 were found to be suitable for further study due to their high protein and calcium content post-fruiting.

In addition to nutritional advantages offered by SMS feeds, there are potential economic benefits, as the high input cost associated with dairy cattle feed could be greatly reduced through use of a waste product such as SMS to supplement feed. Although pelleting and transportation would have associated costs that may mitigate the economic advantage of using a waste product, it is still possible that this alternative feed would be more cost effective than traditional feeds, especially when considering the nutritional advantages offered. Further study is needed to verify this.

Limitations

The two most pressing issues found in this paper were palatability, high fiber content, and high variability. Palatability can be addressed by pelleting the SMS for use as a feed additive; however, high NDF is an issue which may only be solved by replacing cardboard with another lignocellulosic waste. Variability likely resulted from lack of sterile lab space and environmental control, as well as small sample sizes. Future studies should correct these issues and should also collect and analyse fruiting data as this would enhance discussion and expand the scope of the study.

Outlook

Further research must be done to assess the practical applications of the results presented here. A full feeding trial should be conducted to better determine the palatability and digestibility of pelletized SMS, ideally using both pelleted and non-pelleted SMS feeds; methods such as those used in Adamovic et al.[17] may be of use in designing these trials. An economic analysis should also assess the practicality of implementing these recommendations and may help farmers and mushroom producers pursue these changes, which can be an economic risk.

Acknowledgements

This research was conducted as part of the McGill Course BREE 497: Bioresource Engineering Project. The author thanks Dr. Grant Clark for his support and encouragement as project supervisor, as well as the McGill Ecological Engineering Lab and the McGill University Bioresource Engineering Department for helping fund this research.

References

1. Royse, D. J., Baars, J. & Tan, Q. in Edible and Medicinal Mushrooms (eds Diego, C. Z. & Arturo P.-G.) 5-13 (Wiley-Blackwell, 2017).

2. Beaulieu-Fortin, S. Economic Profile of the Mushroom Industry in Canada (2019); https://www150.statcan.gc.ca/n1/pub/21-004-x/2019001/article/00001-eng.htm

3. Grimm, D., Kuenz, A. & Rahmann, G. Integration of mushroom production into circular food chains. Org. Agric. 11, 309-317 (2021).

4. Zhang, C. K., Gong, F. & Li, D. S. A note on the utilisation of spent mushroom composts in animal feeds. Bioresour. Technol. 52, 89-91 (1995).

5. Zisopoulos, F. K., Ramírez, H. A. B., Goot, A. J. v. d. & Boom, R. A resource efficiency assessment of the industrial mushroom production chain: The influence of data variability. J. Clean. Prod. 126, 394-408 (2016).

6. Beyer, D. Impact of the Mushroom Industry on the Environment (2011); https://extension.psu.edu/impact-of-the-mushroom-industry-on-the-environment

7. Mohd Hanafi, F. H. et al. Environmentally sustainable applications of agro-based spent mushroom substrate (SMS): an overview. J. Mater. Cycles Waste Manag. 20, 1383-1396 (2018).

8. Finney, K. N., Ryu, C., Sharifi, V. N. & Swithenbank, J. The reuse of spent mushroom compost and coal tailings for energy recovery: Comparison of thermal treatment technologies. Bioresour. Technol. 100, 310-315 (2009).

9. Grimm, D. & Wösten, H. A. B. Mushroom cultivation in the circular economy. Appl. Microbiol. Biotechnol. 102, 7795-7803 (2018).

10. Phillips, C. J. C. Cow feeding. in The encyclopedia of farm animal nutrition (ed Fuller, M. F.), 120-122, doi:10.1079/9780851993690.0000 (CABI Publishing, 2004).

11. Lucio, B. et al. Protein fraction, mineral profile, and chemical compositions of various fiber-based substrates degraded by Pleurotus ostreatus. BioResources 15, 8849-8861 (2020).

12. Stamets, P. Growing gourmet and medicinal mushrooms. (Ten Speed Press, 2000).

13. Weiss, M. J., Wilson, L. L., Harpster, H. W. & Schisler, L. C. Preliminary Report: Feeding of Mushroom Waste Materials to Livestock (Penn State University, 1980).

14. Wilson, L. L., Turner, M. L., Weiss, M. J. & Harpster, H. W. Mushroom industry wastes studied as livestock feeds. Science in Agriculture 30, 7 (1983).

15 . Rzymski, P., Mleczekb, M., Siwulski, M., Jasinska, A., Budkad, A., Niedzielskie, P., Kalac, P., Gasecka, M. & Budzynska, S. Multielemental analysis of fruit bodies of three cultivated commercial Agaricus species. Journal of Food Composition and Analysis 59, 170-178, doi:10.1016/j.jfca.2017.02.011 (2017).

16. Islam, M. T., Zakaria, Z., Hamidin, N., Bin Mohd Ishak, M. A. & Shi Fern, C. The Management of Humidifying Treatment for Low Contamination Risks During Indoor Cultivation of Grey Oyster Mushroom (Pleurotus pulmonarius). MATEC Web Conf. 97, 01080 (2017).

17. Rinker, D. L. in Edible and Medicinal Mushrooms (eds Diego, C. Z. & Arturo, P.-G.) 427-454 (Wiley-Blackwell, 2017).

18. Adamović, M. et al. The biodegradation of wheat straw by Pleurotus ostreatus mushrooms and its use in cattle feeding. Anim. Feed Sci. Technol. 71, 357-362 (1998).

19. Glück-Thaler, E. Oyster Mushroom Cultivation on Coffee: A Unique Opportunity for Urban Agriculture. Preprint at https://escholarship.mc-

gill.ca/concern/papers/z890rw03f. (2012).

20. Contreras, E. P. Soaking of substrate in alkaline water as a pretreatment for the cultivation of Pleurotus ostreatus. J. Hortic. Sci. Biotechnol. 79, 234-240 (2004).

21. van Hullebusch, E. D. et al. Nature-Based Units as Building Blocks for Resource Recovery Systems in Cities. Water 13, 3153 (2021).

22. Shields, T. How to Build a Shotgun Fruiting Chamber (SGFC) in 5 Easy Steps; https://learn.freshcap.com/growing/how-to-build-a-sgfc/

23. Arevalo, W. E. C. DIY mushroom cultivation: growing mushrooms at home for food, medicine, and soil. (New Society Publishers, 2019).

24. Reed, J. Neutral-detergent fibre. in The encyclopedia of farm animal nutrition (ed Fuller, M. F.), 398, doi:10.1079/9780851993690.0000 (CABI Publishing, 2004).

25. Thunes, C. Horse Hay Analysis: What are 'ADF' and 'NDF'? (2019); https://thehorse.com/164773/horse-hay-analysis-what-are-adf-and-ndf/

26. Shapiro, S. S. & Wilk, M. B. An Analysis of Variance Test for Normality (Complete Samples). Biometrika 52, 591-611 (1965).

27. Zaiontz, C. Basic Concepts for ANOVA; https://www.real-statistics.com/one-way-analysis-of-variance-anova/basic-concepts-anova/

28. Zaiontz, C. T-test analysis tool. https://www.real-statistics.com/students-t-distribution/t-test-analysis-tool/

29. Erickson, P. S. & Kalscheur, K. F. Nutrition and feeding of dairy cattle. Anim. Agric. 157-180 (2020).

30. Garnsworthy, P. C. Compound feed. in The encyclopedia of farm animal nutrition (ed Fuller, M. F.), 116, doi:10.1079/9780851993690.0000 (CABI Publishing, 2004).

31. Goodwill, M. Pelleting. in The encyclopedia of farm animal nutrition (ed Fuller, M. F.), 422, doi:10.1079/9780851993690.0000 (CABI Publishing, 2004).

32. Rouchouse, S. Study of Temperature in the Pelleting of Thermo-Sensitive Products: Changes in Temperature during the Pelleting Process (2020); https://phileo-lesaffre.com/en/study-of-temperature-in-the-pelleting-of-thermo-sensitive-products-changes-in-temperature-during-the-pelleting-process/

33. Nolan, A., McDonnell, K., Devlin, G., Carroll, J. & Finnan, J. Economic Analysis of Manufacturing Costs of Pellet Production in the Republic of Ireland Using Non-Woody Biomass. Open Renew. Energy J. 3 (2010).

34. Salfer, J. Managing high feed costs (2022); https://extension.umn.edu/dairy-news/managing-high-feed-costs 27. Zaiontz, C. Basic Concepts for ANOVA; https://www.real-statistics.com/one-way-analysis-of-variance-anova/basic-concepts-anova/

Submitted: 1/23/2022
Accepted: 03/22/2022

Research Article

[1]McGill University, Montreal, QC, Canada

[2]Harvard University, Cambridge, MA, United States

Email Correspondence

claire.wamboldt@mail.mcgill.ca

Claire Wamboldt[1], Marilyn Ahun[2], Frances E. Aboud[1]

An Analysis of Behaviour Change Techniques used in the Care for Child Development Parenting Program

Abstract

Care for Child Development (CCD) is one of several parenting programs implemented in low- and middle-income countries to improve child cognitive development outcomes by increasing responsive stimulation practices in caregivers of young children. Broadly, these programs have been demonstrated to be effective. However, there is markedly little high-quality evidence for the effectiveness of CCD. Despite this, CCD is promoted by UNICEF and the World Health Organisation as an evidence-based program and is implemented in many countries. We conducted a scoping review, including grey literature and a systematic search of published literature, to obtain an overview of the available evidence. We also performed an analysis of two quantitative outcomes, child cognitive development and caregiver behaviour, to investigate their correlation with behaviour change techniques used in CCD program implementation. We found no significant correlation between any behaviour change techniques and child cognitive development outcomes. There was a significant correlation between the techniques of performance and social support, as well as the total number of techniques used, and caregiver behaviour outcomes. This analysis was limited by the quality of reported data available about the program; of 27 total identified papers, only 14 reported quantitative data regarding either child cognitive development or caregiver behaviour change. Inconsistent reporting of this data required us to use a rating system to perform our analysis; we consequently lost specificity. Even those papers that did report quantitative data were subject to methodological flaws; the measures and study designs used did not always inspire confidence in their results. We concluded that CCD is not one single, well-defined program, and that there is an important distinction to be made between CCD-based and CCD-informed programs. The generic *Care for Child Development* Package (2012) is a framework that contains too many gaps to be easily adaptable. Not enough high-quality studies of this program are available to draw concrete conclusions concerning its effectiveness, in whole or in part.

Introduction

Although 5.2 million children under the age of 5 worldwide died in 2019, almost 50 times that number – close to 250 million children – survived but did not reach their full developmental potential[1,2]. Developmental potential refers to mental development, including cognitive, language, motor, and social-emotional abilities. The degree of this loss of potential is indicated by the difference in children's current development levels and what they would have achieved in a healthier and more nurturing environment. In this nurturing environment, they would have had access to adequate stimulation and nutrition, which promote healthy development[3]. The failure to reach their full developmental potential would translate to an estimated 20% loss in adult productivity[4].

Early childhood, particularly the years from birth to the age of 5, is a critical period of growth and development for children. According to reviews of recent research, healthy inputs such as adequate stimulation and nutrition during these years of life have had positive impacts over the long term[3]. However, parents and health workers in low- and middle-income countries (LMICs) did not believe that mental development started when scientists said it does. The provision of stimulation through play and communication was not common in many LMICs[5,6]. Parents strove to keep children healthy and fed without much understanding about brain development and the importance of communication and play[7]. There was a tendency among caregivers in LMICs to perceive learning and other important skills as being acquired by children after 3 years of age, instead of being developed during infancy. In contrast, neuropsychological evidence supported the need for stimulation from birth[8].

Early childhood development (ECD) interventions include a range of programs that aim to help children reach their full developmental potential. Parenting programs, particularly those centered on the provision of psychosocial stimulation, strive to support caregivers in enabling their children to reach their full abilities by providing them with a nurturing environment. Significant behaviour change is required on behalf of these caregivers to deliver this essential stimulation. These interventions are of such import that attention to ECD and getting all children developmentally on-track has been included in the United Nations (UN) sustainable development goals for 2030[9]. This report presents a well-known ECD program that is used worldwide, *Care for Child Development* (CCD), and analyzes data regarding the behaviour-change techniques used in the implementation of this program. The objective of this program is to help caregivers provide responsive stimulation to their children through play and communication. The effects of the behaviour-change techniques used in CCD to change parental practices are then examined in association with quantitative parent and child outcomes.

Evidence for the Effectiveness of Parenting Programs

Two recent systematic reviews and meta-analyses provided strong evidence for the impact of stimulation interventions on parenting outcomes and child development in LMICs[10,11]. Aboud & Yousafzai (2015) reviewed 21 interventions that promoted parental stimulation and 18 interventions that aimed to improve nutrition in LMICs. They found a medium effect size for interventions that promoted parental psychosocial stimulation on children's cognitive (d=0.42) and language (d=0.47) development[10]. On the other hand, the nutrition interventions had a very small effect on cognitive and language development (d=0.09).

This meta-analysis makes the case for stimulation-focused parenting interventions. Contrary to the previous literature suggesting that nutrition was critical for children's mental development, this review suggests that its impact is less than that of stimulation[10].

Jeong and colleagues' (2021) global systematic review and meta-analysis evaluated the effectiveness of parenting interventions on child development and parenting outcomes[11]. Although they evaluated programs delivered in both high-income countries and LMICs, the results from LMICs were stronger and more relevant for the current scoping review. Pooled effect sizes for interventions in these environments indicated positive benefits of interventions on children's cognitive (*d*=0.41), language (*d*=0.35), motor (*d*=0.26), and socioemotional (*d*=0.24) development. They also found positive benefits on parenting knowledge (*d*=0.60), parenting practices (*d*=0.47), and parent-child interactions (*d*=0.47). This review highlights the importance of teaching responsive stimulation and providing early learning opportunities for children during parenting programs, especially in LMICs. It also demonstrates that parenting programs for caregivers of children in the first three years of life are overall effective for improving child development and parenting outcomes. Due to the results of these two systematic reviews, and the evidence presented above that there is a knowledge gap for parents and caregivers in LMIC, this review focuses on a parenting program as implemented in LMIC.

CCD is a flexible framework with a few core messages about responsive stimulation, focusing on communication and play. Consequently, the CCD program takes many forms in its various implementations. Most use the materials provided by UNICEF (2012): a manual for delivery agents, counselling cards that function as job aides, notes for trainers of delivery agents, a guide for clinical practice, and a framework for monitoring and evaluation. Because CCD is a framework, it requires not only adaptation to the context but also a great deal of fleshing out the details of what to do and tell parents. Materials for those delivering the program and for their trainers are available[12]. They provide illustrations of how to play and communicate with a child of different ages as well as a sequence of actions for the delivery agent: ask caregivers how they play with their child and listen, praise the caregiver, advise on how to improve and problem-solve. Many of their materials and videos appear to show clinicians delivering the messages to caregivers who arrive at a well-baby clinic. However, it is adaptable to a home or group session. One 24-month CCD program combining home and group formats in Pakistan using community health workers found medium effect sizes on cognition (*d*=0.6), language (*d*=0.7), and motor (*d*=0.5) development at 24 months of age. Children in the intervention group also had statistically significantly higher cognition, language, and motor skills at 4 years of age than the children who had not received the intervention[13].

Behaviour Change Strategies in ECD Interventions

Parenting programs seeking to guide caregivers' interactions with their children are ultimately aiming to change behaviour. Until recently, parenting programs provided parents solely with information about what children needed and why. They assumed that knowledge would change practices. However, these programs failed to significantly change caregiver behaviour and improve child development outcomes[14]. Social psychologists introduced program developers to theories and techniques of behaviour change[15]. While this helped the field of ECD advance, program developers are less concerned with testing discrete theories of behaviour change than with finding combinations of techniques that prove to be effective[16].

For example, those implementing ECD interventions began to use Bandura's social-cognitive learning theory to change parental practices. Social-cognitive learning theory posits that people learn from each other via observation, imitation, modeling, practice and feedback[17]. This theory, thus, provides explicit suggestions for delivery agents to demonstrate a new practice to caregivers and have them practice with their children. The use of techniques proposed by social-cognitive learning theory was found to translate to greater behaviour change in caregivers[10]. However, ECD intervention trials are not designed to test Bandura's theory. Rather, they are inspired by it.

The 2015 systematic review of stimulation interventions identified techniques of behaviour change used in interventions to see which led to better child outcomes[10]. The effect sizes of child development outcomes were correlated with the presence or absence of each technique. The use of small media was most highly correlated with mental development gains, followed by performance and then problem solving. Unfortunately, problem solving techniques were infrequently used, and social support even less so; thus, no conclusions could be drawn about these techniques. The authors did conclude that the more of these techniques used in an intervention, the greater the effect on child development outcomes. This evidence suggests that the techniques used in an intervention to change parental behaviour can contribute to the success of that program, and that some techniques are associated with better outcomes than others.

The purpose of this study is to examine behaviour change techniques used in the implementation of CCD and how such techniques relate to outcomes. Consequently, the objectives were to conduct a scoping review to identify publications on the implementation of a CCD program in LMIC, examine its effectiveness in improving parent practices and child development, and relate these outcomes to behaviour change techniques used by CCD programs.

Methods

This study consisted of two parts. First, we conducted a scoping review of the peer-reviewed papers and grey literature reports on CCD. We extracted key information from these reports including evidence for improved parent and child outcomes. Secondly, the extracted data about behaviour change techniques was statistically analysed to see if any techniques were correlated with parent and child outcomes.

Scoping Review

We conducted a scoping review of the published and unpublished reports of CCD implementation. We extracted information regarding the content and implementation of these programs, including the adaptations made to the CCD package. We also extracted information about the behaviour change techniques that were used by the program.

We chose to conduct a scoping review – "a form of knowledge synthesis that addresses an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field by systematically searching, selecting, and synthesizing knowledge to identify relevant studies"[18]. The aim of such a review is not to answer one specific question, but rather to provide an overview of the evidence available regarding the topic in question[19]. One goal of this project was to summarize published and unpublished reports of CCD implementation by consulting sources beyond published, peer-reviewed, journal articles. Thus, a scoping review was more appropriate than a systematic review to examine our broad and exploratory research question, "how is CCD implemented globally?"

To identify peer-reviewed, published journal articles that detail CCD-interventions, a systematic search of two databases was conducted. MEDLINE and Global Health OVID were searched on November 14, 2021. Search terms were designed to capture the various ways that CCD interventions are named when they are reported ("Care for Child Development," "Care for Development," "early childhood development" AND intervention/program/trial, trial "promoting child development"). From the two databases, a total of 497 papers were identified: 253 papers were identified on MEDLINE and 244 on Global Health OVID. To be included in this review, the articles needed to outline a parenting intervention based on CCD either in whole or in part, to be delivered to caregivers of children aged 0-3 years in a LMIC setting and be available in English. A final number of 18 peer reviewed articles were included. In addition to a search of the databases, we identified relevant papers from published systematic reviews and meta-analyses (n=2)[11]. We also looked for work by known re searchers who had conducted or collected published and unpublished reports (n=2)[20,21]. As well, we consulted reports written for earlier reviews of CCD (n=3)[22-24]. This allowed us to survey beyond peer-reviewed journal articles. Grey literature was included because the available peer-reviewed literature is limited. CCD is often implemented for purposes other than research, and the information about these implementations is more often available in internal reports and unpublished documents. Furthermore, previous reviews of CCD were not available as peer-reviewed journal articles but do contain a wealth of information. A total of 27 papers were included in this review.

Data Extraction Process

Data extraction was conducted by all three authors. We began by extracting information from the same papers, and then comparing the data extracted and addressing any inconsistencies. Two of the team members extracted information from different papers, with the third extracting information from overlapping papers to ensure we were consistent in our extraction. The information about the program, its delivery, and its outcomes, as described earlier, was added to an extraction table. The extraction table was designed to report information required by two reporting guidelines: Consolidated Advice on Reporting ECD Implementation Research (CARE) and Consolidated Standards of Reporting Trials (CONSORT)[25,26]. The literature that was reviewed for this review was not all peer-reviewed or published in journals; 18 articles were peer reviewed and 9 were not. The grey literature (n=9) was not expected to conform to rigorous reporting guidelines such as those described above, and some peer-reviewed articles were written before the CARE guidelines were created (n=9). Thus, we were not able to extract information for every column of the extraction table for all sources. This information was then condensed, and is available as Table 1.

Extraction Table Design

We extracted the information that the CARE guidelines recommend reporting. The CARE guidelines include recommendations to include information about how the program was adapted from its original version, how the workforce was trained and supervised, how attendance and acceptability were assessed, and how fidelity and quality of delivery were assessed. They lay out a strategy for improved reporting of implementation in ECD research[26]. These guidelines are recent and were designed with the improvement of the field in mind, we therefore did not expect all papers to meet the requirements of these guidelines. We also summarized the information required by the CONSORT guidelines. They require information on the beneficiaries and sample size, random assignment to intervention and control groups, the design, measurement of outcomes, blinding of assessors, and findings[25]. They provide direction for the reporting of randomised controlled trials (RCTs) so that readers can assess the trial accurately. Including the CONSORT guidelines in our extraction table allowed us to gauge the relative quality of the included articles.

Data Analysis

The behaviour change strategies that were employed in an intervention were extracted as part of our data extraction process. The presence of each of five behaviour change techniques noted in the extraction was coded as 0 (not present) or 1 (present) for each paper. It was possible for each program to use multiple techniques. The techniques that were identified were performance [self or other], problem-solving, social support from peers or family, provision of materials, and use of audio-visual or print media. Performance [other] was defined as the use of demonstrations by either the delivery agent or another beneficiary to model a behaviour, while performance [self] was identified as the caregivers practicing a behaviour themselves with their child and receiving feedback or coaching. Both kinds of performance were combined into one technique for our analysis. We identified a program as using the problem-solving technique if the delivery agent helped the beneficiary identify the facilitators and barriers of a given behaviour, as well as solutions to overcome the barriers. Social support was identified if the intervention leveraged the beneficiary's relationship with family members, community members, and resources as sources of support to facilitate behaviour change.

This technique did not have to be included deliberately as part of the program's implementation to be identified, given that social support can often occur as a by-product of group sessions when the beneficiaries start to encourage each other. However, not all interventions using the group session modality were identified as using social support; the report of the program had to include a description of the role of social support in their program and whether it was intentional or not to be identified as using this technique. Interventions providing materials such as books and play objects to their beneficiaries were identified as using the behaviour change technique of material provision, while those using print media such as pamphlets and posters, or audio-visual media such as radio or TV, to convey messages were identified as using the media technique.

The quantitative outcomes of either child cognitive development or caregiver behaviour change were not reported in a consistent way, so we rated each of the outcomes using an effect size d if available, or a p-value. The ratings assigned to each outcome were based on the reported outcome of the intervention, including effect size and p-value, and were determined by consensus by two authors (CW and FEA).

Results

Fourteen of the 27 papers reported their analysis of a quantitative outcome for child cognitive development and/or caregiver parenting behaviour. The quantitative studies included RCTs (n=4), pilot studies (n=5), and pre-post designs (n=5). The remaining sources were mainly qualitative reports on the implementation of the program from beneficiaries' and implementers' perspective. These reports presented very positive feedback overall, but sample sizes were too small for any significant analysis.

Child Development Outcomes

Thirteen of the 27 analyzed papers reported a child development outcome. Some papers reported more than one quantitative outcome. These included scores on the Bayley Scales of Infant and Toddler Development (BSID-III) (n=2), the Ages and Stages Inventory (ASQ) (n=7), the Malawi Development Assessment Tool (MDAT) (n=4), the INTERGROWTH-21st measure (n=1), and the Early Childhood Development Index (ECDI) (n=1). Direct assessments of the child, such as those using the Bayley, MDAT or INTERGROWTH-21st were of higher quality than a potentially biased parent report (ASQ, ECDI). In some cases, the assessors were not blinded to condition[27]. Of the thirteen articles, ten found at least one significant child outcome and three found no significant child outcome.

To determine the association between behaviour change strategies used and the child outcome, we first categorized the significance of child outcomes. The significance level (p-value) of the outcome was often reported, and sometimes means, standard deviations, and effect sizes were also included. With the information reported, we were not able to calculate an effect size for all the interventions. However, we wanted to determine the relation between the behaviour change techniques employed by a program and an outcome. We therefore assigned an outcome effect rating from 0 to 4 based on significance level or effect size: 0 ($d<0.10$ or $p>0.10$), 1 ($d=0.10$–0.20; or $0.10>p>0.05$), 2 ($d=0.21$–0.50; or $0.049>p>0.01$), 3 ($d=0.51$–0.80; or $0.009>p>0.001$), and 4 ($d>0.80$; or $p<0.001$). The presence of each behaviour change technique noted in the extraction was coded as 0 (not present) or 1 (present) for each paper. The techniques identified were performance [self or other], problem-solving, social support from peers or family, provision of materials, and use of audio-visual or print media.

Table 1. Condensed Extraction Tables

1. Programmatic Features

Author, year, country	Delivery modality	Intensity (duration & dosage)	CCD-Based or CCD-informed	Adaptations to CCD reported	CCD Bundled with another package	Behaviour Change techniques
Ertem et al., 2006, Turkey	Clinic visits	2 weeks, 2 weekly sessions	CCD-based	No	Yes; Reach Up and Learn	Performance [other and self]
Jin et al., 2007, China	Clinic visits	2 30-60 min sessions, 6 months apart	CCD-based	Yes	Yes; IMCI	Media [print]; performance [self &other]; problem-solving
Ertem 2009 Turkey	Training of clinic visit delivery agents	Not reported	CCD-based	No	No	No
Engle, Najimidinova, Faromuzova, 2011; Tajikistan, Kyrgyzstan, and Kazakhstan	Clinic visits	1 5min session	CCD-based	No	Yes; IMCI	No
Jones, 2012, Mozambique	Home visits	3 sessions	CCD-based	No	No	Media [print]; Performance [self]
Lingam, 2014, India and Pakistan	Home visits	27 sessions held over two years	CCD-based	No	No	Media [print], performance [self], problem-solving
Thorne 2014, Kenya	Home visits and group sessions	Not reported	CCD-based	Yes	Yes; health package	No
Yousafzai, 2014, Pakistan	Home visits and group sessions	80 min group sessions, 11-30 min home visits conducted monthly for 21.5-24 months for a total of 20 of both.	CCD-based	Yes	Yes; nutrition education package	Materials, performance [self], problem-solving
Holding, 2015, Mali	Home visits and group sessions	Not reported	CCD-based	Yes	No	Media
PATH, 2015, Mozambique	Clinic waiting room	A few days per week during morning hours	CCD-informed	Yes	No	Media
Rockers 2016, Zambia	Home visits and group sessions	Fortnightly home and group sessions delivered for 12 months; total of 20 group sessions and 23 home visits	CCD-informed	Yes	Curriculum included CCD, and Inter-agency Taskforce on HIV and ECD package	Performance [self], social support [peer]
Bayitondere, 2018, Rwanda	Clinic visits and group sessions	Not reported	CCD-based	No	Yes; PDC	
Fisher 2018, Vietnam	Home visits and group sessions	60-90 min group sessions every 2-4 weeks for 18 months (total of 19). One 45-60 min home visit during first 8 weeks postpartum	CCD-based	Yes	Yes; many	No
Gladstone, 2018, Malawi	Home visits and group sessions	2 45 minute group sessions and 2 30 minute home visits per month for six months.	CCD-based	No	Yes; Children's Health Growth and Development training materials	Materials, Media [AV and print], Social Support [family and community], performance [self, other], problem-solving
Lynch, 2018, Malawi	Home visits	60 min long session every 2 weeks for 6 months; 12 sessions total	CCD-based	Yes	Yes; many	Materials, Media [print], performance [other]
UNICEF, 2019, Paraguay	Home visits and group sessions	Not specified	CCD-based	Yes	No	Performance [other and self], Social support [family]

Table 1. Condensed Extraction Tables (contd.)

1. Programmatic Features (contd.)

Rosales, 2019, Armenia	Home visits and group sessions	4 group sessions for mothers over 14 months, 1 for fathers and 1 for grandmothers	CCD-informed	Yes	Yes; WV Armenia's maternal, newborn, and child health program	Performance [self], Materials, Social Support [Family]
Zhou, 2019, China	Group sessions	3+ available per month for 2 years	CCD-based	No	No	Materials
Barnhart, 2020, Rwanda	Home visits	60-90 minute visits weekly for 3-4 months; 12 sessions	CCD-informed	Yes	Yes; many	Performance, Problem solving, Social support
Betancourt et al., 2020 BMC, Rwanda	Home visits	15-22 sessions in 4-8 months	CCD-informed	Yes	Yes; many	Performance [other and self], materials, problem-solving
Rao 2020, Bhutan	Group sessions at a clinic	Not reported	CCD-based	No	Yes; with health incl. immunization, growth monitoring, developmental screening	No
Shah 2020, India	Clinic waiting room	1 10min session	CCD-informed	Yes	Yes; many	Performance [other and self], Materials
Shi 2020, China	Clinic visits, telephone support	2 120min clinic visits, one at age 2mo and the next at 6mo. Telephone not specified	CCD-informed	Yes	Yes; "The Best Start for Children aged 0-3 Years" and "Early Childhood Education Program for Chinese Children"	Media [print]
Xu 2020, China	Group sessions	42 possible sessions over 24 months	CCD-informed	No	Yes	No
Akhmadi 2021, Indonesia	Training of delivery agents	2 days of training for a total of 10 hours	CCD-based	No	No	No
Dovel 2021, Malawi	Group sessions at a clinic during wait times	8-24 sessions over 24 months; aligned with ART refill schedules	CCD-based	Yes	No	Social Support [peers]
Jensen et al. 2021	Home visits	12 60 min weekly sessions over 3 months, plus 2 120 min booster sessions 3 and 6 months after	CCD-informed	Yes	Yes; adapted from initial intervention with HIV-affected families	Performance [self], Social support [peer and family], problem-solving

2. Delivery Agents

Author, year, country	**Delivery Agents: Sex, Background, Compensation**	**Training: Duration, Trainers, Refreshers, Outcome Evaluation**	**Supervision, Monitoring, Mentoring**	**Integration**
Ertem et al., 2006, Turkey	Professional: Pediatricians at Ankara School of Medicine	Not reported	Researcher monitored clinic visit using Physician Counseling Skills Scale.	Integrated into pediatricians' responsibilities. Structured CDI interview performed after reason for visit was addressed
Jin et al., 2007, China	Professional: Health professional at a clinic	Trained in 4-step communication	Not reported	Performed by a health professional at a clinic.
Ertem 2009 Turkey	Professional: Health professionals (GPs, nurse-midwives); 90% women; 5 years' experience as health providers	Trained 3 days by 6 academics; Outcome evaluation w/ "Knowledge on Child Development" 30-items Pre<Post; Perceived Competence 25 items Pre<Post, rated as they delivered program	Questionnaire completed by trainees 1 year later found that 12% (midwives) were practising; health system barriers to continued implementation	Integrated into health system; received salary; GPs work in clinics; nurse-midwives do home visits
Engle, Najimidinova, Faromuzova, 2011; Tajikistan, Kyrgyzstan, and Kazakhstan	Professionals: Clinic health workers (nurses, pediatricians); Paraprofessionals: Community health workers Volunteers: Volunteers in Tajikistan; M, F	Within-country variation, some a few hours with no clinical practice; Kazakhstan 5 days with guided practice; Tajikistan 3 days with 2hrs practice; Kyrgyzstan 8 hrs with 4hrs practical practice	Kazakhstan & Kyrgyzstan: Trained workers gave 66% more messages on CCD than untrained. Trained no more likely to demonstrate or give mother chance to practice.	Yes, integrated into health system
Jones 2012, Mozambique	Facilitators from agriculture, health, and education backgrounds	2 days training	Facilitators conducting the home visits with mothers and infants were observed directly by researchers	Not specified; program run by Aga Khan foundation
Lingam, 2014, India & Pakistan	Not reported	Not reported	Not reported	Not reported

Table 1. Condensed Extraction Tables (contd.)

2. Delivery Agents (contd.)

Thorne 2014, Kenya (part of AKDN 5-country review)	Volunteer: Community health workers; At most education is high-school leaving certificate	Counsellors trained 4-5 days (insufficient); practice available to some; Trainers were Community Health Extension Workers (paid staff) who got 2 weeks' training (1 week CCD; 1 week practice training)	4x yearly Refresher course and monthly meeting by CHW; 1-time on-the-job supervision by CHEW, some peer support by working in pairs.	Yes, health
Yousafzai et al. 2014 Pakistan	Paraprofessional: Lady Health workers; Female; $85/month paid by government	3 days for responsive stim package; Trainers had Bachelor's recruited and paid by researchers; Trainers trained for 3 months; Refreshers every 6 months	Supervisors same background as trainers; train, supervise, monitor, mentor LHWs on-the-job; supervision & coaching 2x per month. Supervisors observed & rated home & group sessions.	Yes; CCD intervention integrated into the LHWs' job in the health system
Holding, 2015, Mali	Professional: Social workers and community workers	Multi country Training of Trainers on CCD, delivery agents monitored through videotape	Not reported	Yes, delivery agents delivered CCD as part of their pre-existing roles
PATH 2015, Mozambique	Volunteer: Community activists	Not reported	Not reported	Yes, integrated into the health system (delivered in clinic waiting room)
Rockers 2016, Zambia	Volunteer: Group sessions delivered by 'head mothers' who were selected by the communities; Paraprofessional: Child development agents delivered the home visit component	Head Mothers were trained by Child Development Agents ahead of each session. CDA were trained for 5 days.	CDAs provided resources for 'head mothers' before each group session; monitored attendance at group sessions; completed forms on own home visits	No. CDAs were employed full-time by the project. 'Head mothers' were volunteers from the community
Bayitondere, 2018, Rwanda	Paraprofessional: Community health workers	Not reported	Not reported	Yes, integrated into the health system
Fisher 2018, Vietnam	Paraprofessional: Members of local women's union with experience conducting community programmes; Community health workers and kindergarten teachers sometimes co-facilitated.	Trained at the start of each module; 20 times.	Not reported	Women's Union is a highly structured national social organisation; integrated into this organisation
Gladstone, 2018, Malawi	Health Surveillance Assistants (HSAs), who already had 12 weeks training	Training materials from "Caring for the Child's Healthy Growth and Development" were used. Trainers were from the Ministry of Health and Nutrition	HSAs filled out a data collection form. RAs visited two groups and two individual sessions for HSAs to ensure fidelity & data collection.	Yes, integrated into HSA's pre-existing roles
Lynch, 2018, Malawi	Professionals: 14 community workers, including child protection officers, specialist teachers, and community-based rehabilitation workers; allowance of USD $4.20 per visit	3 days training, trainers not reported	Community workers used a logbook. RA checked logs, also visited families and collected data.	Not integrated
UNICEF, 2019, Paraguay	Professional, paraprofessional: "Backpack teachers" do home visits as well as municipal-level community volunteers; M&F	362 master trainers, workforce, and officials trained; Backpack Teachers given 1 day training	Not reported	Yes, "Backpack Teachers" integrated into education sector. Other delivery agents are community volunteers and so not integrated
Rosales, 2019, Armenia	Volunteer, Professional: Facilitators who were community active, educated women, social workers, and nurses Professional: Healthcare professionals	Master trainers from WorldVision (WV) trained ECD experts who then trained the facilitators. WV, Ministry of Health, and ECD experts were trained by external consultant on screening and assessment tools, then trained professionals	Reported "supportive supervision for facilitators" i.e. feedback on performance during clinic visits	Integrated into healthcare system: Primary healthcare providers who parents saw for well-child visits integrated ECD counselling, screening, assessment, and referrals into their services.
Zhou, 2019, China	Volunteer: Volunteers needed at least 8 years of education.	Volunteers were trained by professionals, duration not reported.	Bimonthly supervision.	Not integrated
Barnhart, 2020, Rwanda	Volunteer: Community-based literate volunteers nominated by chief and elected by the community. Compensated for communication, transport to homes, 3hr work per day visiting.	2 weeks of training, two group meetings as continuous training.	Fidelity assessed by supervisors who listened to audio recording of the delivered sessions; Feedback by phone. Two in-person observations of delivery.	Integrated into social protection system: Community volunteers are part of gov't system and beneficiaries are eligible for Government social protection system (cash for work).

Table 1. Condensed Extraction Tables (contd.)

2. Delivery Agents (contd.)

Betancourt, 2020, Rwanda	Paraprofessionals: "Coaches", 3M & 1F; hired by researchers	2 weeks; Trained by researchers; Training included role-plays and active practice;	Programme managers conducted observational monitoring and feedback. Weekly phone supervision by researchers, graduate-level ECD expert supervisor, PhD psychologist.	No
Rao 2020, Bhutan	Paraprofessional: Health Assistants w/ ideal minimum education Secondary 12, some below this	Not Reported	Supervision by health officers with university degree	Yes, Health Assistants in clinics do other health-related work
Shah 2020, India	Professional: Pediatrician, Physical therapist, Research assistant	2 days (8 hours); active learning methods: demonstration, practice, role plays, discussion; reached 85% on Fidelity Checklist during practice	Not reported	Yes, integrated into health system (immunization clinics)
Shi 2020, China	Professional: Child development experts delivered the 2 clinic sessions; compensated allowances were US$130.23 per parenting training session	Not reported	Staff in the Department of Child Healthcare in community health center at Zhanlanlu Hospital were responsible for "supervising and managing" the intervention	Yes, integrated into the health system
Xu 2020, China	Health and family planning personnel, female caregivers, kindergarten teachers. Progressive increase salary = 450 RMB each month.	2 weeks training; "Skills to promote positive relationships and a warm environment"	Ongoing training to answer FAQ.	Not integrated
Akhmadi 2021, Indonesia	Volunteer: Community health workers (F) called "cadres"; Education 70% Junior high (gr 9), 30% senior high; average 10 year experience as cadres; not paid	Training duration 2 days/10 hrs for intervention (n=69); 1 day/5 hrs for controls (n=53); Trainers were Nurses with Masters/PhD; "Caregiver Knowledge of Child Development (2007)" 20 items: Int = Ctrl. Both increased from Pre to Post but scores <14 out of 40; Attitude: Int>Ctrl p=.05; Efficacy, confidence to deliver messages to caregivers: Int=Ctrl.	Supervised by nurses at PHC clinic.	Yes, integrated into health system. Cadres work in community clinics.
Dovel 2021, Malawi	Volunteer: Expert Clients (EC) HIV-positive volunteer community members who provide counselling and support to HIV-positive individuals	2 weeks	Once two ECs from each facility were trained, they were then responsible for ECD implementation at their facilities	Yes, integrated into the health system (Anti-Retroviral medication therapy); session conducted while mothers wait for medication
Jensen et al. 2021	Volunteer: Selected from the local community using a three-step process	3 weeks (120 hours); role-play-based learning, active coaching practice; trained by researchers and graduate-level experts	Supervisors were graduate-level experts trained by researcher; Weekly phone supervision, monthly group refreshers, weekly peer support groups. Fidelity assessed by supervisors who listened to audio recording of the delivered sessions; Feedback by phone. Two in-person observations of delivery.	Not integrated

Table 1. Condensed Extraction Tables (contd.)

3. Outcomes for Sources (n=14) that reported a quantitative outcome

References	**Country** **HDI**	**Child Outcome**	**Caregiver Behaviour Outcome**	**Ratings**	
				Child	**Caregiver**
Ertem et al., 2006	Turkey 0.820	Not Reported	Increase the prop. of participants with HOME scores greater than or equal to 38.	0	1
Jin et al., 2007	China 0.761	DQ gains d=0.27		2	2
Engle et al. 2011	Tajikistan 0.668 Kazakhstan 0.825 Kyrgyzstan 0.697	ASQ measure strongest in Tajikistan. Intervention>control on Gross motor (p=.009), Communication (p=.03), Problem solving, fine motor, personal-social ns	Family Care Indicators (FCI): ns	1	0
Yousafzai et al., 2014	Pakistan 0.557	BSID-III *d*= 0.6 for cognition, 0.7 for language, 0.5 for motor development at 24mo of age.	HOME *d*=0.86	3	4
Rockers et al., 2016	Zambia 0.584	INTERGROWTH-21 = 0.11	Reports positive impact on caregiver behaviour	1	1
Gladstone, 2018	Malawi 0.483	Change in MDAT: Gross motor= 0.53, fine motor=0.89, language=1.89, social=0.97.	Change in FCI: Availability of play things=1.52, reported interaction with children=0.53	2	2
Rosales, 2019	Armenia 0.776	BSID-III Cognitive scale: d=0.22 Language scale: d= 0.17 Motor scale: d= 0.17	Reports that parents from intervention communities demonstrated better support for learning compared to control sites	2	2
Zhou, 2019	China 0.761	ASQ adjusted odds ratio of suspected delay 0.69 p<.005, Intv 18% drop; Ctrl 10% drop, esp among children 12-24 m		3	
Betancourt, 2020 BMC	Rwanda 0.543	MDAT ns ASQ ns	HOME intervention M (SD) pre-23.73 (5.29) to post- 26.02 (3.59) to follow up-27.34 (4.48) p<.05, OMCI M(SD) ns	0	1
Barnhart 2020	Rwanda 0.543	MDAT ns ASQ ns	HOME M(SD): Post Interv. *d*=1.45; FU *d*=0.87 MICS M(SD): Post Interv *d*=0.55; FU *d*=1.00	0	4
Rao, 2020	Bhutan 0.654	ECDI those who did vs not participate in C4CD b-.068, SE.032, p<.035		2	
Shah, 2020	India 0.645		Items from MICS p <.05 for; look at picture book, told stories, play. p<.01 for; sing, took out of home, name/count		3
Shi, 2020	China 0.761	Overall ASQ *d*= 0.58		3	
Jensen et al., 2021	Rwanda 0.543	ASQ Gross motor: Post d=0.136; Follow-up d=0.14 MDAT Gross motor: Post d=-0.02; Follow up d=-0.02; ASQ Fine Motor: Post d=0.14; Follow up d=0.00 ASQ Communication: Post d =0.18; Follow Up d=0.13 MDAT Language Post d= 0.05; Follow Up d=0.05 ASQ Prob-Solv: Post d=0.24; Follow Up d= 0.13 ASQ Social-Emotional: Post d= 0.19; Follow Up d=0.06 MDAT all ns (see above)		1	

Child outcomes were analyzed in relation to behaviour change techniques used by the program. Table 2 shows statistics for the Human Development Index of the countries, the behaviour change techniques used by the programs, the total number of these techniques, and the rating of the child cognitive development outcome. To control for the relative development of the countries that the programs were implemented in, the Human Development Index (HDI) was also included in our analysis. The Human-Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and having a decent standard of living[28].

Higher means for behaviour change techniques implied more frequent use. Performance (self and other combined) was the most common behaviour change technique for these programs, followed by problem solving, materials, and finally media and social support. Programs used less than two techniques on average (*M*=1.846). The mean rating for child development outcomes was 1.54, a rating corresponding to a point between a small and a medium effect size.

Table 2. Statistics for Reported Child Development Outcomes and Behaviour Change Techniques

Variable	Mean	Std Dev	Minimum	Maximum
HDI	0.65	0.12	0.48	0.82
Performance	0.62	0.51	0	1
Problem Solving	0.46	0.52	0	1
Materials	0.31	0.48	0	1
Media	0.231	0.44	0	1
Total # of Techniques	1.85	1.28	0	4
Child Outcome (rating)	1.54	1.13	0	3

Table 3 presents the correlations between child outcomes and behaviour change techniques. Given the small sample size, we used $p<0.10$ as the threshold for significance. No significant correlations were found between the reported child development outcome and the behaviour change technique(s) employed by the program.

Table 3. Pearson Correlation and p-Values for Child Development Outcomes and Behaviour Change Techniques

Variable	Correlation r (p)
HDI	0.25 (0.42)
Performance	-0.34 (0.26)
Problem Solving	-0.18 (0.57)
Social Support	0.07 (0.83)
Materials	0.28 (0.35)
Media	0.23 (0.44)
Total # of Techniques	0.00 (0.99)

Caregiver Behavioural Outcomes

Ten of the 27 papers analyzed a quantitative caregiver behaviour outcome. The measures for these outcomes included the Home Observation Measurement of the Environment (HOME) (n=4), the Family Care Indicators (FCI) (n=2), Observations of Mother-Child Interactions (OMCI) (n=1), and the Multiple Indicator Cluster Survey (MICS) of parenting practices (n=1). These measures all assess parenting practices related to provision of psychosocial stimulation for the child, such as play materials, play activities, and responsive interactions. In some studies, the information reported was insufficient to calculate an effect size, so ratings were assigned on the same scale as the child development outcomes.

The presence or absence of each behaviour change technique was once again coded as 0 if absent and 1 if present. For these papers, performance [self and other] was once again the most common behaviour change technique, followed by problem solving, materials, social support, and finally media (see Table 4). On average, programs used slightly more than two techniques (*M*=2.20). The mean rating for the parent behaviour change outcome was 2.00, corresponding to a small to medium effect size (*d*=0.20-0.50).

We considered p-values less than 0.10 as significant due to the small sample size (see Table 5). The strongest correlation was between the total number of behaviour change techniques used and the caregiver behaviour change outcome (r=0.75, $p<0.02$). Social support (r=0.69, $p<0.03$) and performance (r=0.59, $p<0.08$) were the two specific techniques related to parental outcomes.

Table 4. Statistics for Reported Caregiver Behaviour Change Outcomes and Behaviour Change Techniques

Variable	Mean	Std Dev	Minimum	Maximum
HDI	0.64	0.12	0.48	0.82
Performance	0.80	0.42	0	1
Problem Solving	0.50	0.53	0	1
Social Support	0.30	0.48	0	1
Materials	0.40	0.52	0	1
Media	0.20	0.42	0	1
Total # of Techniques	2.20	1.23	0	4
Parent Outcome (rating)	2.00	1.33	0	4

Table 5. Pearson Correlation and p-Values for Caregiver Behaviour Change Outcomes

Variable	Correlation r (p)
HDI	-0.31 (0.39)
Performance	0.59 (0.07)
Problem Solving	0.47 (0.17)
Social Support	0.69 (0.03)
Materials	0.32 (0.36)
Media	-0.20 (0.58)
Total # of Techniques	0.75 (0.01)

Discussion

The goals of this review were twofold. Firstly, we conducted a scoping review to overview the global implementation of CCD. Secondly, we examined the behaviour techniques used in the implementation of CCD for a correlation with both child cognitive development and caregiver behaviour change outcomes. We found no significant correlations between behaviour change techniques and child cognitive development outcomes; however, parental behaviour changes were associated with the techniques of performance, social support, as well as the total number of techniques employed by an intervention.

The correlations with parental behaviour change support the claim made by others that the techniques of behaviour change employed by a program influence the outcomes[10]. The lack of correlation between child development outcomes and behaviour change techniques is inconsistent with this claim. This suggests that CCD implementation of these techniques may be inadequate to improve child outcomes. It could also be due to low sensitivity and bias inherent in parent-report measures of child development, such as the ASQ-III, used as a measure in five of the 13 studies reporting on child development outcomes. The results of this study support the use of the techniques of performance [self and other], social support, as well as an increase in the total number of behaviour change techniques, in further implementation of CCD.

Conclusion

Programs derived from the *Care for Child Development* package are very variable. The available evidence about their effectiveness is based on low quality research and shows weak impacts. However, it must be noted that it is inherently difficult to conduct research on an unstructured program such as this one. Of the 27 studies identified in our scoping review, only 14 reported effectiveness statistics with regards to child development and/or caregiver behaviour outcomes. Using the available evidence, we determined that caregiver behaviour outcomes were correlated with the behaviour change techniques of performance and social support, as well as the total number of behaviour change techniques used in the intervention. These techniques should be used in future implementation of CCD. We did not find evidence that behaviour change techniques used had a significant impact on the child development outcomes reported. We believe that this lack of effect is not proof that behaviour change techniques cannot influence such outcomes. Instead, it supports the idea that implementation of these techniques is not sufficient, or too variable, to impact child cognitive development. Furthermore, some of the measures used to quantify child cognitive development are subject to bias, most clearly in the case of those that depend on parental report.

One of the most interesting questions about the CCD package is whether its flexible nature is a strength or a weakness. Because it isn't a prescribed program, it can be implemented in a way that matches a cultural context. LMIC are varied, and such adaptability is vastly preferable to a "one size fits all" approach. However, the cost of a flexible program is that it requires expert assistance to be implemented. CCD does not contain enough information to be implemented by anyone. Therefore, despite its adaptability, it cannot be implemented anywhere. More randomized controlled trials that use valid measures should be conducted in order to make claims that CCD is "evidence based." Furthermore, to improve the package globally, a venue for those who have experience implementing CCD to share their expertise with those just starting out would be a huge asset. As of now, the program's adaptability comes at the cost of a need for technical support from UNICEF and/or child development experts.

Acknowledgements

I am very thankful to have had the chance to work with Dr. Aboud, a McGill Psychology Department faculty member, and Dr. Ahun, a postdoctoral fellow at Harvard University, on this project. Dr. Ahun has been commissioned by UNICEF to undertake a review of the *Care for Child Development* program; this project was completed alongside preparatory work for that undertaking.

References

1. Lu, C., Black, M. & Richter, L. M. Risk of Poor Development in Young Children in Low-Income and Middle-Income Countries: An estimation and analysis at the global, regional, and country level. Lancet Glob. Health 4, e916-e922 (2016).

2. Sharrow, D., Hug, L., Liu, Y. & You, D. Levels and Trends in Child Mortion, New York, 2020).

3. Britto, P. R. et al. Nurturing Care: Promoting early childhood development. Lancet 389, 91-102 (2017).

4. World Health Organisation. Improving Early Childhood Development: WHO guideline (World Health Organisation, Geneva, 2020).

5. Bornstein, M. H. & Putnick, D. L. Cognitive and Socioemotional Caregiving in Developing Countries. Child Dev. 83, 46-61 (2012).

6. Bornstein, M. H., Putnick, D. L. & Lansford, J. E., Deater-Deckard, K. & Bradley, R. H. A Developmental Analysis of Caregiving Modalities Across Infancy in 38 Low- and Middle-Income Countries. Child Dev. 86, 1571-1587 (2015).

7. Lindland, E. et al. Expanding the Basics: Mapping the Gaps between Expert, Decision-Maker and Public Understandings of Early Childhood Development in Kenya (FrameWorks, 2018).

8. Fox, S. E., Levitt, P. & Nelson, C. A. How the timing and quality of early experiences influence the development of brain architecture. Child Dev. 81, 28-40 (2010).

9. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development (2015); https://sdgs.un.org/2030agenda

10. Aboud, F. E. & Yousafzai, A. K. Global Health and Development in Early Childhood. Annu. Rev. Psychol. 66, 433-457 (2015).

11. Jeong, J., Franchett, E. E., Ramos de Oliveira, C. V., Rehmani, K. & Yousafzai, A. K. Parenting Interventions to Promote Early Child Development in the First Three Years of Life: A global systematic review and meta-analysis. PLoS Med. 18, e1003602, (2021).

12. UNICEF. Care for Child Development (2012); https://www.unicef.org/documents/care-child-development

13. Yousafzai, A. K., Rasheed, M. A., Rizvi, A., Armstrong, R. & Bhutta, Z. A. Effect of Integrated Responsive Stimulation and Nutrition Interventions in the Lady Health Worker Programme in Pakistan on Child Development, Growth, and Health Outcomes: a cluster-randomised factorial effectiveness trial. Lancet 384, 1282-1293 (2014).

14. Aboud, F. E. Evaluation of an Early Childhood Parenting Programme in Rural Bangladesh. J. Health Popul. Nutr. 25, 3-13 (2007).

15. Glanz, K. & Bishop, D. B. The Role of Behavioral Science Theory in Development and Implementation of Public Health Interventions. Annu. Rev. Pub. Health 31, 399-418 (2010).

16. Briscoe, C. & Aboud, F. Behaviour Change Communication Targeting Four Health Behaviours in Developing Countries: A review of change

techniques. Soc. Sci. Med. 75, 612-621 (2012).

17. Bandura, A. Social foundations of thought and action: A social cognitive theory. (Prentice-Hall, 1986).

18. Colquhoun, H. L. et al. Scoping Reviews: Time for Clarity in Definition, Methods, and Reporting. J. Clin. Epidemiol. 67, 1291-1294 (2014).

19. Munn, Z. et al. Systematic Review or Scoping Review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med. Res. Methodol. 18, 143 (2018).

20. Karuskina-Drivdale, S., Kawakyu, N. & Mulhanga., F. A Playbox Intervention in Health Facility Waiting Rooms in Mozambique: improving caregivers' knowledge, skills, and communication with health professionals. Int. J. Birth Parent Educ. 6 (2019).

21. Rosales, A. et al. Behavior Change Communication Model Enhancing Parental Practices For Improved Early Childhood Growth And Development Outcomes In Rural Armenia: A quasi-experimental study. Prev. Med. Rep. 14, 100820 (2019).

22. Lucas, J. E., Richter, L. M. & Daelmans, B. Care For Child Development: An intervention in support of responsive caregiving and early child development. Child Care Health Dev. 44, 41-49 (2018).

23. Philbrick, W. C., Patel Priya C. & Yousafzai A. Care for Child Development (CCD): An approach to enhance nurturing care in the XXI Century. (Harvard School of Public Health, UNICEF, RTI International; Tanzania, 2017).

24. Thorne, C. Care For Child Development in Five Countries. (Aga Khan Foundation, 2014).

25. Schulz, K. F., Altman, D. G. & Moher, D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. BMJ 340, c332 (2010).

26. Yousafzai, A. K., Aboud, F. E., Nores, M. & Kaur, R. Reporting Guidelines for Implementation Research On Nurturing Care Interventions Designed To Promote Early Childhood Development. Ann. N. Y. Acad. Sci. 1419, 26-37 (2018).

27. Jin, X. et al. "Care for Development" Intervention In Rural China: A prospective follow-up study. J. Dev. Behav. Pediatr. 28, 213-218 (2007).

28. United Nations Development Programme. Human Development Index (2020); http://hdr.undp.org/en/content/human-development-index-hdi

29. Easthall, C., Song, F. & Bhattacharya, D. A Meta-Analysis of Cognitive-Based Behaviour Change Techniques As Interventions to Improve Medication Adherence. BMJ Open 3, (2013).

Submitted: 01/13/2022
Accepted: 03/19/2022

Research Article

Connie Corbin[1]

[1]The Open University, Milton Keynes, United Kingdom

Exploring Differing Host Cutaneous Microbiome and Immune Responses Contributing to Chytridiomycosis Susceptibility in Amphibians

Keywords

Chytridiomycosis susceptibility, immune response development, cutaneous microbiome defences

Email Correspondence

connie_corbin@outlook.com

Abstract

Chytridiomycosis is an emerging infectious disease caused by *Batrachochytrium dendrobatidis*, a fungal pathogen affecting the skin of amphibians. Chytridiomycosis is differentially affecting amphibian species and populations across the world, causing severe declines and extinctions. It is spread by contact or zoospores travelling in water sources. It is not yet fully understood why susceptibility is so variable. Understanding differing susceptibility is crucial for realising any successful conservation efforts. Multiple factors appear responsible for the differing susceptibility. The two factors being examined in this literature review are ineffective immune responses and the limitations of the cutaneous microbiome. Relevant papers' significance and limitations are discussed with their provenance and objectivity taken into consideration. The immune system of amphibians comprises innate and acquired defences. The innate immune system has been found to be counterproductive in some cases, much like how an allergic reaction in humans can be detrimental to health. An adaptive immune response has yet to be confirmed in previously exposed individuals. The cutaneous microbiome plays an initial role in the defence against harmful zoospores by making the environment unsuitable or by producing deadly secretions; this consequently prevents colonization by the zoospores. Differing levels of secretions have been measured in response to disease. Most amphibian immune systems and microbiomes are not adapted to deal with chytridiomycosis, and the fungi are adapted to exploit this weakness. There are many difficulties in studying this disease, such as recreating a natural habitat in laboratory conditions, which is vital to get accurate microbiome data. The variety of species and global spread of this disease is incredibly wide ranging with many factors to consider. Many studies are only focused on one aspect of the disease, so a holistic and global approach would be more beneficial.

Introduction

Chytridiomycosis is a skin disease caused by the fungus *Batrachochytrium dendrobatidis (Bd)*. As amphibian skin health is vital for critical functions such as respiration and drinking, chytridiomycosis is a great threat to the amphibian population. The disease was observed in the 1970's and formally described by scientists in 1998. The origins of this disease are still debated, with scientists claiming it originated from Africa[1] or Asia[2]. Chytridiomycosis is now thought to be responsible for the severe decline and extinction of amphibian species such as *Rheobatrachus vitellinus* and *Rheobatrachus silus* (northern and southern gastric brooding frogs) in Australia[1].

Chytridiomycosis has spread globally, now primarily affecting amphibians in Europe, Australia, and the Americas. Some species or populations such as *Litoria verreauxii*, a vulnerable Australian species, are susceptible to the disease whilst others such as *Rhinella marina*[1], a common cane toad, remain unaffected[2]. Amphibians cannot be grouped into affected and non-affected species, as susceptibility variations can occur intra-species. This demonstrates the complexity of establishing susceptibility theories and formulating conservation methods.

A new fungal strain, *Batrachochytrium salamandrivorans (Bs)*, emerged in 2010[3] from the recombination of strains that until now were separated geographically. *Bs* only affects salamanders[4].

It is generally accepted that global geographical movement by humans and environmental change are the primary causes of chytridiomycosis emergence and spread[5]. The introduction of invasive species, such as super spreader *Xenopus laevis*[6], has distributed *Bd* around the globe without they themselves being negatively affected. Wildlife research remains limited with much research conducted on captive species. This can be problematic, especially when considering the microbiome which is highly variable in the same species when the environment is changed. Estimates of *Bd* infection can be greatly influenced by which body part of the frog is swabbed[7].

Emerging infectious disease research looks at the epidemiologic triangle (the components contributing to disease): the host, the pathogen, and the environment. Many factors have the potential to affect amphibian susceptibility to chytridiomycosis; these include environmental conditions, host behaviour, pathogen behaviour, habitat conditions, geographical location, interspecies or intra-species interactions and the wildlife trade. The two factors explored here are host microbiome and immune responses. Research could be used to protect vulnerable species and predict the spread of disease. No effective methods of disease management have been identified.

Methods

A review of papers was conducted by searching on a variety of databases: Science Direct, Web of Science, PubMed and JSTOR. Key words such as chytridiomycosis, susceptibility, microbiome and immune responses were used to narrow down articles relating to the area of interest. Thirty-six papers and websites were read and analysed and chosen for their relevance and objectivity, including those with contrasting results to ensure a fair and unbiased approach was achieved. Instances where methods and sample sizes vary significantly are addressed. There are many key authors on this topic, but this review ensures that conclusions are drawn from a variety of scientists in a variety of locations.

Amphibian and *Batrachochytrium* Biology and Interaction

Amphibians are found on all continents apart from Antarctica. They are vertebrate ectotherms, relying on external factors to control their temperature[8]. Their habitats vary widely. They have a moist skin used for respiration. Most species metamorphose, growing from young to adult form such as tadpole to frog, by changing their body structure[9]. Amphibians are an indicator species, used to assess ecosystem health due to their sensitivity to change[10].

Batrachochytrium dendrobatidis and *Batrachochytrium salamandrivorans* are fungi which are parasitic in amphibians, leading to chytridiomycosis disease. *Bd* fungus anchors itself to the epithelial cells of the amphibian skin and gets pushed deeper via germ tubes, spreading the infection to further cells[4]. *Bd* reproduces by uniflagellate zoospores which possess a tail-like structure that allows locomotion. They are initially contained in a zoosporangium casing. As the infected epithelial cells differentiate, the fungi spores are carried to the skin surface and distributed to the environment.

Zoospores have a membrane and can live for weeks in water. They reproduce asexually and spread by direct contact, through water or moist surfaces. In laboratory conditions, zoospores can travel two centimetres in 24 hours. In fast moving water, they are able to spread faster than this. With moisture, zoospores can live for seven weeks, but when desiccated for three hours, they die. It is unsure whether *Bd* uses keratin, a protein layer of amphibian skin, as a nutrient source or a protected place to live. Ideal conditions for *Bd* are between 17°C to 25°C and a pH between 6 to 7. Outside of these parameters, *Bd* can still live, but either growth is slow or development stops[11].

Tadpoles cannot die from chytridiomycosis because they have only a small amount of keratin present until they are in adult form[12]. They can distribute spores by keratin shedding. The disease presents itself by weight loss, skin shedding, lethargy, red skin, convulsions, mouth discoloration, and feeding changes[4]. Skin diseases are devastating in amphibians as it affects vital functions such as respiration, water uptake, and osmosis. The susceptibility to chytridiomycosis is thought to affect the conservation status of the species[13].

The Amphibian Immune System

The amphibian immune system consists of components and processes similar to humans, having an innate and acquired immune system. Innate refers to an element of the system that an organism is born with, such as post injury tissue inflammation, as a protective mechanism. Acquired or adaptive (used interchangeably) immunity is a specific immunity or response developed after infection or exposure to a disease or virus, such as varicella zoster virus or chickenpox. Acquired immune response includes B and T lymphocytes, which are leukocytes (white blood cells) common in most organisms. They are controlled by specific receptors and determine the body's response to infection. Host immune system suppression is a method of survival for fungal pathogens.

No effective adaptive immune response to chytridiomycosis has been confirmed[3]. The following stages should be occurring in the immune system upon skin infection[14]:

1. Innate & adaptive components are triggered at the site of infection and in organs such as the liver and spleen.
2. Early innate immune responses include increased hepatic gene expression, such as increased CRP (C-reactive protein) productivity, as this is the liver's response to pathogens. This is an inflammatory protective response.
3. Innate responses include activation of macrophages, neutrophils, and Langerhans cells along with inflammatory signalling to produce more leukocytes.
4. Cell surface receptors, like the major histocompatibility complex (MHC) (proteins that assist the adaptive immune system in recognising disease), activate along with cytokines and anti-microbial peptides (chemicals expressed on the skin).
5. Adaptive immune responses, such as lymphocyte maturation (B and T cell receptors); and antibodies, antigens, and cytokine production; activate.

Immune systems develop and adapt over time[15]. One species 20 years ago is not sufficient to make assumptions of that same amphibian immune system now. Repetition of research is vital to stay accurate, adaptive and to observe changes over long periods.

Over-activation of Hormones and Regulatory Gene Responses

Serotonin is a hormone which acts on immune system receptors as an immunoregulator in amphibian skin. Claytor et al. studied levels of serotonin produced by *Litoria verreauxii*, a vulnerable Australian species[16]. Frogs able to maintain stable serotonin production during *Bd* infection were more likely to survive than those whose levels rose. *Bd* excretes tryptophan, an essential amino acid, in susceptible amphibians, which could be the trigger of serotonin level rise. Savage et al. came to a similar inference about excessive immune responses when looking at the gene immune responses of *Rana yavapaienis*[17]. This species is of particular interest as susceptibility is unique to the individual frog, even when living in the same population. A highly reactive immune system was likely to lead to mortality[17]. Over-activation was ineffective; instead, survival improved in amphibians that maintained a non-reactive balance in MHC variants, a T cell adaptive immune response, produced upon infection.

The amphibian immune system is reacting to *Bd*, but these reactions are counterproductive[16,17]. Amphibian survival rate decreased when these immune responses were upregulated. This is seemingly counterintuitive and not what was initially predicted. The adaptive immune response is producing more B and T lymphocytes, as it should, but *Bd* can easily destroy these. The reaction is futile and energy consuming. Savage et al. did not attempt to explain why some *Rana yavapaienis* reacted differently to others, which would be a valuable study[17]. Research relating to gene responses can be applied to a wide variety of amphibian species due to strong similarities in their genes.

Leukocytes and Lymphocyte Responses

Ellison et al. and Grogan et al. discovered susceptible amphibians have weak and late leukocyte responses which are insufficient to protect against chytridiomycosis[18,19]. Key components of the tricarboxylic acid cycle, the second stage of cellular respiration, were depleted upon *Bd* infection[19]. These components are a nutrient pathway for the immune system, required for effective function.

Ellison et al. compared the transcriptome, all coding and non-coding genes, of *Atelopus zeteki* in three different tissues, including the spleen, that are important to immune function. Comparisons were performed on naïve frogs (previously uninfected) and on frogs which previously survived *Bd* infection[18]. Interleukins, which are immune inflammatory proteins, were disproportionately represented in infected naïve frogs but were minor in previously infected frogs. The spleen showed the greatest difference in immune responses of naïve and previously infected frogs. Three hundred differentially expressed genes were found between the naïve and previously infected frogs, showing potential acquired immunity. It must be noted that the sample size of previously infected exposed frogs (n=2) was small due to the susceptibility of death from chytridiomycosis in this species. This same study should be repeated on different species.

Evidence of late-stage infection (white blood cell response) was found in the form of neutrophil-associated genes in *Xenopus tropicalis*[14]. This reaction was too late to have any significant effect on survival rates. Some immune responses were detected early in the spleen, such as interleukins; however, these responses were too weak to impact survival. The most significant finding of this experiment concerned cytochrome p450 in the liver of infected frogs. 51 out of the 70 p450 genes showed decreased expression in infected frogs. Rosenblum et al. commented that very few of the expected immune responses occurred on *Bd* infection of *Xenopus tropicalis*[14]. There was no significant difference between naïve and previously infected frogs, contrary to the conclusion made by Ellison et al.[18] This could be a result of the small sample size or a difference in response of the species studied. Contrary findings are important in deciphering factors of differing susceptibility between species.

Lymphocytes can be inhibited by metabolites produced by *Bd*. *Bd* was shown to adapt its metabolism to alter the skin environment of the host, inhibiting immune responses[20]. Researchers compared metabolites produced by *Bd* to a closely related pathogen, *Homolaphlyctis polyrhiza*, which did not produce these metabolites. They were trying to establish why this form of chytrid fungi was so effective in inhibiting the host's immune defences. The metabolites were methylthioadenosine (MTA), tryptophan, kynurenine (Kyn). MT, and Kyn, which independently inhibit lymphocytes; when acting together, they have a combined greater impact on immune inhibition, even at low concentrations. The exact mechanism MTA uses to inhibit lymphocytes in amphibians is unknown. Much of these conclusions are assumptions that mechanisms are akin to those exhibited by mammalian MTA. Further study on amphibians would be required for confirmation. This study also found that tryptophan was converted to Kyn which, when released, can interrupt the effector lymphocytes, inhibiting natural immune responses[20].

Amphibian Cutaneous Microbiome

The cutaneous microbiome is the community of bacteria and fungi that lives on the skin of most organisms. The symbiotic relationship between host and microbiome is believed to be an important factor in disease prevention and susceptibility in amphibians. Cutaneous microbiomes vary significantly between species and environments. They are subject to changes, especially during transition from juvenile to adult. The topography of the bacterial communities on the amphibians also varies greatly. *Bombina orientalis* in the wild has higher skin bacterial diversity in the dorsal regions; in captivity, however, the richness is higher in the ventral regions[21]. *Bombina orientalis* in the wild is resistant to decline from *Bd*[22]. These observations could be factors in susceptibility variation and important to consider during research.

Given that *Bd* fungi lives on amphibian skin, the cutaneous microbiome plays an important protective and defensive role. A study highlighted three preventative interactions by the cutaneous microbial community, reworded here[3]:

1. Zoospore colonization ability is reduced as the skin is already colonised by bacteria which can block adhesion sites for the pathogen.
2. The microbiome can produce secretions, changing the skin pH, making the environment unsuitable for zoospore survival.
3. Some of the bacteria species can produce enzymes and secondary metabolites which directly kill the invading microorganisms.

Defence Abilities of the Cutaneous Microbiome

Susceptible species are thought to have a less rich cutaneous microbiome than those less susceptible. Researchers tested frog species with varying susceptibilities from Panama[23]: *Atelopus certus, Colosthetus panamansis, Espadarana prosoblepon, Craugastor fitzingeri , and Strabomantis bufoniformis*. They found that 11 amplicon sequence variants (groups of bacteria) made up the core microbiome. Four were affected by *Bd*, suggesting that the presence of certain bacteria could be a good indication of *Bd* susceptibility[23]. Research found microbiomes of *Rana muscosa* and *Rana sierrae* did not recover from previous forms after *Bd* infection, likely due to some of the bacteria being affected and unable to recolonise[24]. Amphibians were only tested for recovery over 48 days. A useful follow-up study would have been examining if these amphibians are now more susceptible to mortality from chytridiomycosis with degraded microbiomes. Captive frogs were used in both these studies[23,24]. Microbiomes of wild species may have differing microbial communities which presents an experimental obstacle when studying this defence mechanism. Natural interactions would influence the bacteria able to colonise the skin. A study of a variety of species and their cutaneous microbiome community would be necessary to gather enough data to perform a test for significance of certain bacteria present against susceptibility.

The unsusceptible salamander species, *Plethodon cinereus*, was found to have *Janthinobacterium lividum* bacteria on its skin[25]. This provides an antifungal layer able to produce secondary metabolites, such as violacein, which inhibit the growth of *Bd*. When the microbiome of *P. cinereus* is altered, symptoms of chytridiomycosis worsen[3]. Researchers reduced the cutaneous bacteria on *P. cinereus* and compared this to the same species with an undisturbed microbiome[3]. They found that salamanders with removed bacteria lost significantly more body weight. Supporting results on the same species show a reduction of *Bd* prevalence where more *J. lividum* was present[26]. Treating susceptible species Rana muscosa with *J. lividum* increased their survival rate by 40%[27]. Studies on other species have found contrasting results. When *Atelopis zeteki* was treated with *J. lividum*, the survival rate did not improve. *J. lividum* may be a species-specific protective mechanism, or it may be part of a group of bacteria required[28].

When amphibians are in their optimal environment, their microbiome is most effective in dealing with *Bd* infection. Optimal temperature varies between species and populations depending on their natural environment. *Eleutherodactylus coqui*, susceptible frogs endemic to Puerto Rico, were infected with *Bd* during their optimal conditions (warm and wet) and during less ideal conditions (cool and dry)[29]. In their prime season, they were able to recruit skin bacteria that helped to fight off infection and return to the same pre-infection levels of bacterial richness and diversity. In the cool dry season, they were unable to do this, and infections rose as a result.

The microbiome of frogs freshly caught in the wild changes upon placement in a laboratory setting. Recreating wild conditions, including seasonal changes, adds value to research. The researchers used mesocosms, setting their experiment up in the natural environment, then used rRNA amplicon sequencing to detect changes in frog condition. The finding highlights that seasons have a direct effect on the cutaneous microbiomes' ability to protect against *Bd*. In the warm wet season, the frogs were significantly more able to fight infection and return to pre-infection condition[29].

Roback and Richards-Zawacki demonstrated the complexity of reaching conclusions about the microbiome[30]. They tested antifungal bacterium *Stenotrophomonas maltophilia* on *Acris crepitans*. The individual frogs with more antifungal bacteria survived longer when infected with *Bd* but only at temperatures of 14°C and not at 26°C. All amphibian species have specific and greatly varied optimum temperatures, indicating this would be a useful study to repeat on a wide variety of species for comparison against *Bd* susceptibility. This study again provides indication of the multifactorial nature of the disease.

Antimicrobial Peptides

Antimicrobial peptides (AMPs) are part of the early innate immune response in most mammals and amphibians. The immune system produces antimicrobial peptides, which are chemicals expressed on the skin to protect from infection. They can interact with the cutaneous microbiome.

Antimicrobial peptides are important in the development of host immunity and susceptibility. They are effective in decreasing the survival of zoospores on the skin of *Speleomantes*, a European salamander genus with 8 species. AMPs, although part of the innate immune system, have been found to adapt further after *Bd* exposure, making them more effective[31]. The combination of AMPs and their interactions with skin microbes affects the response to *Bd*. Both *Dendropsophus labialis* and *Rheobates palmatus*, frog species living symbiotically (beneficially interacting) in the Columbian Andes, both host *Bd* without succumbing to chytridiomycosis[32]. By testing 158 bacterial isolates (populations of microbes) found on these amphibians, 80% displayed antifungal properties. Some of the iso lates, however, promoted the growth of *Bd*. The combination and ratio of AMPs could be key in determining the susceptibility of certain amphibians. Susceptibility differences could be due to AMPs working in synergy with bacteria from the cutaneous microbiome[33]. The metabolite 2,4-diacetylphloroglucinol is produced by *Pseudomonas fluorescens*, a cutaneous bacterial species found on *Rana muscosa*. This bacteria worked with AMPs to inhibit *Bd* growth at a concentration 4-fold lower than if either acted alone. This demonstrates the importance of observing multiple aspects of host defence when considering susceptibility as evidence points to a multi-factorial explanation.

All amphibian species and populations have varying abilities within their immune processes due to their varying habitats and life cycles. This makes it difficult to draw any overarching conclusions about the ability to adapt AMPs. Age, life history, and environmental habitat would need to be considered for repeatable experiments to be valuable.

Differential Recovery Impacts

Some species or populations have made recoveries after a decline from chytridiomycosis while others have not. Understanding differential recovery could lead to a greater understanding of susceptibility.

Australia has the longest spanning history of scientifically recorded chytridiomycosis data. Scheele et al. looked at these data to establish how the epidemic has affected species long term from 1978 to 2017[24]. This type of study is vital in tracking disease trajectory to identify if the biological data supports the numbers in terms of extinctions and recoveries. Out of 238 Australian species, chytridiomycosis was believed to be responsible for 43 species extinctions or declines. However, 11 of these species are showing signs of recovery[24]. None of the research can empirically confirm whether host genetic adaptation or a decrease in *Bd* virulence is the contributing factor to recovery. Research certainly suggests host resistance is developing in the species studied in Panama[15].

Decades have passed since some studies were conducted, and species' immune systems and genetic alterations are now more readily examined. There are signs of potential adaptations and resistance. Researchers revisited an area in Panama where 12 species of frog had previously been recorded at critical levels due to chytridiomycosis[15]. The populations of *Atelopus varius* and *Colostethus panamansis* had recovered to pre-epidemic levels. *Bd* was still present in the current populations at a lower level. When previously uninfected captive frogs were exposed to *Bd*, both historic and contemporary isolates, they reacted the same as they originally did upon *Bd* emergence. However, wild samples previously exposed to *Bd* exhibited greater inhibitory effects when tested through skin secretions samples. This could indicate a developed host resistance.

Heritability of *Bd* resistant genes was demonstrated in *Bufo spinosus*[35]. Trait changes do not develop in isolation so researchers looked at the fungal burden of *Bd* and development time of *Bufo spinosus*. An associated trait was found to be to be significantly heritable, suggesting that adaptation to *Bd* could occur in this species. To conduct their research, a semi-natural environment was used in this study, and amphibians were exposed to temperature and varied environmental conditions. These findings are supported by the genetic adaptations found in the loci of *Litoria dayi*[36]. Selection was occurring for resistance to *Bd* infection. Loci are specific locations on a chromosome where a gene is located. This adaptation is occurring by heritable Bd resistant genes being selected for. This study is an unrepeated, novel finding. The researchers acknowledge that the conclusion assumes that the microbiome does not affect the outcome, which is a big assumption given the role the microbiome plays in disease defence. This study requires repetition with microbiome analysis to confirm conclusions with more certainty.

This research indicates that developments are occurring in some species to prevent or stimulate recovery from chytridiomycosis, whether through acquired immunity or genetic adaptation toward *Bd* resistance. These developments must be compared to environmental changes and other factors to determine the predominant susceptibility factors.

Conclusions and Recommendations

Although most often considered individually, this review demonstrates the importance of linking and understanding multiple factors which can contribute simultaneously to the differing susceptibility of amphibians to chytridiomycosis. Meeting the objectives of the review, the emergence, major theories, and limitations of previous theories and studies have been discussed. Combining and reviewing research reveals different factors work together to affect susceptibility such as immune responses, including AMPs interacting with the cutaneous microbiome.

Some amphibian species are still declining, whilst others are showing signs of recovery. Analysis on species recovery has revealed possible evidence of innate and acquired immunity along with *Bd* resistant heritable genes. These recent studies require corroboration and repetition. Knowledge on recovery can reveal information on susceptibility.

Limits and uncertainties discussed reveal wildlife experiments are difficult to replicate in a laboratory. This is particularly relevant in cutaneous microbiome research which is highly susceptible to change and variability. Some studies evaluated reveal particular bacteria with strong species specific *Bd* resistance. This review proposes further research in this area may produce significant findings.

The global nature of chytridiomycosis and the variety of species involved require collaboration from researchers. The weakness in current research lies in the lack of connections to other areas of biology, even in review papers where the opportunity is more practical. The sharing of information in a database would establish possible susceptibility associations not yet linked.

References

1. Berger, L., Roberts, A.A., Voyles, J., Longcore, J.E., Murray, KA. & Skerratt, L.F. History and recent progress on chytridiomycosis in amphibians. *Fungal Ecol.*19, 89-99 (2016).

2. Romero-Zambrano, GL., Bermúdez-Puga, S.A., Sánchez-Yumbo, A.F., Yánez-Galarza, J.K., Mauricio Ortega-Andrade, H.M., Naranjo-Briceño, L. Amphibian chytridiomycosis, a lethal pandemic disease caused by the killer fungus Batrachochytrium dendrobatidis: New approaches to host defence mechanisms and techniques for detection and monitoring. *Bionatura.* 6 (2021).

3. Becker, M.H., Harris, R.N. Cutaneous bacteria of the Redback Salamander prevent morbidity associated with a lethal disease. *PLoS ONE.* 5, 6 (2010).

4. Mutschmann, F. Chytridiomycosis in Amphibians. *J. Exot. Pet Med.* 24, 276-282 (2015).

5. Skerratt, L.F. et al. Spread of Chytridiomycosis has caused the rapid global decline and extinction of frogs. *EcoHealth.* 4, 125-134 (2007).

6. James, T.Y. et al. Disentangling host, pathogen, and environmental determinants of a recently emerged wildlife disease: lessons from the first 15 years of amphibian chytridiomycosis research. *Ecol. Evol.* 5, 4079-4097 (2015).

7. Puschendorf, R., Bolanos, F. Detection of Batrachochytrium dendrobatidis in Eleutherodactylus fitzingeri: effects of skin sample location and histologic stain. *J. Wildl. Dis.* 42, 301-306 (2016).

8. Pough, H.F. Amphibian Biology and Husbandry. *ILAR Journal.* 48, 203-213 (2007)

9. Zug, G.R & Duellman, W.E. Amphibian. *Britannica.* https://www.britannica.com/animal/amphibian (2020).

10. Wagner, N., Zughart, W., Mingo, V., Lotters S. Are deformation rates of anuran developmental stages suitable indicators for environmental pollu

tion? Possibilities and limitations. *Ecol. Indic.* 45, 394-401 (2014).

11. Piotrowski, J.S., Annis, S.L., Longcore, J. Physiology of Batrachochytrium dendrobatidis, a Chytrid Pathogen of Amphibians. *Mycologia.* 96, 9-15 (2014).

12. Marantelli, G., Berger, L., Speare, R., Keegan, L. Distribution of the amphibian chytrid Batrachochytrium dendrobatidis and keratin during tadpole development. *Pac. Conserv. Biol* 10, 173-179 (2004).

13. Amphibia Web. Species List. https://amphibiaweb.org/lists (2021).
14. Rosenblum, E.B. et al. Genome-wide transcriptional response of Silurana (Xenopus) tropicalis to infection with the deadly chytrid fungus. *PLoS ONE.* 4 (2009).

15. Voyles, J. et al. Shifts in disease dynamics in a tropical amphibian assemblage are not due to pathogen attenuation. *Science.* 359, 1517-1519 (2018).

16. Claytor, S.C. et al. Susceptibility of frogs to chytridiomycosis correlates with increased levels of immunomodulatory serotonin in the skin. *Cell. Microbiol.* 21. (2019).

17. Savage, A.E. et al. Sustained immune activation is associated with susceptibility to the amphibian chytrid fungus. *Mol. Ecol.* 29, 2889-2903 (2020).

18. Ellison, A.R. et al. Fighting a losing battle: vigorous immune response countered by pathogen suppression of host defences in the chytridiomycosis-susceptible frog Atelopus zeteki. *G3 Genes Genomes Genetics.* 4, 1275–1289 (2014).

19. Grogan, L.F. et al. Evolution of resistance to chytridiomycosis is associated with a robust early immune response. Mol. Ecol. 27, 919–34 (2018).

20. Rollins-Smith, L.A., Fites, J.S., Reinert, L.K., Shiakolas, A.R., Umile, T.P & Minbiole, K.P.C. Immunomodulatory metabolites released by the frog-killing fungus Batrachochytrium dendrobatidis. *Infect. Immun.* 83, 4565–70 (2015).

21. Bataille, A., Lee-Cruz, L., Tripathi, B. Kim, H., Waldman, B. Microbiome Variation Across Amphibian Skin Regions: Implications for Chytridiomycosis Mitigation Efforts. *Microb. Ecol.* 71, 221-232 (2016).

22. Fu, M. & Waldman, B. Ancestral chytrid pathogen remains hypervirulent following its long coevolution with amphibian hosts. *Proc. Royal Soc.* 286 (2019).

23. Robella, E.A., Hughey, M.C., Medina, D., Harris, R.N., Ibanez, R. & Belden, L.K. Skin bacterial diversity of Panamanian frogs is associated with host susceptibility and presence of Batrachochytrium dendrobatidis. *ISME Journal.* 10, 1682-1695 (2016).

24. Jani, A. et al. The amphibian microbiome exhibits poor resilience following pathogen-induced disturbance. *ISME Journal.* 15, 1628-1640 (2021).

25. Lauer, A., Simon, M.A., Walke, J.B. & Andre E. Common cutaneous bacteria from the eastern red-backed salamander can inhibit pathogenic fungi. *Copeia.* 3, 630-640 (2007).

26. Muletz, C.R., Myers, J.M., Domangue, R.J., Herrick, J.B & Harris R. Soil bioaugmentation with amphibian cutaneous bacteria protects amphibian hosts from infection by Batrachochytrium dendrobatidis. *Biol. Conserv.* 152, 119-126 (2012).

27. Becker, M.H., Brucker, R.M., Schwantes, C.R., Harris, R.N. & Minbiole, K.P.C. The bacterially produced metabolite Violacein is associated with survival of amphibians infected with a lethal fungus. *Appl. Environ. Microbiol.* 75, 6635–6638 (2009).

28. Becker, M.H., Richards-Zawacki, C.L., Gratwicke, B. & Belden, L.K. The effect of captivity on the cutaneous bacterial community of the critically endangered Panamanian golden frog (Atelopus zeteki). B*iol. Conserv.* 176, 199–206 (2014).

29. Longo, A.V. & Zamudio, K.R. Environmental fluctuations and host skin bacteria shift survival advantage between frogs and their fungal pathogen. *ISME Journal.* 11, 349-361 (2016).

30. Roback, J.M. & Richards-Zawacki, C.L. Temperature dependant effects of cutaneous bacteria on a frog's tolerance of fungal infection. *Front. Microbiol.* 9 (2018).

31. Fischer, M.C. & Garner, T.W.J. Chytrid fungi and global amphibian declines. *Nat. Rev. Microbiol.* 18, 332-343 (2020).

32. Flechas, S.V. et al. Microbiota and skin defense peptides may facilitate coexistence of two sympatric Andean frog species with a lethal pathogen. *ISME Journal.* 13, 361-373 (2019).

33. Myers, J.M., Ramsey, J.P., Blackman, A.L., Nichols, A.E., Minbiole, K.P.C. & Harris, R.N. Synergistic Inhibition of the Lethal Fungal Pathogen Batrachochytrium dendrobatidis: The Combined Effect of Symbiotic Bacterial Metabolites and Antimicrobial Peptides of the Frog Rana muscosa. J. *Chem. Ecol.* 38, 958-965 (2012).

34. Scheele BC. et al. After the epidemic: Ongoing declines, stabilizations and recoveries in amphibians afflicted by chytridiomycosis. *Biol. Conserv.* 206, 37-46 (2017).

35. Palomar, G., Bosch, J. & Cano, J.M. Heritability of Batrachochytrium dendrobatidis burden and its genetic correlation with development time in a population of Common toad (Bufo spinosus). *Evolution.* 7, 2346-2356 (2016).

36. McKnight, D.T., Carr, L.J., Bower, D.S., Schwarzkopf, L., Alford, R.A. & Zenger, K.R. Infection dynamics, dispersal, and adaptation: understanding the lack of recovery in a remnant frog population following a disease outbreak. *Heredity.*125, 110-123 (2020).

Submitted: 01/16/2022
Accepted: 03/19/2022

Research Article

Jennifer Jiang[1], Yun Xiao[1]

[1]McGill University, Montreal, QC, Canada

Email Correspondence

jennifer.jiang2@mail.mcgill.ca

Evaluation of Whole Cell Biosensors for Usability in On-site Detection of Two Major Classes of Antibiotics in Agricultural Soil and Water

Abstract

Human health is heavily influenced by the environment. In recent years, the contamination of soil and water by antibiotics has become a major public health issue. This is because of the selective pressure from antibiotics in the environment that results in the proliferation of antibiotic-resistant bacteria. A major contributor to the emergence of antibiotic resistance is the indiscriminate use of antibiotics in the agriculture and medical industry, followed by insufficient antibiotic-removal treatment of wastewater from these industries, resulting in the antibiotic accumulation in the environment. Limiting the use of antibiotics must be followed by extensive surveillance to track antibiotic residue levels in agricultural soil and water samples. In recent years, there has been a growing interest in the use of whole cell biosensors to monitor levels of antibiotics in agricultural samples; this is because whole cell biosensors are portable, cheaper, and simpler to operate and interpret compared to traditional methods of antibiotic detection. This review article compares the potential of existing β-lactam and tetracycline whole cell biosensors for on-site analysis of agricultural soil and water samples based on practicality, performance, robustness, and range of detection. Despite the lack of data regarding the performance of these biosensors under varying pH and temperature conditions, this review weighs the benefits and drawbacks of each biosensor to determine a promising candidate for use in on-site detection of β-lactams and tetracyclines. Of the β-lactam biosensors examined, only a *Bacillus subtilis*-based biosensor was able to detect β-lactams in water samples with high sensitivity and specificity while producing a strong and stable signal. However, this biosensor was not tested in soil samples, has a relatively long response time, and requires a spectrophotometer to view the signal. Engineering the reporter gene to produce a colorimetric signal will increase its potential in on-site detection. Of the tetracycline biosensors examined, a compact paper strip biosensor was found to be sensitive and highly practical when tested in both soil and water samples. Thus, we determined it to be the best candidate for on-site detection. This biosensor, however, also suffers from relatively lengthy response times. The realization of these biosensors as tools for antibiotic surveillance depends on further experimentation using on-site samples.

Introduction

The rapid emergence of antibiotic resistant bacteria is the greatest imminent threat to global health[1]. Antibiotic resistance is a naturally occurring process that is greatly expedited by antibiotic use, as antibiotics kill susceptible bacteria and allow resistant bacteria to survive and replicate. The pervasive overuse of antibiotics in agriculture and medicine are two major contributors to antibiotic resistance[1,2]. Conventional wastewater treatment cannot entirely remove antibiotics from sewage produced by the medical industry. Furthermore, the lack of effective antibiotic-removal treatment in agricultural systems gives rise to antibiotic accumulation in agricultural water and soil[2].

To monitor and treat antibiotic overuse, it is necessary to develop methods to support widespread and continuous surveillance of antibiotic levels in agricultural soil and water. Currently, several antibiotic detection methods are available. Traditional chemical assays use liquid chromatography and mass spectrometry to extract antibiotics from a sample[3]. These methods can be extremely sensitive, being able to detect antibiotics at a concentration as low as 0.05 ng/mL[4]. However, traditional methods require expensive equipment, experienced technicians, and complex, lengthy processing, making them impractical for on-site antibiotic detection in agricultural soils and water[4]. Conventional biosensors that use aptamers or antibodies as antibiotic recognition elements may be more portable, but are expensive and only stay sensitive for a small range of ion concentration, pH, and temperature, thus limiting their suitability for on-site detection of antibiotics[4].

Whole cell biosensors have been developed to address the limitations of chemical-analytical detection methods and conventional biosensors. A whole cell biosensor reprograms existing signalling pathways in living cells to respond to the critical levels of antibiotics by producing a visible output. Whole cell biosensors for antibiotics are usually constructed via fusing a reporter gene to a promoter, which acts as an antibiotic recognition element[5]. Whole cell biosensors are sensitive, specific, cheap, and portable. Furthermore, they produce easily interpretable results rapidly, do not require specialized equipment, and accurately report the concentration of bioavailable antibiotics[5]. These characteristics make whole cell biosensors an appealing tool for on-site analysis of agricultural soil and water samples.

This review aims to evaluate existing whole cell biosensors for two representative classes of antibiotics commonly used in agriculture and medicine: tetracyclines and β-lactams. The whole cell biosensors' usability in on-site analysis of agricultural soil and water samples will be examined. This review will consider the biosensors' [1] practicality; [2] performance, using metrics such as specificity, signal stability, and response time; [3] response format (responding in a dose-dependent manner is preferred over binary classification, since a quantitative response would provide useful information for subsequent antibiotic removal and sample treatment in these soil and water); [4] range of detection (they must be sensitive enough to detect the range of antibiotic concentration predicted to select for resistance bacteria (β-lactam: 0.25-4 ng/mL, tetracycline: 1-16 ng/mL)[6]); and [5] robustness (whether they retain their sensitivity in pH and temperature ranges presented by farm water and soil samples). While countries around the world would have varying pH, temperature range, and antibiotic concentration values in soil and water[7,8,9,10], they do not deviate greatly from the values obtained in Canada, which are shown in Table 1[11,12,13,14,15,16,17].

Table 1. pH, temperature, and antibiotic concentrations in agricultural soil and water samples in Canada

	pH	temperature (°C)	[tetracycline] (ng/mL)	[β-lactam] (ng/mL)
water	6.5-8.5 [11]	0-25 [12]	0-400 [13]	0-42.2 [14]
soil	5.5-8.0 [15]	0-20 [16]	0-249 [13]	0-6720 [17]

Biosensors for β-lactams

β-lactams are one of the most prescribed antibiotic classes and the most commonly used bactericidal agent in agriculture[18]. They are bactericidal agents that kill bacteria by inhibiting the production of peptidoglycans in bacterial cell walls. Bacterial β-lactam resistance arises mainly based on the synthesis of β-lactamase enzyme, which cleaves the β-lactam ring and inactivates the antibiotic. A number of whole-cell sensors are available for detection of β-lactams. Lautenschläger et. al. developed a *Bacillus subtilis (B. subtilis)*-based whole cell biosensor by fusing the promoter P*blaZ* with the luciferase reporter gene *luxABCDE* downstream of the β-lactam-activated BlaR1/BlaI signalling pathway[19].

Without the presence of β-lactam, BlaI represses the *blaZ* pathway by binding to the promoter P*blaZ*[20]. As β-lactam enters the whole cell biosensor, it binds to BlaR1 and activates BlaR1's proteolytic activity[20]. BlaR1 degrades BlaI and frees the target promoter, allowing the expression of the downstream luciferase gene, which emits bioluminescence. The functionality of this sensor was tested on ten β-lactam derivatives, representing all four subclasses of β-lactams—penicillins, cephalosporins, monobactams, and carbapenems, as well as on the cyclic polypeptide antibiotic bacitracin as a negative control[19]. The sensor detected all ten β-lactam derivatives with high specificity in *Streptomyces* soil isolates and water samples and had a lower detection limit of 1 ng/mL[19]. *Streptomyces* are known to produce a large variety of antimicrobial compounds, among them β-lactams[19]. The *Streptomyces* soil isolates were screened by the biosensor for β-lactam production through a modified disk diffusion assay[19]. β-lactam derivatives induced a luciferase signal within two hours that remained stable for several hours, while the control bacitracin did not induce a luciferase signal[19].

While the luciferase signal needs to be viewed using a spectrophotometer, decreasing the portability of this assay, the development of sensitive and robust hand-held luminometers would solve this problem. Alternatively, changing the reporter to β-galactosidase and using a chromogenic enzyme substrate would allow the biosensor to produce a signal that is visible to the eye. β-galactosidase reporter genes have been incorporated in whole-cell biosensors for on-site detection of bacitracin, another type of antibiotic, for as low as 49.3 ng/mL[21]. Its application in the detection of β-lactams or other types of antibiotics is a promising direction of research. This study did not test the functionality of the biosensor in varying pH and temperature conditions encountered in the field, but the authors claim that the sensor can be used to analyse weakly acidic samples[19]. B. subtilis can grow in pH 4 to 9.5, and its optimal temperature range is 25 to 35 °C; these numbers can serve as a rough estimate for the pH and temperature range of the biosensor[22]. This *B. subtilis*-based biosensor detects β-lactam derivatives with adequate sensitivity and specificity in water samples and bacterial soil isolates, as its lower detection limit falls within the range of antibiotic concentration predicted to select for resistance. It is likely to function within the pH range of most agricultural samples, and is unlikely to require temperature control, suggesting that this sensor is robust enough to be used for on-site monitoring. However, the response time was relatively slow at two hours compared to other biosensors discussed in this review, and the luciferase signal must be detected using specialized equipment. The biosensor should be tested in various pH and temperature conditions, as well as soil samples, after making a β-galactosidase reporter gene substitution to produce a colorimetric output.

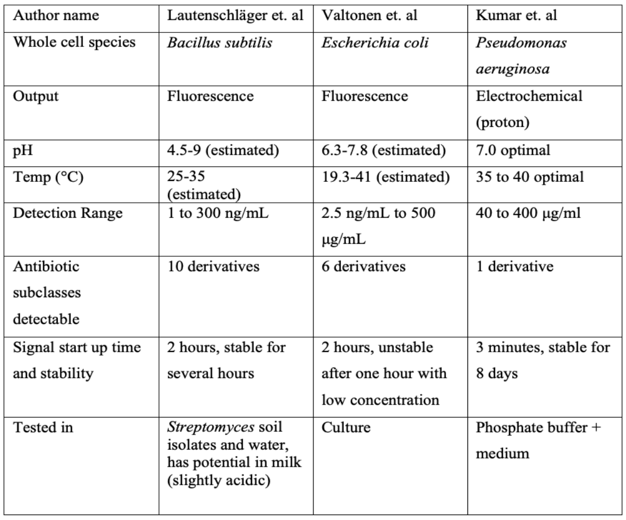
Valtonen et al. developed an *Escherichia coli (E. coli)* sensor with a luciferase reporter under the control of the β-lactam-responsive element *ampR/ampC*[23]. The inducible β-lactamase promoter *ampC* is under transcriptional control of a regulator encoded by *ampR*. The presence of β-lactam leads to breakdown of microbial cell wall murein structures. These murein products bind to *ampR*, and in turn *ampR* activates the *ampC promoter* and its downstream gene, the luciferase[23]. The biosensor was able to detect six β-lactam derivatives (ampicillin, piperacillin, imipenem, cephapirin, cefoxitin, and oxacillin) with high specificity. The biosensor had a detection range of 2.5 ng/mL to 250 µg/mL.

The signal was generated in two hours but was unstable, especially when the concentration of β-lactam was low. As with the *B. subtilis*-based sensor, acquiring a suitable handheld luminometer or the replacement of the luciferase reporter with β-galactosidase reporter would remove the requirement for a spectrophotometer. The authors only tested the biosensor in laboratory conditions and did not analyse soil or water samples. For an approximation of the pH and temperature range of the biosensor, *E. coli* can grow in pH 6.3 to 7.8 and a temperature of 19.3 to 41 °C[24]. Under these conditions, *E. coli* can survive in soil and water for approximately 90 days[24]. The biosensor is therefore likely to be robust enough to be used for on-site monitoring. The biosensor can be freeze-dried and used instantly after being rehydrated, without any growth step, eliminating the laboratory cultivation period prior to use[23]. Further experimentation is needed to determine the sensor's functionality for soil and water samples. The slow and unstable signal output is a major disadvantage of this biosensor.

Kumar et. al developed a *Pseudomonas aeruginosa (P. aeruginosa)*-based potentiometric biosensor for one derivative of β-lactam antibiotics, cephalosporin, with a detection range of 40 to 400 µg/mL[25]. A layer of biosensor was immobilized to a cellulose acetate membrane, and permeabilized with lysozyme[25]. The lysozyme-permeabilized biosensor cells hydrolyse a β-lactam ring of cephalosporin, producing cephalosporin acid[25]. The protons generated in this reaction are detected by pH electrodes[25]. The study did not test whether this biosensor can be used to analyse soil samples. The low sensitivity of this biosensor is another major drawback, as the lower detection limit of the biosensor would be too high to detect cephalosporin in most agricultural water samples. Furthermore, the biosensor is the least robust of all those presented in this review; to achieve maximum sensitivity, the temperature of the biosensor must be kept at 35 to 40 °C, and the pH must be kept at 7. Another limitation of this biosensor is that it can only detect one antibiotic derivative. Despite these drawbacks, the biosensor's strength is in its short response time: a strong and extremely stable signal is produced within three minutes, giving the biosensor the potential to be used in real-time monitoring of cephalosporin levels. To improve its range of usage, water evaporation methods can be used to concentrate the sample, meeting the detection limit. However, the evaporation process takes away the advantage of this sensor: its short response time. The low sensitivity coupled with strict pH and temperature requirements of this biosensor impede its practical application in on-site monitoring.

Of the three of β-lactam biosensors presented in this review, only the *B. subtilis*-based whole cell biosensor was tested using water samples, and only the *P. aeruginosa*-based biosensor was tested for functionality in varying pH and temperature conditions. None of the biosensors was tested on soil samples directly. With incomplete data on the robustness and performance of all three biosensors in water and soil samples, it is difficult to make a definitive judgement with regards to which biosensor is most suitable for on-site detection of β-lactams. The *B. subtilis*- and *E. coli*-based biosensors were able to detect β-lactams with high sensitivity and specificity, as its detection range covers the range of antibiotic concentration that selects for resistance[6]. They could be modified slightly to produce a colorimetric output, which is convenient for on-site monitoring in a low-tech environment. Both biosensors had relatively longer response time compared to the *P. aeruginosa* biosensor, but only the *B. subtilis*-based whole cell biosensor produced a stable signal. While the *P. aeruginosa* biosensor had a significantly shorter response time, its low sensitivity makes it unsuitable for analysis of agricultural water samples and its diffusion-dependent construct makes it unsuitable for soil sample analysis. Based on the available data, the *B. subtilis*-based biosensor is the best candidate for on-site detection of β-lactams in agricultural soil and water.

Table 2. Properties of three β-lactam biosensors



Author name	Lautenschläger et. al	Valtonen et. al	Kumar et. al
Whole cell species	*Bacillus subtilis*	*Escherichia coli*	*Pseudomonas aeruginosa*
Output	Fluorescence	Fluorescence	Electrochemical (proton)
pH	4.5-9 (estimated)	6.3-7.8 (estimated)	7.0 optimal
Temp (°C)	25-35 (estimated)	19.3-41 (estimated)	35 to 40 optimal
Detection Range	1 to 300 ng/mL	2.5 ng/mL to 500 µg/mL	40 to 400 µg/ml
Antibiotic subclasses detectable	10 derivatives	6 derivatives	1 derivative
Signal start up time and stability	2 hours, stable for several hours	2 hours, unstable after one hour with low concentration	3 minutes, stable for 8 days
Tested in	*Streptomyces* soil isolates and water, has potential in milk (slightly acidic)	Culture	Phosphate buffer + medium

Biosensors for tetracyclines

Tetracyclines are a class of bacteriostatic antibiotics that inhibit bacterial growth by inhibiting protein synthesis[26]. Tetracyclines are the most frequently used antibiotics in agriculture because they can be synthesized with high purity and are cheap to produce[26]. However, they are not absorbed well by animals' intestines and therefore contribute to significant long-term contamination of groundwater through animal waste[26]. As a result, the prevalence of tetracycline resistance is very high, reaching 67% for E. coli in several European countries[27].

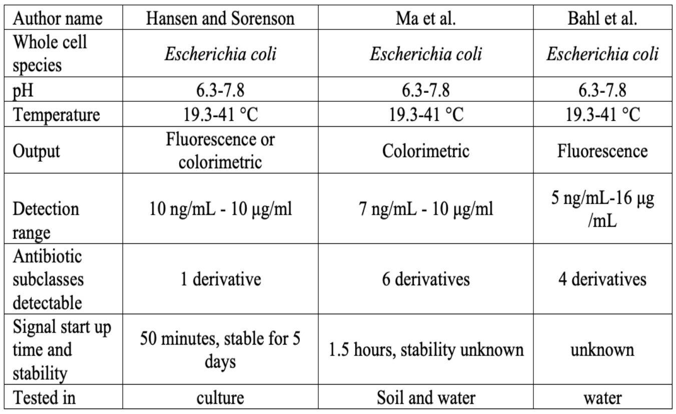
Various whole cell tetracycline biosensors are available. Hansen and Sorenson developed three *E. coli*-based biosensors by fusing the tetracycline inducible *tet* promoter with different reporter genes: *lacZYA*, which encodes β-galactosidase, Green Fluorescent Protein (GFP), and *luxCDABE* luciferase[28]. All three biosensors were quite sensitive, being able to detect as low as 10 ng/mL of the tetracycline derivative oxytetracycline[28]. The *lacZYA* biosensor had a relatively short response time of 50 minutes and produced a stable signal that lasted for 5 days, while GFP and *luxCDABE* biosensors had longer response times of three hours[28]. The authors did not report the stability of the GFP and *luxCDABE* signals. Unlike the GFP and *luxCDABE* biosensors, the *lacZYA* biosensor produces a colorimetric output and does not require specialized equipment. The authors did not analyze soil samples and only tested the biosensors under laboratory conditions. They noted that the GFP and *luxCDABE* biosensors could not function well at temperatures higher than 30 °C[28]. E. coli can grow in pH 6.3 to 7.8 and a temperature of 19.3 to 41 °C[24], but this does not mean that the biosensor can function well within this whole range. The most promising biosensor, the *lacZYA* biosensor, should be tested using soil and water samples under varying pH and temperature conditions.

Ma et al. developed a paper strip *E. coli*-based whole cell biosensor for tetracycline by liquid-drying *E. coli* cells that expressed the *lacZ* reporter gene under the control tetracycline-inducible *tet* promoter onto strips of filter paper[3]. The paper strips were photographed and analysed digitally to determine quantitatively the color intensity of the blue area of the paper[3]. The biosensor could detect six tetracycline compounds (tetracycline, oxytetracycline, chlorotetracycline, deoxytetracycline, minocycline, and methacycline) with high specificity[3]. The authors claimed that the biosensor is more robust than typical whole cell biosensors, having the ability to accommodate a wider range of pH, temperatures, ionic strengths, and presence of other contaminants such as fertilizers, because it is a self-contained paper strip[3]. However, the authors did not support most of these claims with experimental data. The biosensor is more sensitive than the biosensor by Hansen and Sorenson, with a detection limit of 7.5 ng/mL in water and in soil samples[3]. Samples were collected from tetracycline-contaminated soils, mixed with EDTA solutions acting as sensitizers, sonicated with ultrasound, centrifuged, and filtered[3]. The processed soil extracts were mixed with Lysogeny broth and tested using the paper strip biosensor[3]. If the soil extract processing can be simplified, then the paper strip format makes the biosensor highly portable and convenient, removing any requirement for well-trained personnel. The biosensor had a longer response time than the biosensor by Hansen and Sorenson, generating a clear signal 90 minutes after incubation with low concentrations of tetracycline[3]. However, the colorimetric output, relatively ease of use, and compatibility with soil samples makes this biosensor effective for on-site monitoring of tetracyclines. Additional testing is needed to validate the specificity and robustness of this biosensor. Significant attention should also focus on simplifying the soil extract processing to highlight the advantage of this biosensor: portability and ease of use.

Bahl et. al developed an *E. coli* whole cell biosensor by fusing the tetracycline-induced promoter *tet* with the GFP reporter gene[4]. This biosensor was able to achieve the lowest detection limit among all tetracycline biosensors being compared in this review: 5 ng/mL[4]. The flow cytometry was used to detect and quantify four tetracycline compounds[4]. While the concept of using flow cytometry for antibiotic detection is certainly novel and promising, the complicated design and bulkiness of flow cytometers limits the feasibility of implementing such a method in the field for on-site detection of tetracycline. Moreover, this method requires cells to be removed from the detection site and into the flow cytometer; fluorescent signals might be lost during this transportation[4]. Development of miniaturized flow cytometry devices would potentially allow this sensitive method to be used on site. Furthermore, the experimenters did not test the ability of this biosensor to analyse environmental water and soil samples. Analysis of soil samples using flow cytometry would require the separation of the biosensor bacteria from soil particulates, further reducing the efficiency of this method[29].

For the three tetracycline whole cell biosensors discussed in this review, although all their lower detection limit falls within the range of concentration that selects for resistant bacteria, there is still a trade-off between sensitivity and adaptability. The flow cytometry method developed by Bahl et. al is slightly more sensitive than the others, but far less practical for on-site monitoring. The highly portable paper strip biosensor is the most practical, has the greatest potential, and has a greater sensitivity than the biosensor produced by Hansen and Sorenson. Regrettably, of the three biosensors presented in this review, only the paper strip biosensor was tested using soil and water samples, and none of the studies thoroughly tested the ability of the biosensors to accommodate pH and temperature conditions encountered in the field.

Table 3. Properties of three tetracycline biosensors



Author name	Hansen and Sorenson	Ma et al.	Bahl et al.
Whole cell species	*Escherichia coli*	*Escherichia coli*	*Escherichia coli*
pH	6.3-7.8	6.3-7.8	6.3-7.8
Temperature	19.3-41 °C	19.3-41 °C	19.3-41 °C
Output	Fluorescence or colorimetric	Colorimetric	Fluorescence
Detection range	10 ng/mL - 10 µg/ml	7 ng/mL - 10 µg/ml	5 ng/mL-16 µg /mL
Antibiotic subclasses detectable	1 derivative	6 derivatives	4 derivatives
Signal start up time and stability	50 minutes, stable for 5 days	1.5 hours, stability unknown	unknown
Tested in	culture	Soil and water	water

Reporter genes for on-site detection

Three different reporter genes have been employed in the construction of whole cell biosensor in this review: luciferase, GFP, and β-galactosidase. β-galactosidase hydrolyzes lactose, and in colorimetric assays, lactose analogues (X-gal or ONPG) become colored after hydrolysis[30].

As production of β-galactosidase increases, the intensity of the color increases in the assay[30]. On the other hand, GFP emits green fluorescence after being excited by light in the blue to ultraviolet range. Increasing production of GFP increases the intensity of green light emission[30]. Similarly, increasing the production of luciferase increases the bioluminescence intensity[30]. While the researchers of these whole cell biosensors discussed in this review seem to favor β-galactosidase as reporter due to its colorimetric output identifiable by the naked eye, fluorescent protein and luciferase-based reporters need no additional substrates and are more suitable for quantitative measurements over a dynamic range[30].

Both reporters need specialized luminescence reader for quantitative measurements, and while GFP reporters must be excited, luciferase reporters do not. GFP and luciferase also have the advantage of short half-life, that luminescence of the proteins directly correlates with gene expression, for applications in real-time measurements[30]. For preliminary on-site detection of antibiotics in soil and water, the β-galactosidase reporter is more suitable since it can easily be used in a low-tech environment without specialized personnel or equipment. However, for more precise, quantitative determination of antibiotic concentration in samples to provide meaningful data for subsequent antibiotic removal as well as soil and water sample treatment, the luciferase-based reporter is more suitable.

Conclusion

In this review, the suitability of various whole cell biosensors for on-site detection of two different classes of antibiotics, β-lactams and tetracyclines, in farm water and soil was compared on the basis of practicality, performance, range of detection, and ability to accommodate varying pHs and temperatures encountered in the field. Of the β-lactam biosensors, only the *B. subtilis*-based biosensor had both sufficient sensitivity as well as signal stability. However, the *B. subtilis*-based biosensor has a relatively long response time and needs to be modified in order to produce a colorimetric output or to be used with a suitable hand-held luminometer. The *P. aeruginosa*-based biosensor had a shorter response time, but would not be able to detect β-lactams at the levels found in farm water. There is therefore a trade-off between sensitivity and response time for β-lactam biosensors, with sensitivity being the more important consideration. It should be noted that recent methods of antibiotic detection in soil involving mass spectroscopy have a response time of within four hours[31]; thus, the whole cell biosensors discussed in this study all have an advantage in response time over mass spectroscopy methods.

The three tetracycline biosensors had comparable sensitivities, but all had their advantages and disadvantages. The paper strip biosensor is significantly more convenient for on-site monitoring but has a longer response time and requires soil extract. The biosensor by Hansen and Sorenson had a slightly shorter response time than the paper strip biosensor, but had lower sensitivity. The biosensor by Bahl et al. was slightly more sensitive than the paper strip biosensor, but its use of flow cytometry makes it impractical for on-site detection. Few biosensors were tested for pH and temperature sensitivity, as most of the biosensors were only tested in laboratory conditions. Furthermore, only the paper strip biosensor was tested using soil and water samples, and the performance of the other biosensors for these types of samples is unknown. While this review attempts to infer the performance of several biosensors in soil and water samples using pH and temperature ranges for the biosensor bacteria species, the biosensor must be ultimately tested using soil and water samples in the field for accurate analysis. The development of portable equipment for quantitative measurements of the biosensor reporter signal would also improve the practicality of these biosensors. Overall, though the data supports the value of whole cell biosensors as a tool for detection of antibiotics in agricultural samples, more experimentation is needed before the available biosensors can be adequately compared and implemented for on-site detection.

Acknowledgements

The authors declare no conflict of interest.

References

1. Manyi-Loh, C., Mamphweli, S., Meyer, E. & Okoh, A. Antibiotic use in agriculture and its consequential resistance in environmental sources: Potential Public Health Implications. *Molecules* 23, 795 (2018).

2. Ebrahimi, S. M., Dehghanzadeh Reyhani, R., Asghari-JafarAbadi, M. & Fathifar, Z. Diversity of antibiotics in hospital and municipal wastewaters and receiving water bodies and removal efficiency by treatment processes: A systematic review protocol. *Environmental Evidence* 9, (2020).

3. Ma, Z., Liu, J., Sallach, J. B., Hu, X. & Gao, Y. Whole-cell paper strip biosensors to semi-quantify tetracycline antibiotics in environmental matrices. *Biosensors and Bioelectronics* 168, 112528 (2020).

4. Bahl, M. I., Hansen, L. H. & SÃ¸rensen, S. Ã. J. Construction of an extended range whole-cell tetracycline biosensor by use of thetet(m) resistance gene. FEMS *Microbiology Letters* 253, 201–205 (2005).

5. Berepiki, A., Kent, R., Machado, L. F. & Dixon, N. Development of high-performance whole cell biosensors aided by statistical modeling. *ACS Synthetic Biology* 9, 576–589 (2020).

6. Bengtsson-Palme, J. & Larsson, D. G. J. Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. *Environment International* 86, 140–149 (2016).

7. Ferguson, B., Lin, M.-H. & Gresshoff, P. M. Regulation of legume nodulation by acidic growth conditions. Plant Signaling & Behavior 8, (2013).

8. Lembrechts, J. J. *et al.* Global maps of soil temperature. *Global Change Biology* (2022). doi:10.1111/gcb.16060

9. Osbiston, K., Oxbrough, A. & Fernández-Martínez, L. T. Antibiotic resistance levels in soils from urban and rural land uses in Great Britain. *Access Microbiology* 3, (2021).

10. Liu, C., Tan, L., Zhang, L., Tian, W. & Ma, L. A review of the distribution of antibiotics in water in different regions of China and current antibiotic degradation pathways. *Frontiers in Environmental Science* 9, (2021).

11. Ph in drinking-water - world health organization. Available at: https://www.who.int/water_sanitation_health/dwq/chemicals/ph.pdf. (Accessed: 6th March 2022)

12. Government of Canada. Canada.ca (2021). Available at: https://www.canada.ca/en/health-canada/services/publications/healthy-living/guidelines-canadian-drinking-water-quality-guideline-technical-document-cyanobacterial-toxins-document.html. (Accessed: 3rd March 2021)

13. Conde-Cid, M. *et al.* Tetracycline and sulfonamide antibiotics in soils: Presence, fate and environmental risks. *Processes* 8, 1479 (2020).

14. Kulkarni, P. *et al.* Antibiotic concentrations decrease during wastewater treatment but persist at low levels in reclaimed water. *International Journal of Environmental Research and Public Health* 14, 668 (2017).

15. Grandel D. *Nutrient Stewardship* Available at: https://nutrientstewardship.org/implementation/soil-ph-and-the-availability-of-plant-nutrients/. (Accessed: 8th March 2021)

16. Williams & Gold. Williams, G. P. Canada.ca (2021). Available at: https://nrc-publications.canada.ca/eng/view/object/?id=386ddf88-fe8d-45dd-aabb-0a55be826f3f. (Accessed: 8th March 2021)

17. Cycoń, M., Mrozik, A. & Piotrowska-Seget, Z. Antibiotics in the soil environment—degradation and their impact on microbial activity and Diversity. *Frontiers in Microbiology* 10, (2019).

18. Pandey, N. Beta lactam antibiotics. *StatPearls [Internet]*. (2022). Available at: https://www.ncbi.nlm.nih.gov/books/NBK545311/. (Accessed: 8th

March 2021)

19. Lautenschläger, N., Popp, P. F. & Mascher, T. Development of a novel heterologous β-lactam-specific whole-cell biosensor in bacillus subtilis. *Journal of Biological Engineering* 14, (2020).

20. Pence, M. A. et al. Beta-lactamase repressor Blai modulates Staphylococcus aureus cathelicidin antimicrobial peptide resistance and virulence. *PLOS ONE* 10, (2015).

21. Fantino, J.-R., Barras, F. & Denizot, F. Sposensor: A whole-bacterial biosensor that uses immobilized bacillus subtilis spores and a one-step incubation/detection process. *Journal of Molecular Microbiology and Biotechnology* 17, 90–95 (2009).

22. Krulwich, T. A., Agus, R., Schneier, M. & Guffanti, A. A. Buffering capacity of Bacilli that grow at different ph ranges. *Journal of Bacteriology* 162, 768–772 (1985).

23. Valtonen, S. J., Kurittu, J. S. & Karp, M. T. A luminescent escherichia coli biosensor for the high throughput detection of β-lactams. *Journal of Biomolecular Screening* 7, 127–134 (2002).

24. Wilks, J. C. & Slonczewski, J. L. Ph of the cytoplasm and periplasm of Escherichia coli: Rapid measurement by Green Fluorescent Protein Fluorimetry. *Journal of Bacteriology* 189, 5601–5607 (2007).

25. Kumar, S., Kundu, S., Pakshirajan, K. & Dasu, V. V. Cephalosporins determination with a novel microbial biosensor based on permeabilized pseudomonas aeruginosa whole cells. *Applied Biochemistry and Biotechnology* 151, 653–664 (2008).

26. Chopra, I. & Roberts, M. Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and Molecular Biology Reviews* 65, 232–260 (2001).

27. Grossman, T. H. Tetracycline antibiotics and resistance. *Cold Spring Harbor Perspectives in Medicine* 6, (2016).

28. Hansen, L. H. & SÃ¸rensen, S. Ã. J. Detection and quantification of tetracyclines by whole cell biosensors. *FEMS Microbiology Letters* 190, 273–278 (2000).

29. Khalili, B., Weihe, C., Kimball, S., Schmidt, K. T. & Martiny, J. B. Optimization of a method to quantify soil bacterial abundance by flow cytometry. *mSphere* 4, (2019).

30. Jiang, T., Xing, B. & Rao, J. Recent developments of biological reporter technology for detecting gene expression. *Biotechnology & Genetic Engineering Reviews* 41–76 doi:10.7313/upo9781904761679.002

31. Meng, M. et al. Simultaneous extraction and determination of antibiotics in soils using a method based on quick, easy, cheap, effective, rugged, and safe extraction and liquid chromatography with Tandem Mass Spectrometry. *Journal of Separation Science* 40, 3214–3220 (2017).