Something Smells Fishy: Modeling Mass Mortality Events Under Climate Change

———————————————

A Thesis

Presented to

The Division of Mathematics and Natural Sciences

Reed College

———————————————

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

———————————————

Aaron Till

May 2018

Approved for the Division
(Biology)

———————————————
Sam Fey

Acknowledgements

I want to thank...

Sam Fey, for being relentlessly positive and guiding me through the entire process

Andrew Bray, for being a second advisor and a great professor

Keita, Hulali, and Tamara, for their patience in my presentations

Emerson Webb, for help understanding random forests

Alex Grant, for help with Latex and for being a friend

My girlfriend - Georgia,
and my housemates - Cooper, Dustin, Josh, Laura, Mitra, and Crash

 for their love and support

and my family - Mom, Dad, and Elaine,

 for making this whole thing possible

List of Abbreviations

BOM   Bureau of Meteorology
CF   Simplistic classification tree based analysis
CSIRO   Commonwealth Scientific and Industrial Research Organization
DNR   Department of Natural Resources
GLM   Generalized least means (logistic regression analysis)
MME   Mass mortality event
NLDAS   North American Land Data Assimilation
OOB   Out of bag estimation of error
RF   Random forest classification tree based analysis

Table of Contents

List of Tables

List of Figures

Abstract

There exists a rich body of literature on how climate change will affect abiotic conditions around the world (Solomon et al., 2009; McNutt, 2013). There is, however, more uncertainty about how those changes will affect living organisms. In this thesis I use data on mass mortality events (large population die-offs) of freshwater fish, as well as data on the modeled thermal metrics of Midwestern lakes, to estimate the change in fish die-off frequency in the coming century. Examining the data with logistic regression analysis shows a significant positive relationship between environmentally linked summertime mass mortality events and water temperature without showing a corresponding significant negative relationship between winterkill events and water temperature. By modeling this data with both logistic regression methods as well as classification-tree based methods, I show that fish MMEs will likely become more prevalent in the coming century, especially between 2082 and 2099. This research indicates that global warming may have catastrophic consequences for freshwater ecosystems and argues for the necessity of reducing greenhouse gas emissions.

Introduction

The discipline of Ecology primarily focuses on how populations interact with typical environmental conditions. However, atypical conditions also play a significant role in shaping biological dynamics. When a stable environment is punctuated by disease or extreme weather, regardless of how abnormal the event, the quality and composition of the biota may be permanently altered (Wethey et al., 2011). For this reason, how organisms respond - or fail to respond - to uncommon events can be as important as how organism behave under commonly occurring conditions.

Mass mortality events (MMEs) occur when many individuals of a species, or across multiple species, fail to survive an atypical condition and suffer abnormally high casualties over a short window of time (Fey et al., 2015). The term MME has a relatively ambiguous definition within the scientific literature as it is sometimes used to describe very small die-offs or events spread out over a long period of time. Depending on the objective of the research intentional events such as culling may or may not be included. In this thesis I will define MMEs according to a previously described method of fish kill analysis (La et Cooke, 2010): An MME must involve at least 25 macro-organisms dying over the course of 48 hours from any unusual event that is not part of the species natural process (such as synchronized post-spawning death in anadromous fish) and is not the result of predation by animals.

MMEs have many ecological ramifications. Understanding the cause and characteristics of an MME is key to managing natural populations and preventing the recurrence of mass mortality. In this thesis I aim to explore not just the direct causes of MMEs, but also the indirect set of environmental circumstances that lead to them.

1.1 Ramifications of MMEs

There are a variety of reasons why MMEs are important to study. Not only do these events have negative consequences for both the affected species and the wider ecosystem, but they also may signal changes in an underlying abiotic stressor. Below I outline six common concerns related to MMEs.

1.1.1 Loss of Genetic Diversity

Although MMEs do not necessarily lead to population extirpation (localized extinction), all MMEs do lead to a reduction in genetic diversity. Because MMEs eliminate a significant portion of a population the gene pool will experience a bottleneck effect

following an event. This has repercussions for the health of the population as there is less variation available for adaptation and there may develop a risk of inbreeding depression if the population remains low. Having a large standing pool of genetic information is the primary way populations can adapt to new threats with sufficient speed (Barrett et Schluter, 2008), and if MMEs suppress the gene pool the risk of extinction increases.

1.1.2   Loss of Numerical Advantage

Some species may rely on large numbers not just for genetic diversity but also for direct survival advantage, this is known as the Allee effect (Stephens et Sutherland, 1999). A dense population may be required to fend off predators, mate effectively, tackle large prey, or cooperate in other eusocial means. One example of this is environmental engineering. A cold snap in the English Channel (1962-1963) wiped out populations of reef building marine invertebrates such as mollusks and oysters that had been growing their own environments for decades. It took forty years for these organisms to recolonize the full extent of their pre-mortality range (Wethey et al., 2011). Other species may be unable to recover at all if their population dips below a critical threshold. For example, the passenger pigeon (E. migratorius) may have been hunted below the density required for population stability (Butcher, 1992), thus leading to its extinction. This is only a concern for some species as often the benefit of reduced competition can outweigh the cost of cooperation loss.

1.1.3   Increased Stochasticity of the Population Landscape

After removal of the mortality cause a population in the absence of Allee effects should spring back to some degree due to the bounty resulting from reduced competition. However, regular mortality events may lead to a loss of stability in the population. This is because consecutive negative jumps in population size will result in extirpation even with proportionate positive jumps. The time until extinction of a population is directly proportional to the size of the population, its long run growth rate, and a stochastic factor representing the variability of death and growth (Lande, 1993). Therefore, in addition to having a large gene pool species must also have stable population sizes (moderate boom and bust cycles) in order to have a long evolutionary lifespan.

1.1.4   Trophic Impacts on Biological Communities

Following an event MMEs may disrupt the existing food web. For instance, prey of the afflicted species may exhibit unrestricted growth resulting in damaging ecological consequences such as overgrazing. Also, predators of the affected species may often experience scarcity. This does not only apply to wildlife communities as reduction in game species is an economic concern to humans. In places around the world where hunted or fished species comprise a majority of dietary protein recurring MMEs may lead to food insecurity (Allison et al., 2009).

1.1.5 Non-trophic Impacts on Biological Communities

Organisms that do not directly predate or get preyed on by the afflicted species may also experience the negative ramifications of an MME. For example, the carrion may become a biohazard if it accumulates in high density, forcing sympatric populations away from the affected area.

1.1.6 Symptoms of Chronic Abiotic Stress

MMEs are also a concern not just because of what results from them, but because of what they indicate. The causal factor of an MME, if not addressed, may lead to more MMEs and possibly extinction. Some causal factors of MMEs may persist after an event, reducing the fecundity of the remaining individuals and thereby preventing a population correcting boom. The occurrence of multiple MMEs may indicate a serious underlying problem. Causal factors of MMEs should be identified and rectified in order to prevent future mortality and loss of biodiversity.

1.2 Human Contributions to MMEs

MMEs are particularly alarming because evidence shows they may be on the rise (Fey et al., 2015). There is an observed increase in occurrence of MMEs since 1940. While this trend may be a result of greater awareness and documentation of the phenomenon, there is also an increase in the average magnitude of MMEs when holding constant the increase in occurrence. This increase in magnitude is not true across all animal taxa but has proved significantly positive for birds, fishes, and marine invertebrates (see figure 1.1). It is attributed to an increase in multiple stressor events combining disease and environmental shocks (Fey et al., 2015).

There are multiple ways in which humans may cause a mass mortality event, such as habitat destruction, invasive species introduction, or mismanagement of toxic waste. In a summary of fish kills in the U.S. between 1980 and 1989 the majority were found to have anthropogenic causes (La et. Cooke, 2011).

Even when humans are not the direct culprit, human activities may play an indirect part in mass mortality events by contributing to abnormal weather conditions. Extreme weather is one of the most common triggers of MMEs. Disturbingly, extreme weather events are increasing (Moreno et Mller, 2011). The primary reason for this is climate change. The Earth's atmosphere is experiencing an increased greenhouse effect due to human burned fossil fuels, and temperatures are expected to rise, even in the most conservative model of reduced CO_2 emission, by 0.3 to 1.7 Celsius by the end of the century (IPCC, 2013). In addition to an increase in extreme weather events this warming will lead to a host of secondary environmental complications.
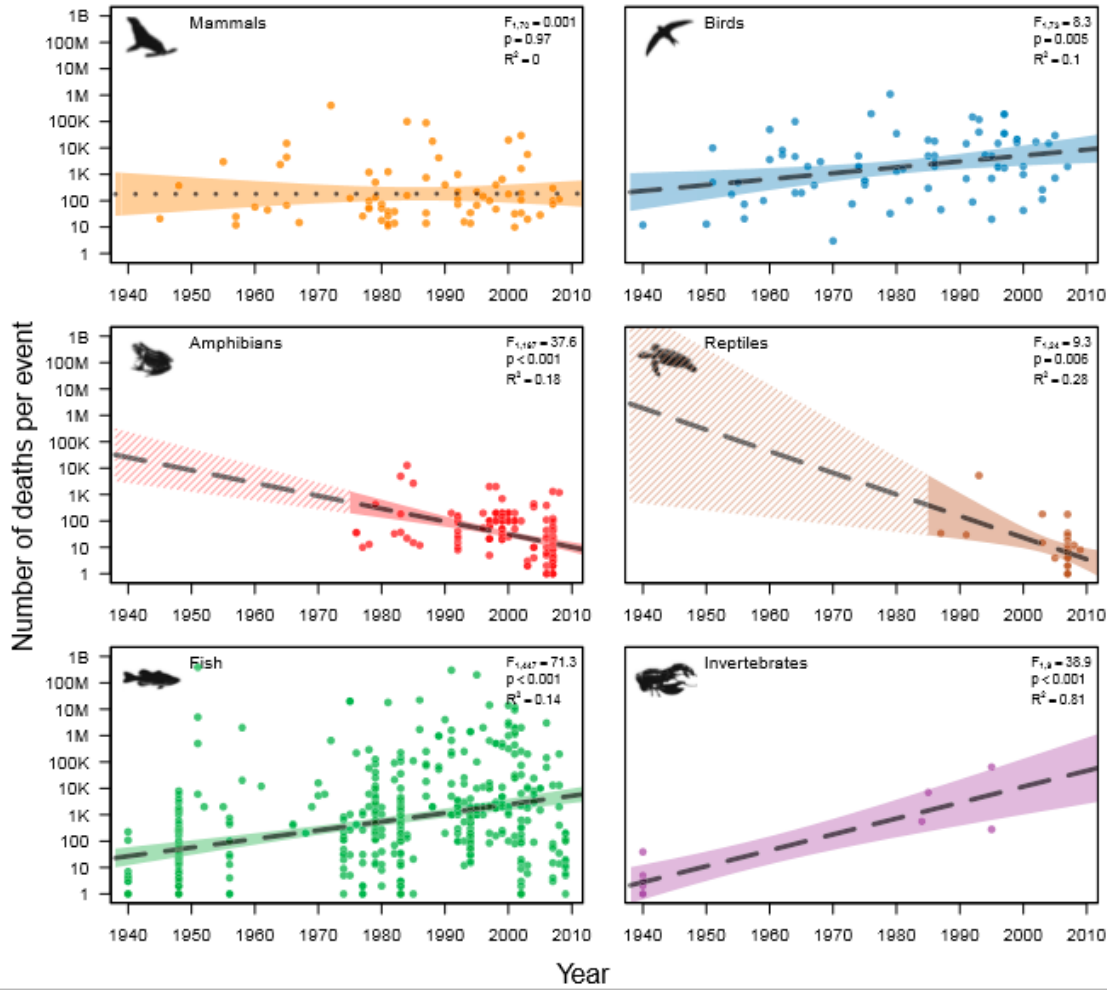
Figure 1.1: The observed change in MME magnitude across kingdoms. The increase in magnitude of MMEs for various animal kingdoms since 1940. Individual points indicate a single MME (n = 727). Dashed lines indicate significant trends. Shading indicates 95% confidence intervals. Hatched shading indicates extrapolation. Figure from Fey et al., 2015.

1.3 Freshwater Ecosystems Under Climate Change

Freshwater resources are among the environments most vulnerable to climate change, (Kundzewicz et al., 2008). Aquatic ectotherms will experience increased metabolic stress. Higher rates of evaporation will lead to drought and low water levels in some areas while melting glaciers and/or increased atmospheric water will lead to flooding in others. Additionally, as temperature increases the solubility of oxygen decreases. This means that less oxygen will be dissolved in aquatic environments - potentially leading to hypoxia. Heat stress, drought, flooding, and oxygen depletion are just a few of the many ways human caused climate change may lead to MMEs (Brookes et. Carey, 2011; Lovejoy et Hannah, 2005; Smith et Piedrahita, 1988; Xenopoulos et al., 2005).

Cyanobacteria blooms may also be increasing in freshwater environments, in part due to climate change (Brookes et Carey, 2011). Phytoplankton (including Cyanobacteria) are healthy features of aquatic systems and contribute positively to dissolved oxygen levels in moderate density, but can lead to hypoxia at too high of a concentration (Smith et Piedrahita, 1988). Thus, as temperature increases directly reduce the amount of dissolved oxygen, eutrophication may further decrease the oxygen supply of freshwater sources. There is no question that the biotic composition of lakes around the globe will be dramatically changed in the coming century.

Evidence shows that the surface of lakes have begun a variable but rapid increase in temperature that is expected to continue into the future (O'Reilley et al., 2015).

Freshwater species may be among the most susceptible to increasingly variable temperatures. Fish populations do not typically experience abnormal environmental events (Anderson et al., 2017). Additionally, the majority of fish are poikilothermic (their body temperature matches the ambient water temperature), because most lack a method of regulating internal temperature (Nilsson et Lefevre, 2016). Because water is climatically stable, a gradual increase in temperature will have a great effect on the creatures adapted to this stability. Freshwater environments are hotspots of climatic change as the point at which the experienced temperature exceeds the range of historical variance may be reached much sooner than in highly variable terrestrial environments. This same reasoning is used to argue that the tropics (which are also adjusted to a baseline of environmental stability) may be particularly susceptibility to the effects of climate change (Mora et al., 2013).

This vulnerability of freshwater species is already becoming apparent among fish that have evolved in cooler waters. As new extreme temperatures are reached these fish are no longer adapted to their current environments. It is estimated that an increase in water temperature of 1C can lead anywhere from a 3 to 30% increase in organism metabolic rate, and a 160 km shift in ideal range (Lovejoy et Hannah, 2005). Migration for freshwater species is difficult. River species may be driven to smaller tributary streams at higher elevations (genetically fragmenting larger populations), and southerly lakes and ponds may become inhospitable environments for the coldwater species they currently contain long before warm water species have a chance to colonize.

The extreme susceptibility of freshwater ecosystems to change, lakes in particular,
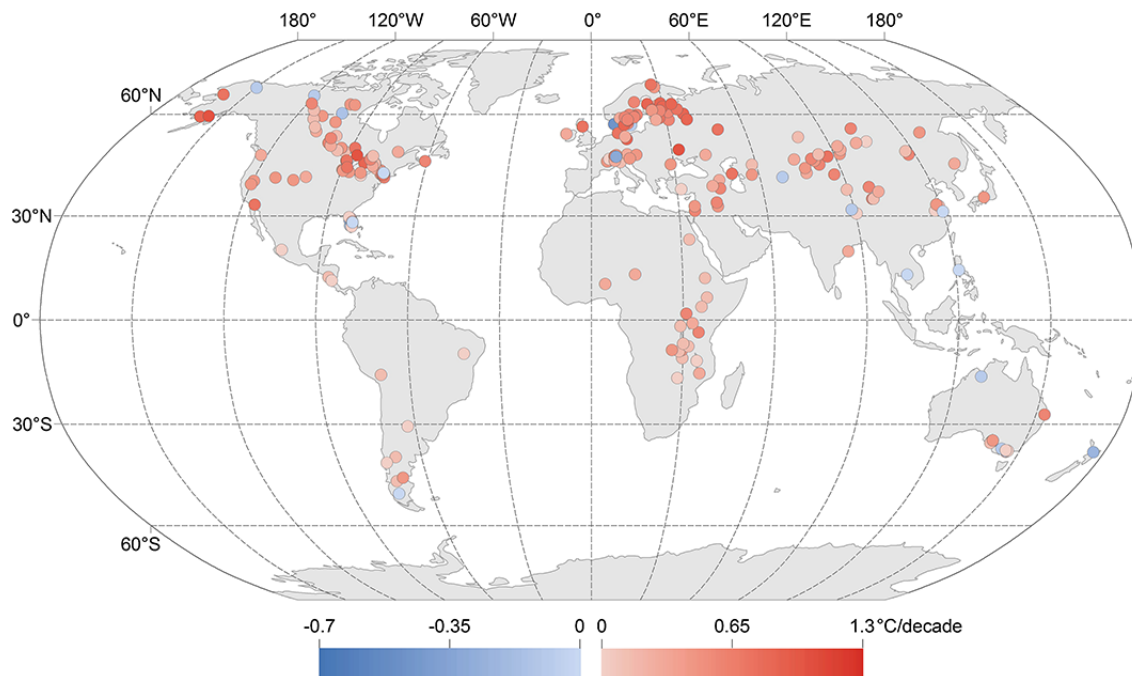
Figure 1.2: Heatmap of global lake warming. Heatmap of change in lake temperature between 1985 and 2009. Although there is a great degree of variation, the majority of lakes are warming (warming and cooling heat scales are not equivalent). Figure from O'Reilly et al., 2015.

may be an asset in measuring the effect of climate change. Lakes are useful because they have well defined boundaries and are well studied, they respond directly to temperature variations, and they are distributed worldwide in various environments (Adrian et al., 2009). Lakes may function as early indicators of intensifying conditions.

1.4   MMEs of Wisconsin Freshwater Fish

The Wisconsin Department of Natural Resources has compiled a database of fish kills reported in the state between 1988 and 2014 (although the majority of data were recorded after 2003). Using this resource in conjunction with temperature data modeled for 10,774 Midwestern lakes (Winslow et al., 2017), I am able to build a more complete picture of how MMEs relate to climate change. I am interested in environmental stressors such as temperature and dissolved oxygen, and in infectious agents (primarily *Columnaris*).

There exist three primary cause categories of fish MMEs.

1.4.1   Anthropogenic Causes of Fish MMEs

These primarily consist of chemical spills from industrial areas or nutrient/chemical runoff from agricultural areas, although there are many other ways in which fish may be inadvertently killed by humans, such as private dumping, waste mismanagement,

Figure 1.3: The composition of MME cause types. The distribution of causes or suspected causes of reported incidents in the Wisconsin dataset. *Columnaris* may also account for a significant percentage of the unknown infections category. Dissolved oxy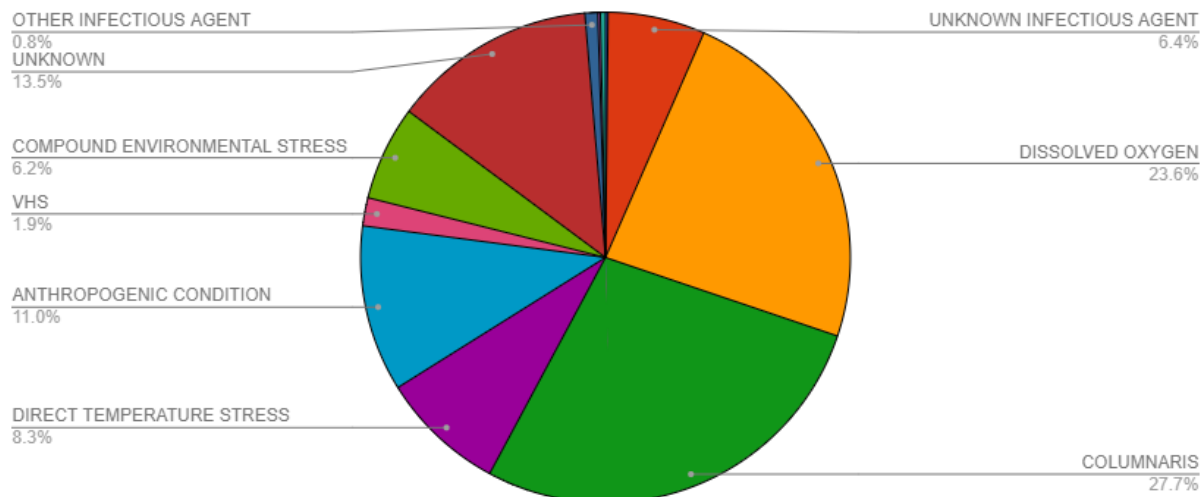gen involves either winterkill through ice induced stagnation and reduced photosynthesis, or summerkill through increased temperature or eutrophication. Some incidents attributed to direct temperature stress likely involve reduced oxygen levels. Most events involve compound environmental stressors but MMEs are only attributed to this category if there is no clear primary cause. Anthropogenic conditions involve spills, runoffs, or any disturbance primarily attributed to humans. Other infectious agents include any known pathogen that is not VHS (viral hemorrhagic septicemia) or *Columnaris*.

electrocution, overcrowding, or trapping through damming and water rerouting. Most fish deaths attributed to toxicity likely have an underlying anthropogenic cause. These types of fish kills, although in certain countries historically comprising the majority of fishkills (La et Cooke, 2011), are not the primary type of fishkill observed in Wisconsin and tend to have easily identifiable sources and solutions. Although disastrous, human caused MMEs are the least likely type of event to repeat if public awareness increases and safety measures are implemented.

1.4.2 Environmental Causes of Fish MMEs

Environmental threats to fish are multifaceted and unpredictable. As climate change intensifies they will also become more numerous (Moreno et Mller, 2011; Xenopoulos et al., 2005). In any given freshwater system global warming may lead to increased temperatures and dissolved oxygen stress, low water levels from increased evaporation, high water levels from increased rainfall and glacial melt, highly volatile water cycles from a simultaneous increase in evaporation and rainfall acidification, and/or eutrophication (Lovejoy et. Hannah, 2005). Not all environmentally linked die-offs

are necessarily climate related. In the dataset there is also a large proportion of winterkill MMEs - die-offs caused by low oxygen levels because of the inability of plants to photosynthesis and the inability of atmospheric oxygen to mix with water after prolonged ice cover (Greenbank, 1945).

1.4.3   Infectious Agents

Fish are susceptible to pathogens as water is an effective medium of transfer. Disease has always been a cause of mass mortality but its role may become even more prominent under climate change. Otherwise resilient fish populations may become susceptible to infection in light of weakening but non-lethal stressors. More troubling, some pathogens may become increasingly virulent as temperatures rise. Flavobacterium columnare, *Columnaris*, is a common bacterial pathogen affecting freshwater fish. The virulence of *Columnaris* directly increases with water temperatures, ranging from nonlethal at 9.4C up to 100% lethal at 20.5C (Holt et al., 1975). This is likely due to the improved ability of Flavobacterium columnare to adhere to the gills of fish in warmer waters (Decostere et al., 1999).
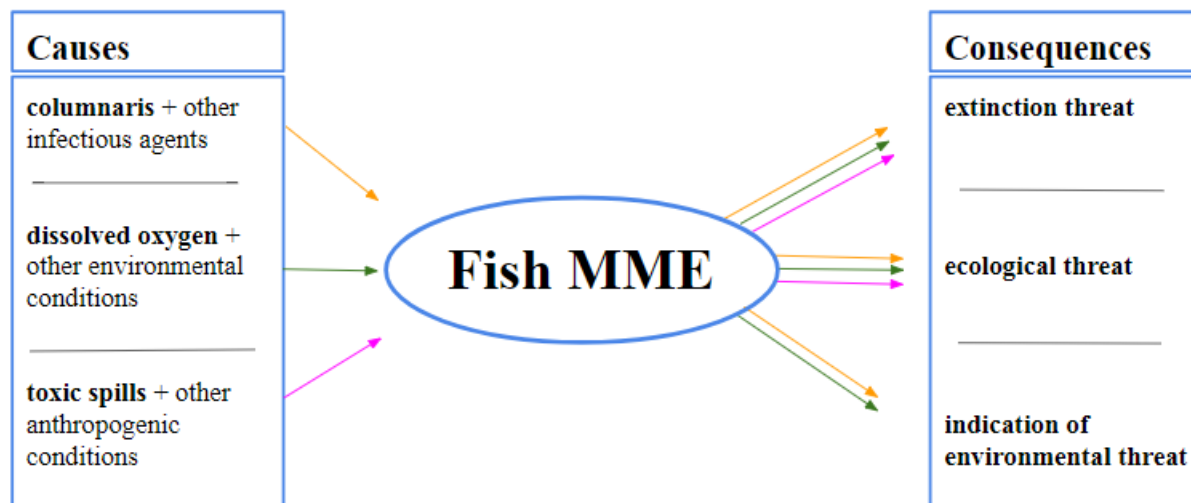
Figure 1.4: The relationship of causes and consequences of MMEs. A graphical representation of possible causes and consequences of fish mass mortality events. Determining if infectious and environmental related MMEs are signals of underlying global warming is the primary goal of this thesis.

Data on MMEs are difficult to gather as MMEs are inherently rare and unpredictable. Data on fish MMEs in particular are hard to obtain as the true magnitude or even existence of a kill is often not-apparent if corpses decay or become trapped underwater rather than wash up on shore. The Wisconsin dataset offers a valuable opportunity to study the rate of occurrence and variation in cause type of fish MMEs, and paired with extensive modeled data on Midwestern lake temperature (Winslow et al., 2017), it has the potential to shed light on an underlying pattern of fish mortality.

I hypothesize that both summertime dissolved oxygen stress (a direct result of higher temperatures) and *Columnaris* (which becomes more virulent in warmer waters) will both increase in occurrence in lakes that see warming. Additionally, I hypothesize that there will not exist a corresponding decrease in cold related die-offs such as winterkill. This is based on the observation of a 0.4C increase in temperature in equatorial zones between 1961 and 1991. Comparing the 20 year periods following each date we see up to a 76.3% increase in major heat related events but only an 8.6% decrease in cold related events (Buckley et Huey, 2016). Understanding how these types of MMEs change with temperature will help predict when MMEs are most likely to occur.

Methods

2.1 Data Tidying and Transformation

There are two primary datasets used in this thesis.

1. The MME dataset. This includes raw data on fish die-offs collected by the Wisconsin Department of Natural Resources primarily between 2004 and 2014.

2. The Thermal Metrics dataset (Winslow et al., 2017).The Thermal Metrics dataset (Winslow et al., 2017). This resource is unparalleled in scope and ambition. It includes temperatures changes in 10774 Midwestern lakes modeled for both hindcast data (1979 - 2015) based off the North American Land Data Assimilation (NLDAS). It also includes forecast data for two future periods 2042 - 2059 and 2082 - 2099 built with a global climate model from the Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM)).

The MME dataset is comprised of 592 events. These are citizen reported kills investigated by agents within the department of natural resources. Encoded information includes the start and end date of the investigation (a proxy for the time of occurrence), details on the waterbody type, the area affected, the predicted cause, the confidence in prediction, the degree of mortality (total or partial extirpation), the investigator, the enforcement action if applicable, and a wide host of qualitative descriptions. 74 of the events reported to the DNR by private citizens did not meet the qualifications of an MME (i.e, at least 25 individuals dying over the course of 48 hours from any unusual event) and were excluded from the analyses. The majority of these exclusions are investigations of individual dead fish. An additional 89 events were automatically excluded from the analyses because no thermal metrics are available for them within the Winslow et al. dataset. These events likely take place in streams or exceptionally small lakes (below the 4 hectare requirement of the Winslow et al. dataset).

I analyzed the qualitative descriptions of each mortality event to determine the species affected. Some events do not include descriptions of the involved species, but the majority specify at least one (with two specifying as many as nine different species). Additionally, I created a secondary column for cause type, using the qualitative descriptions to further distinguish between summerkills, winterkills, anthropogenic kills, and infectious agent events.

Integrating the datasets is possible by linking the WBIC location numbers in the MME dataset with the site ID numbers in the Thermal Metrics dataset. Additionally, I used a DNR dataset of Wisconsin lakes to slim the thermal metrics dataset down to only Wisconsin lakes, and to incorporate raw statistics on lake size and depth. When merging datasets in R, I first created a limited data set (all = FALSE) of just the mortality events with thermal descriptions, and then re-merged the MME dataset with the thermal metrics dataset (all = TRUE) to incorporate the control lakes. In this way I created a dataset with each row equivalent to a year at a particular site. Every row contains thermal data but only a small fraction of rows contain MME data (these are the sites that have experienced an event). Because MMEs are highly dependent on their month of occurrence I gathered the dataset with the TidyR package so that each observation was a month of a particular year at a particular site.

2.2   Data Visualization

There are numerous interesting questions to ask when presented with this data, and multiple ways to visualize the answers. I attempt to explore MMEs and their relationship with temperature change in a holistic, unbiased manner.

2.2.1   General Exploration

There are 2,235 Wisconsin lakes with temperature data used in this study (many were filtered out as they did not match with the DNR lake dataset being either outside Wisconsin or too small). For these 2,235 lakes there are data on size and depth (in most cases), surface and bottom monthly maximum and means, stratification (in most cases), and ice duration. Of these 2,235 lakes 258 saw an MME at some point between the year 2004 and 2014 (see figure 2.1).
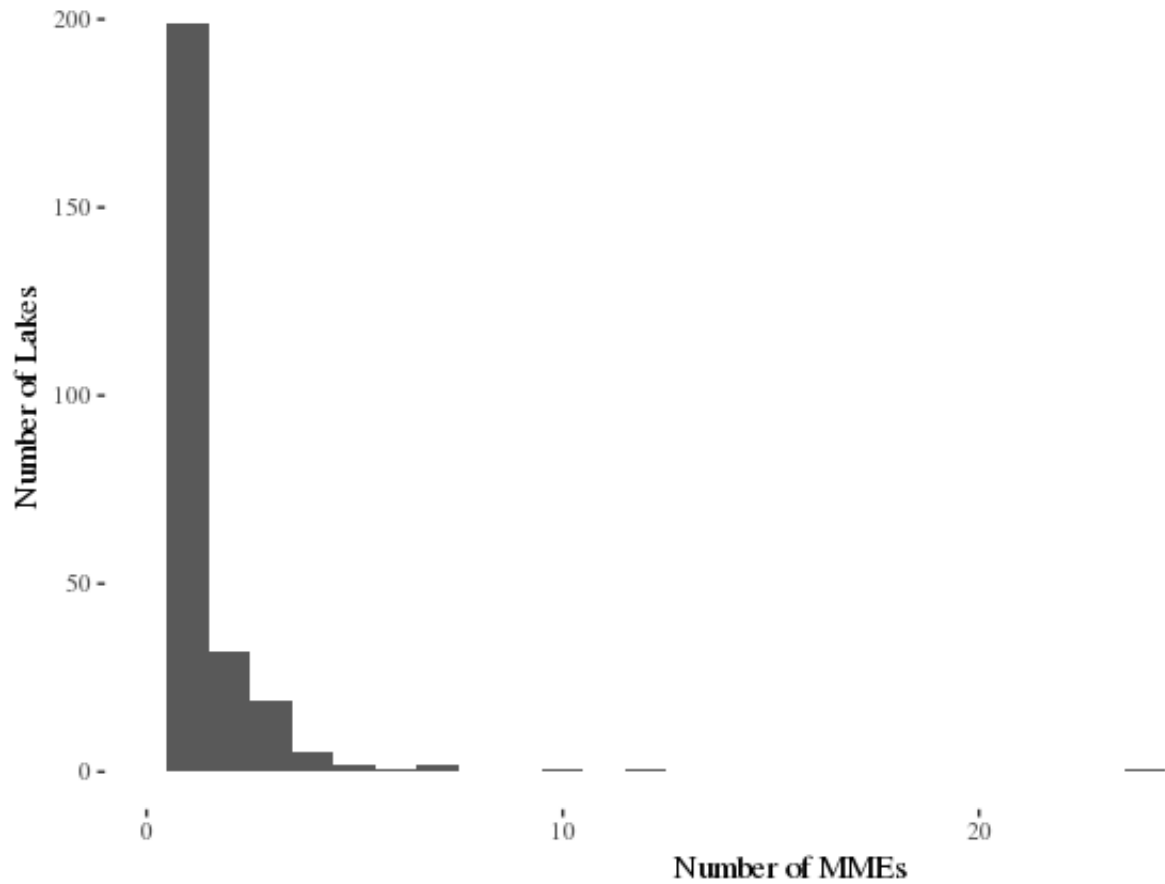
Figure 2.1: The distribution of MME frequency by lake. The distribution of MME frequency across lakes with events. The majority of lakes that saw an MME experienced only a single event during the decade. However, there is a long right tail with one lake experiencing up to thirty events.

The majority of MME data are limited to a ten year span from 2004-2014 (entire data range illustrated below, data before 2004 is removed in all other sections). Examining the consistency of the data across years is important for determining whether a year to year analysis will be possible. If year to year data collection methods were consistent than it would be useful to examine the relationship between MME frequency and yearly temperature spikes, however, this is not the case.

An alternative method of determining whether temperature affects MMEs is to compare the number of events across months (illustrated on bottom of figure 2.2). Doing this shows that there is a strong correlation between early summer events and MMEs.
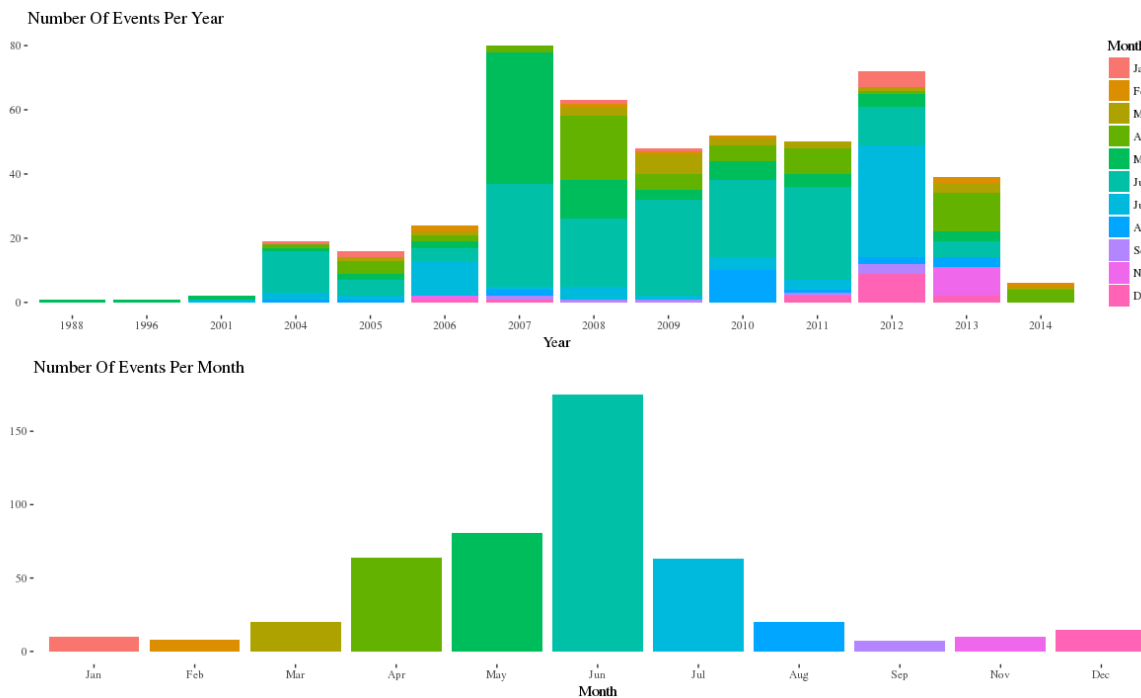
Figure 2.2: MME occurrence by year and month. The number of MMEs per year (top) colored by the month in which they occurred, and the number of MMEs per month (bottom). This reveals the pulse of MMEs in the early summer with the majority of events occurring in June.

Areas of potential error are revealed by figure 2.2: although it is certainly possible that the middle years (2007 - 2013) had conditions more favorable for MMEs, it is not likely that mortality events were cataloged completely for every year and the skew may be a result of increased documentation. For instance, the dataset only includes observations in 2014 from January through May which almost certainly does not reflect the occurrence of MMEs in that year. Another potential source of bias is natural rather than human induced. Because many winter mortality events occur beneath the ice layer, they are not discovered until months later when the ice begins to melt. This likely accounts for the influx of reports in May and April and the observed discrepancy between MME counts in the spring and fall.

There are many temperature metrics cataloged in the thermal metric dataset (Winslow et al., 2017). The four of interest are maximum surface temperature, mean surface temperature, maximum bottom temperature, and mean bottom temperature.

Most MMEs occur in summer months and the average maximum temperature is significantly higher for lakes with MMEs (see figure 2.2 and table 2.1). Given this, is it possible that MMEs are caused by temperature spikes? One way to determine this is to compare the average yearly peak temperatures of lakes that did not have an event to the average yearly peak temperature of lakes that did have an event (figure 2.3).

(No-event) - (MME) = Difference	Maximum Temperature	Mean Temperature
Surface	(14.0) - (21.2) = -7.2°C	(10.4) - (16.9) = -6.5°C
Bottom	(10.3) - (16.6) = -6.3°C	(8.6) - (14.0) = -5.4°C

Table 2.1: Temperature metrics. The average temperature for all months where MMEs did not occur minus the average temperature for months where they did occur. The 4x4 matrix represents the four metrics for thermal evaluation, maximum surface temperature by month, mean surface temperature by month, maximum bottom temperature by month, and mean bottom temperature by month. The greatest thermal separation of non-event and MME lakes is observed in the maximum temperature metric, suggesting that some MMEs may be a result of abnormal temperature spikes.

It is important to note that table 2.1 does not control for the summer-weighted distribution of MMEs which drives up the average temperature in all categories. Although MMEs do occur more often at hotter times of the year, table 2.1 does not necessarily indicate that they occur in the hottest lakes.

Figure 2.3 (follwoing page) includes all types of MMEs regardless of month of occurrence, including types that wouldn't be expected to have heat related causes such as winterkill events. For this reason it is important to also consider MMEs at a higher resolution in order to determine what kill types may become more likely at abnormal peak temperatures.
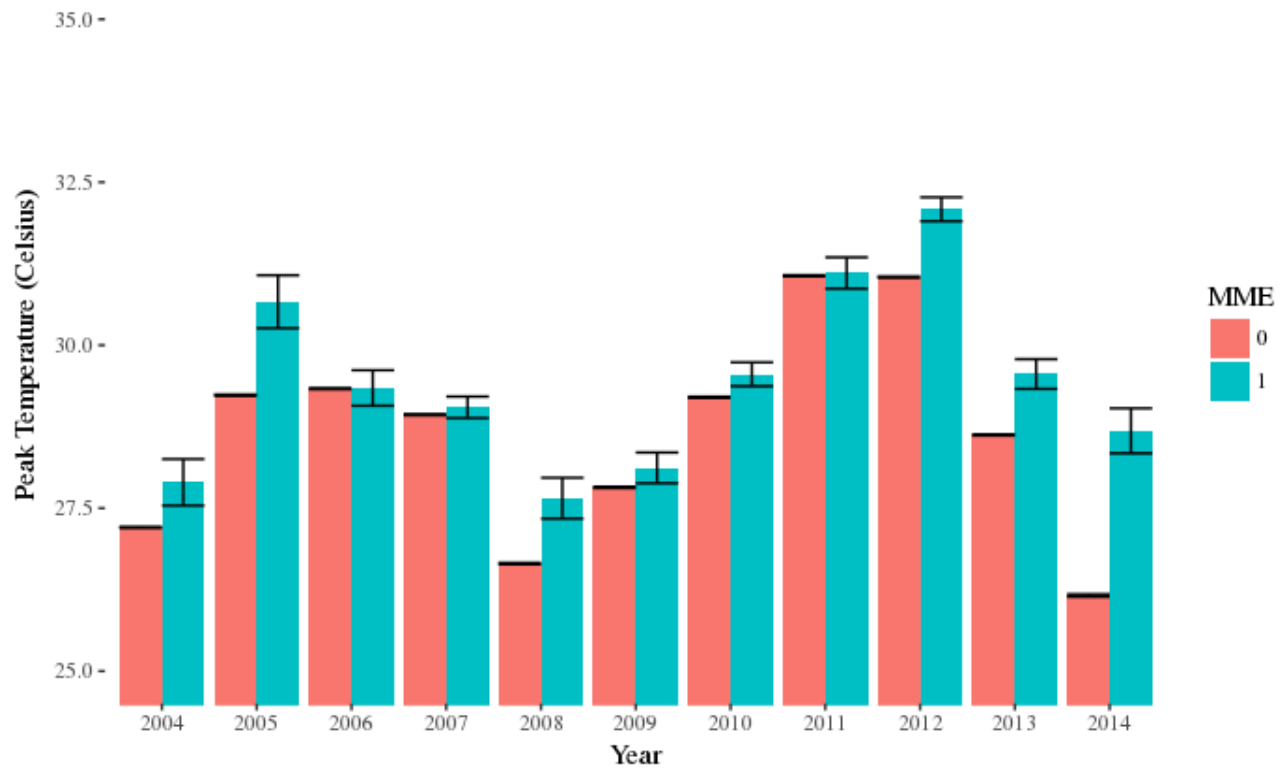
Figure 2.3: Peak temperature by year for MMEs and non-events. The average peak temperature for all months in every lake, grouped by year and MME occurrence. Lakes without MMEs are shown in orange and lakes with MMEs are shown in teal. Error bars (+/- 1 S.E.) are larger for lakes with MMEs due to the smaller sample size. As this figure only considers peak temperature by year and not by month we do not see the discrepancy between MME and non-MME that is indicated in table 2.1. Instead, the mean peak temperature for all events is 29.6°C and for all non-events it is 28.2°C, a difference of 1.4°C. This is a significant difference (for an ANOVA of the difference in temperature between MME and non-event lakes F = 53.63 and p <0.001).
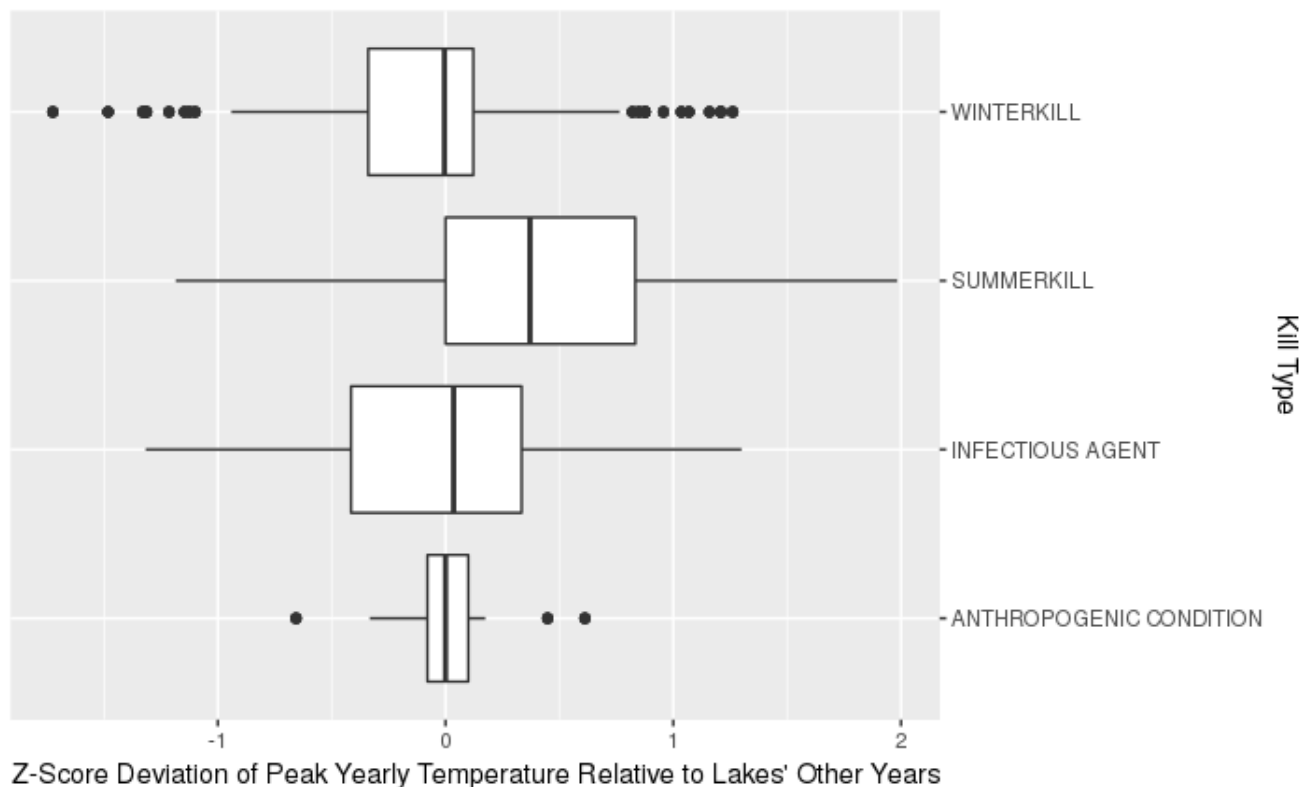
Figure 2.4: Yearly peak zscore by killtype. The different kinds of MMEs by Z-score (peak temperature during the year of an MME - mean peak yearly temperature for the given lake / standard deviation). The type of MME is classified based off qualitative descriptions of the kills, investigators personal assessments, and month of kill occurrence. Winterkill and summerkill may refer to MMEs with unknown causes that occurred during the winter (Oct - May) or summer (Jun - Sep) months and were not obviously caused by humans or pathogens. A Z-score of 0 implies that, for that mass mortality event, the peak temperature in that year was equivalent to the expected peak yearly temperature of that lake. This histogram shows that lakes with certain types of MMEs have experienced abnormally hot peak temperatures (particularly those with summerkills). Our control, lakes with anthropogenically caused MMEs, have an effective Z-score of zero implying that they do not tend to have above average peak temperatures. An ANOVA reveals that there is larger significant difference between groups than within groups (F = 14.555, p <0.001). A post-hoc Tukey reveals that there is a significant difference between summerkill and winterkill (p <0.001), summerkill and infectious agents (p <0.001), and possibly between summerkill and anthropogenic kills (p = 0.056). Other kill types were not significantly different from each other in regards to peak yearly temperature.

2.2.2   Summerkill Exploratory Analysis

Any non-human/non-pathogen mortality event occurring between June and September (when most Wisconsin lakes are finally ice free) is grouped in the all encompassing category of Summerkill. The majority of these events involve multiple abiotic stressors, although the most common cause is dissolved oxygen depreciation either from rising water temperatures or eutrophication. One problem is that different fish species have different physiological attributes such as oxygen requirement (Moore, 1942). A critical thermal maximum is difficult to determine due to the differing adaptive capacities between stenothermal and eurythermal species. Additionally, due to variation in size and stratification between different lakes, max surface temperature is not always the best proxy for lake heat lethality. Instead, consistent, higher average temperatures may have a greater likelihood of permeating all levels of the lake. Furthermore, extreme max temperatures are partially captured in mean temperature values. For this reason mean temperature rather than max temperature may have a more significant effect on summerkill probability in the logistic regression.
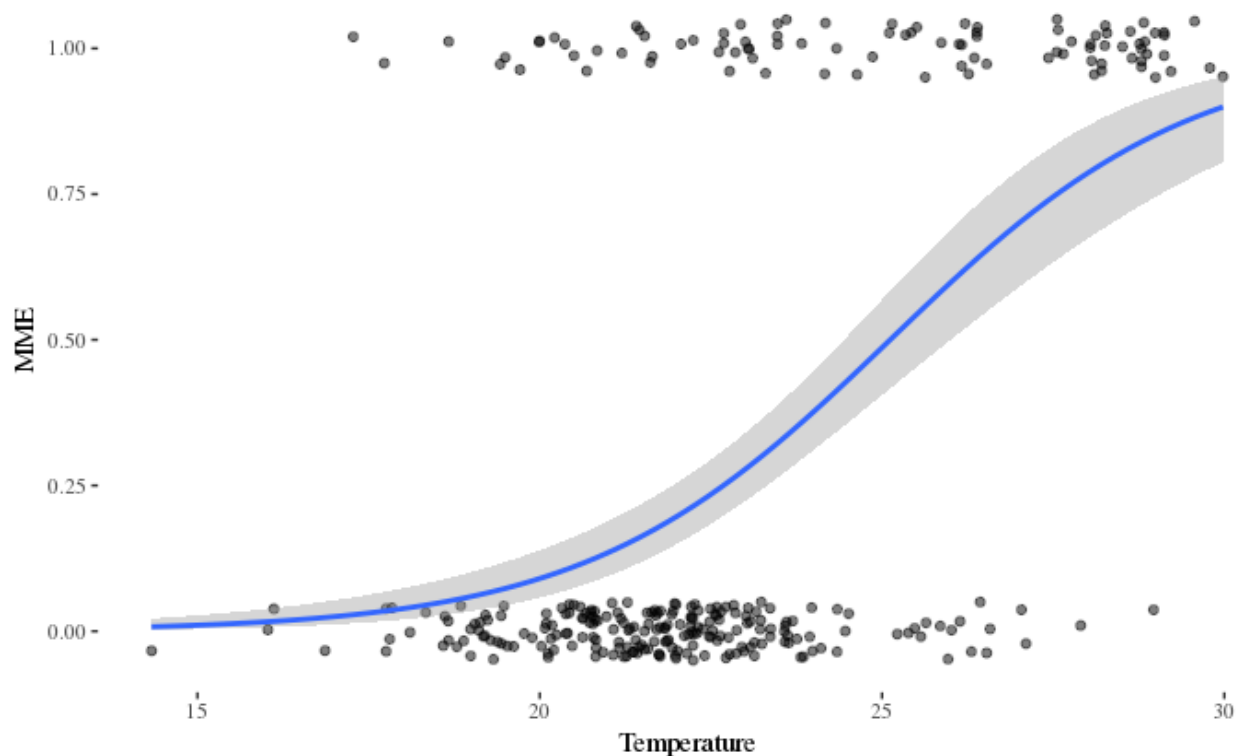
Figure 2.5: Logistic regression of summerkill by temperature. The mean surface temperature for the months of June, July, August, and September for lakes with events (94 datapoints) against a random sample (set.seed = 1234) of 500 lakes without events (59 summertime non-events). This sample has a beta coefficient of 0.45114 and p <0.001 for a logistic regression of temperature by MME. This is similar to the logistic regression performed on all lakes where the beta coefficient is 0.41933 and p <0.001.

Figure 2.5 shows that there is a significant relationship between mean surface temperature and the probability of a summerkill MME. Running a logistic regression on all summer events (including human and infectious agent kills) reveals a less robust but still significant trend (beta coefficient of 0.02136 and p <0.001). This shows that there is a relationship between hotter max temperatures in the summer months and the occurrence of some types of MMEs.

As temperatures increase it is feasible to expect more of these heat related events. It is unclear however, how other types of MMEs change with climate change. Will we see a reduction in winter kills? Are pathogens at all temperature dependent? In order to determine how the pattern of MMEs will change in the coming century it is necessary to investigate other killtypes.

2.2.3 Winterkill Exploratory Analysis

One of the most disruptive climatic variables for freshwater fish is ice cover. When lakes become frozen-over it is difficult for light to reach aquatic autotrophs below the surface and photosynthesis is halted. Additionally, ice prevents atmospheric gas exchange with the water surface. Without the influx of oxygen from photosynthesis respiration may quickly deplete the dissolved oxygen in the lake depending on the size of the system and the density of aquatic life within (Greenbank, 1945).

Winterkill cannot be predicted by either temperature or ice duration (see figure 2.6). One explanation for why there does not appear to be a direct relationship between these thermal metrics and winterkill may be localized variability. Wisconsin species inhabiting lakes with annual winterkill conditions have demonstrated behavioral adaptations to winter hypoxia (Magnuson et al., 1984; Petrosky et al., 1973). Unlike climate induced heat stress, lake freeze-overs are a historic phenomena that have selected on the morphology and behavior of northern fish. Unless ice duration increases beyond its historical intensity it is unlikely that we will see an increase in winterkill. If ice duration decreases, as it is expected to do, it will likely need to exceed the currently modeled minimum number of ice duration days before we will see any significant reduction in MMEs. It is not implausible however, that this could happen (see figures 1.7 and 1.8).
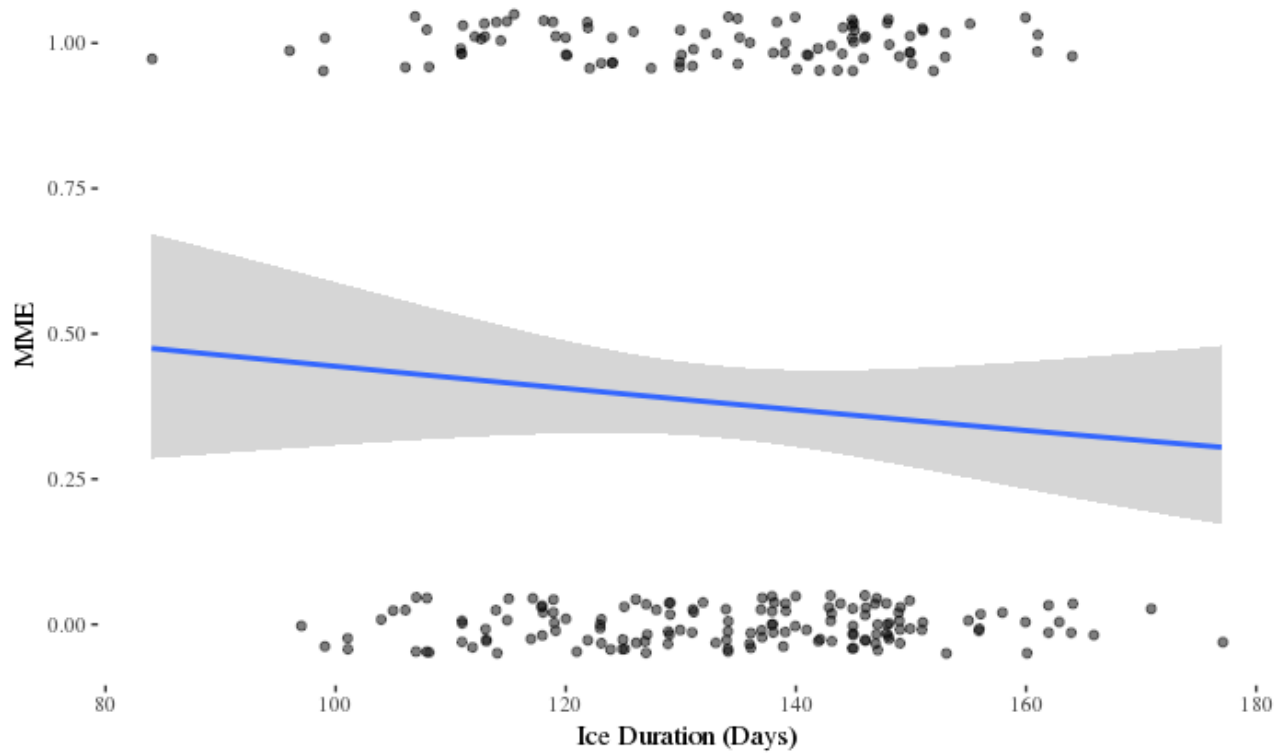
Figure 2.6: The effect of ice duration on MME occurrence in lakes that were or were not expected to have had a winterkill between the months of October and May. All 90 lake-years that experienced a suspected winterkill MME were paired with a random sample of 500 non-event lake-months (146 individual lake-years) for this visualization. The beta coefficient for ice duration of this visualization is -0.00785 with an insignificant $p = 0.334$. When using the full control dataset the coefficient increases to -0.01702 and $p = 0.00447$. This is contrary to our expected result and may reflect a correlation with a third, undetermined variable. Ice duration does not effectively predict winterkill.
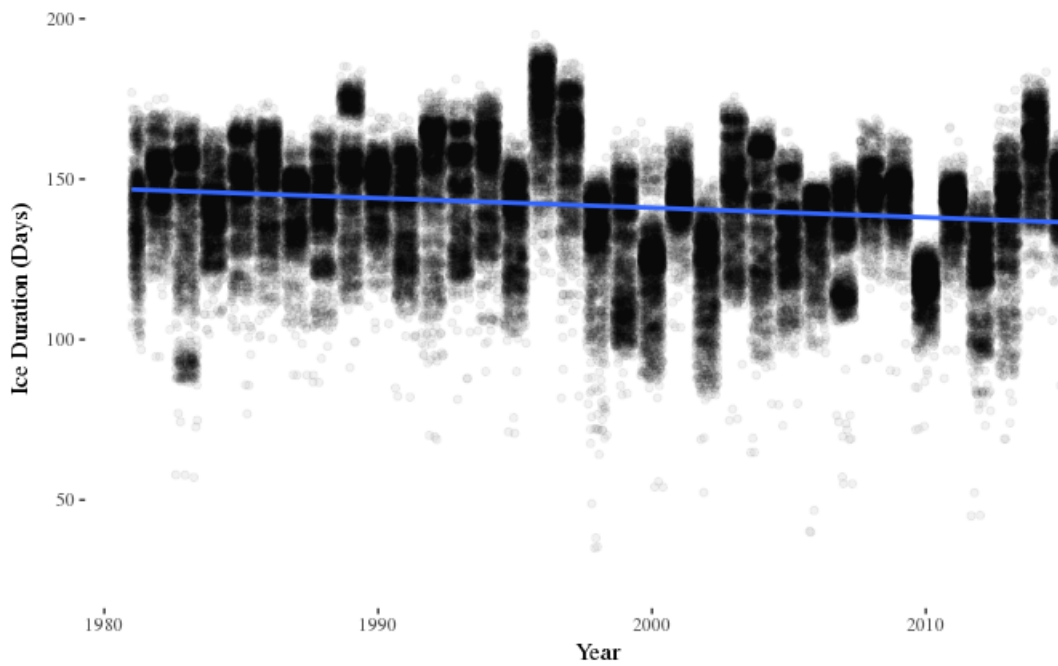
Figure 2.7: Trend in contemporary ice duration. A linear regression line of best fit for ice duration and year. The data are from the entire range of historic thermal metric modeling, a point represents a particular lake at a particular year. This indicates that there is in fact a negative trend towards length of ice duration, beta coefficient = 0.3619 and p <0.001.
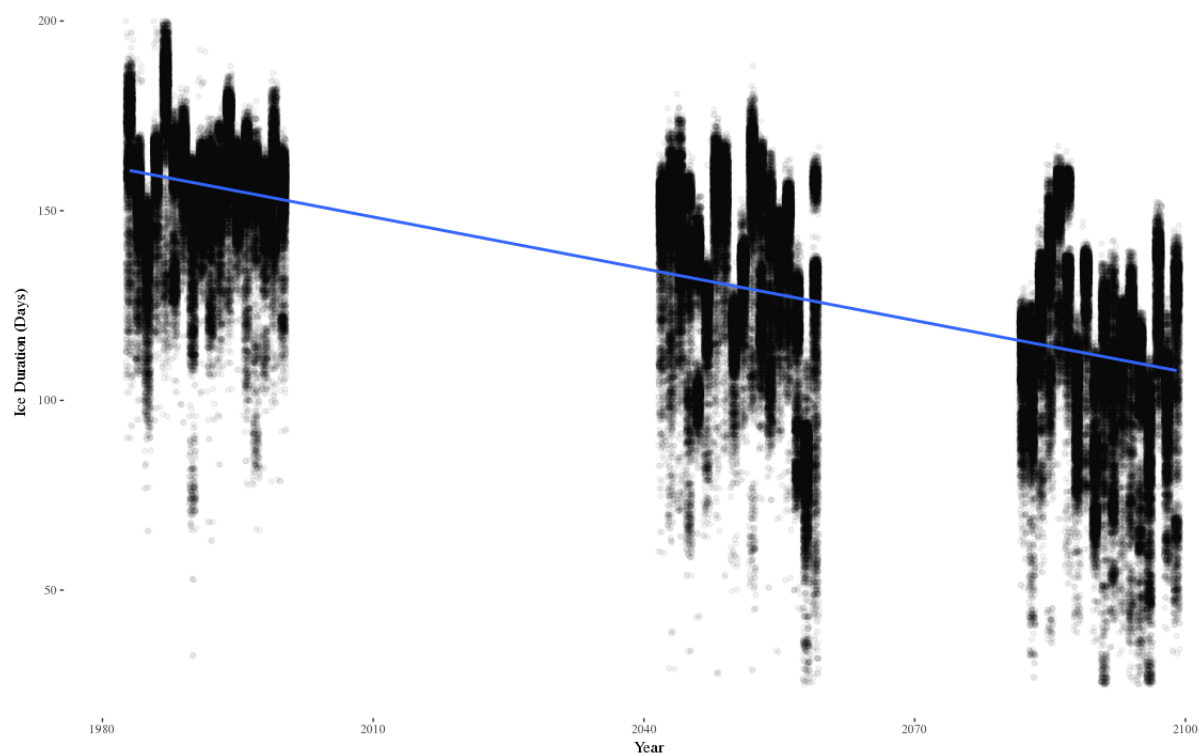
Figure 2.8: Trend in future ice duration. A linear regression line of best fit for ice duration and year including future modeled data for the periods 2042-2059 and 2082-2099. Future modeled metrics (Winslow et al., 2017) based on temperature data from the Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM)). A point represents a particular lake at a particular year. This confirms the negative trend of ice duration and indicates that there may be several lakes that experience below the minimum number of ice duration days in the contemporary extrapolated data. It is impossible to tell if these lakes will see a reduction in winterkill probability.

2.2.4   Infectious Agent Exploratory Analysis

It is not implausible that there is a relationship between temperature and waterborne pathogens. Thermal stress may weaken fish immunities and warmer waters may increase the virulence of some bacteria such as Columaris (Holt et al., 1975). Of the 193 fish kills suspected to be caused by infectious agents approximately 157 display symptoms of *Columnaris*. The remaining kills are primarily VHS (viral hemorrhagic septicemia).

Because the ability of *Columnaris* to adhere to the gills of fish is linearly proportional to the temperature of the water (Decostere et al., 1999) mean temperature is likely the best predictor of infectious agent kills. Additionally, pathogenic kills occur year round and so comparing them to non-events without controlling for the month of occurrence will result in a bias due to the summer-weighted distribution of pathogenic kills. Because of this it is necessary to visualize temperature as a Z-score deviation.

Z-score temperature deviations do not appear to affect the number infectious agent kills. Additionally, non-adjusted monthly surface temperature does not have a significant effect on infectious agent kills when filtering by season. One possible explanation for this is that the temperature threshold for lethal *Columnaris* outbreaks is very low, just 20.5C (Holt et al., 1975). The distribution of *Columnaris* may be a far more important and obscuring factor. Assuming that the temperature of a given lake is at least 20.5C it will likely see a mortality event around the time that it is exposed to the pathogen. After the event low populations may leave it less likely to experience an event for the remainder of the year, regardless of how hot the temperature becomes.

This is not to say that *Columnaris* will not become more frequent under climate change, it may still benefit as it will have more habitats reaching thermal optimum earlier in the year. However, it is not likely that there is an accurate way to predict pathogenic kills from temperature data without being able to control for bacteria distribution and rates of transmission.

2.2.5   Predictive Variable Exploration

There are a wide range of predictive variables available in the datasets for use in MME classification. The relationship between the various predictors and the model outcome are likely complex and non-linear owing to the threshold nature of the response variable. Because of this a tree-based approach may be more accurate for the final modeling component.

Tree-based approaches to data modeling work by binning the data at various decision nodes and using a greedy algorithm to select the variable with the greatest separation in classification result. For a classification problem this results in a tree of nodes with predictive variables at every split and the values of the response variable along the bottom at the conclusion of every path.

Although tree-base methods are easy to understand they suffer from a high degree of variability depending on the order at which the nodes are explored. To account for this and increase the predictive power many variants of the tree based approach
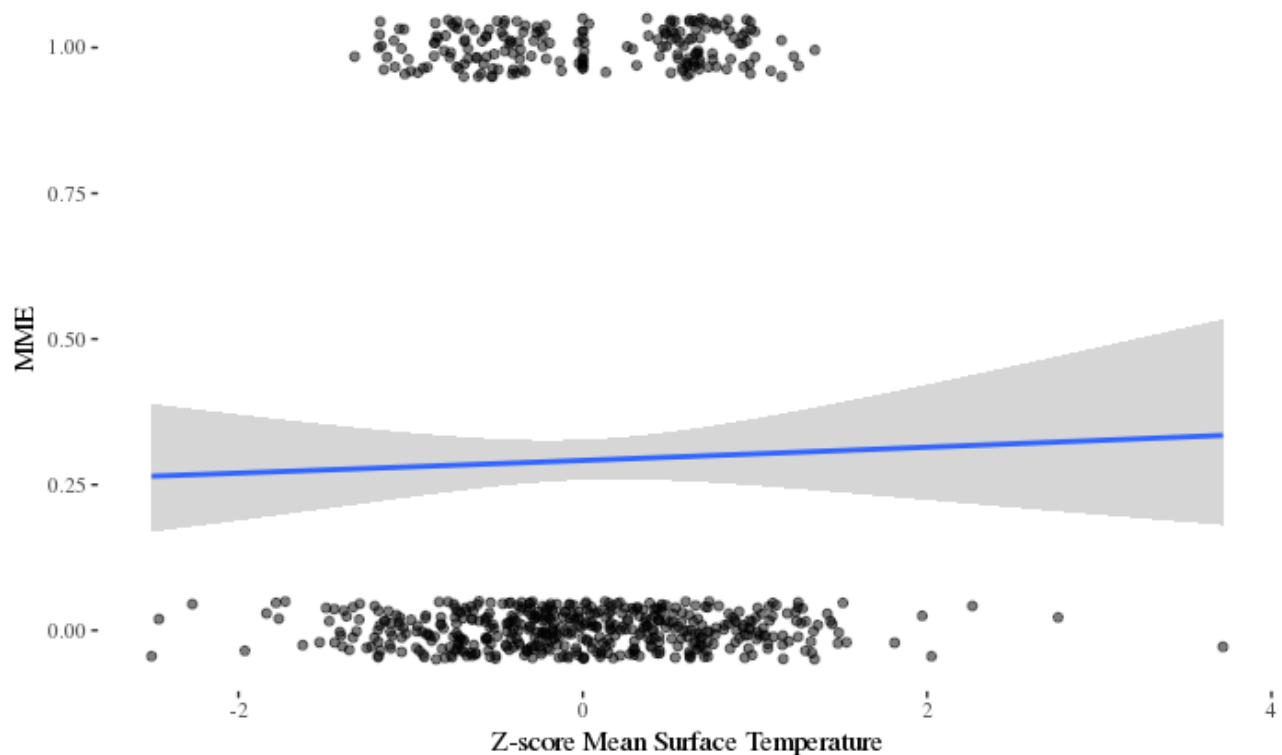
Figure 2.9: Logistic regression of infectious agent kills by temperature. A regression of MME and Z-score mean surface temperature. The Z-score calculates how different the temperature in a given lake at a given month is from the average temperature of that month in that lake. Control lakes are taken from a random sample of 500 lakes. This graph has a beta coefficient for temperature of 0.05381 and an insignificant p = 0.627. Running a logistic regression with the full control dataset reveals a beta coefficient of 0.1784 and similar insignificant p = 0.843. The beta coefficients of both models are exceeded by their standard error. Temperature deviance is not an effective predictor of infectious agent kills.

employ bagging. This process involves building multiple trees with replacement. One type of classification tree with bagging is the random forest method (Ho, 1995).

Random forest modeling is unique from other types of tree-based modeling because at each node, rather than selecting from the entire range of predictive variables, only a fraction of them are considered. This is useful in scenarios where there is a high degree of asymmetry in the importance of the predictive variables. If the model considered every variable at every node it would overwhelmingly favor the most predictive (which in our example is temperature). By only considering a fraction of the variables the random forest model is forced to build a more diverse forest, possibly increasing the predictive potential of the final product (James et al., 2013).

Because random forest models sample only a subset of the variables at each node they can account for high collinearity between variables in a way that logistic regression cannot (James et al., 2013). This is potentially useful in this dataset as the

majority of predictors used are variations of temperature (and likely highly correlated).

As versatile as random forests are, overfitting is still a risk. Overfitting occurs because there is no measurable penalty in adding predictors (they will either have a negative or positive affect on the OOB error estimation), and so predictive models may accidentally be tailored too closely to the training data rendering them ungeneralizable. To account for this, the random forest model should employ a conservative number of predictors and should be validated with misclassification tests.

One of the great abilities of random forest is its use in feature selection. Rather than maximizing predictive value, models can be built to compare different predictors by tracking accuracy contributions during the modeling process. When using random forest in this way correlation should be avoided as it can interfere with the reliability of the accuracy metrics.

When using random forest variable importance plots for feature selection (see figure 2.10), we see that one of the most important factors in determining whether a lake will have an MME at any given time is whether it or similarly numbered lakes have had MMEs in the past (WBIC identification number). Although there are certainly lakes that are more prone to MMEs, this variable is one of the most susceptible to data collection bias. If a lake is populated and has the infrastructure for MME investigation it is far more likely to end up in the dataset, therefore as a metric for predicting fish moralities I will exclude lake WBIC number.

Another promising candidate for prediction is Schmidt stratification values (see figure 2.10). This is because their accuracy and Gini scores are high and relatively aligned. Peak temperature may also have true predictive value (see figure 2.10). Cumulative days above 0 (and by extension its correlates, above 5 and above 10) should not be included in the predictive model due to the low accuracy and questionability of Gini score (see figure 2.10). Variance after ice for 30 days (and 60 days), and ice duration are possibly not worth the risk of overfitting to warrant inclusion due to their respective mismatch in accuracy and Gini scores or low total accuracy and Gini scores (see figure 2.10).
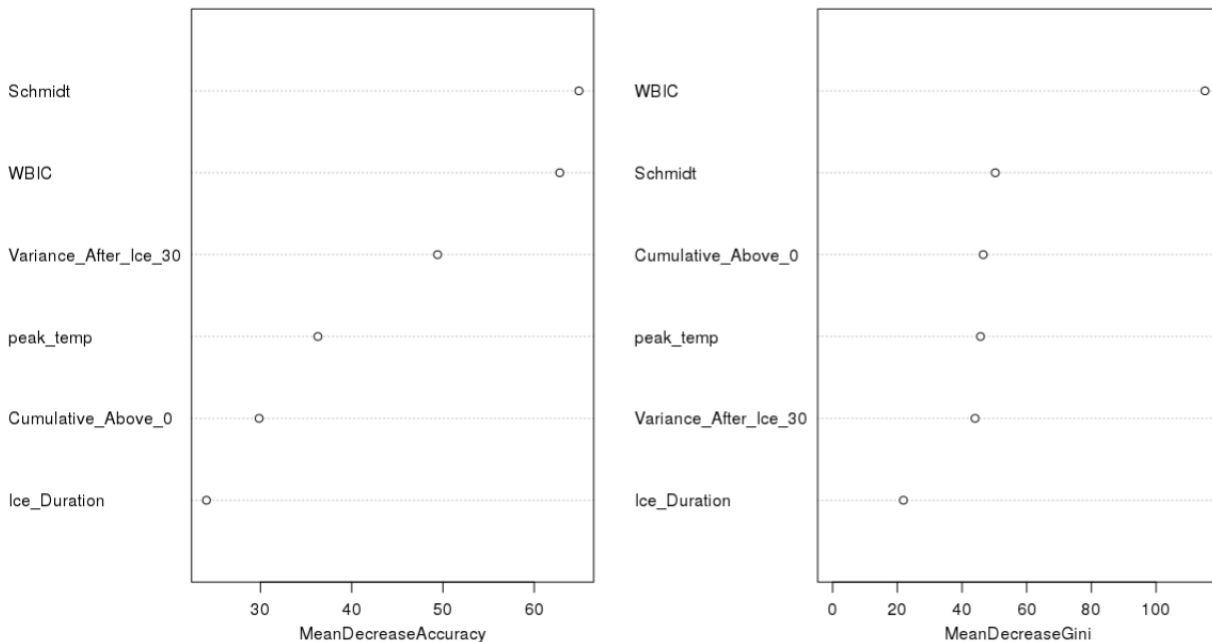
Figure 2.10: The variable plot for the random forest output with mostly non-correlated variables that affect lakes on the yearly scale. This random forest has an OOB estimation of error at 0.15%. Mean decrease in accuracy measures how much the models overall predictive power will decrease if that variable is removed. Mean decrease in Gini measures how much each variable contributes on average to the homogeneity of each split in the decision tree. A larger Gini score implies that at any given node that variable is better at segregating the data between MME and non-MME points. Ice_Duration measure the number of days with ice cover. WBIC is the lake identification number. Cumulative_Above_0 is the sum of the days above 0C for the entire year (above 5°C or 10°C not included due to collinearity). Variance_After_Ice_30 is a measure of the variability of surface temperature 0 - 30 after ice off (30 to 60 days after not included due to collinearity). Schmidt is the sum of daily Schmidt stability values (a measure of resistance to mixing due to the stratification of the water). Peak_Temp measure the max temperature observed in that year. The reason that mean decrease in Gini and mean decrease in accuracy may not match in their relative positions is because despite mitigation efforts there is still correlation between some of these predictors. Correlation can render Gini scores inaccurate and/or inflate the mean accuracy of insignificant predictors.
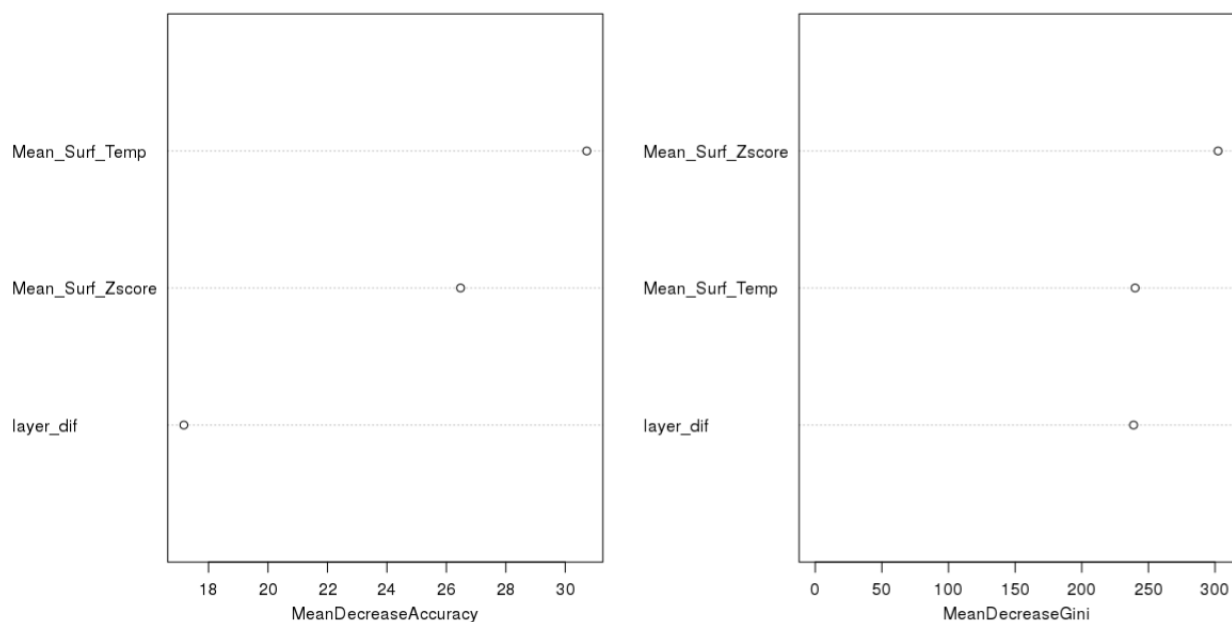
Figure 2.11: Random forest variable importance plot - monthly variables. Random forest variable importance plot - monthly variables. The variable plot for the random forest output with mostly non-correlated variables that affect lakes on the monthly scale. This random forest has an OOB estimation of error at 0.16%. Mean decrease in accuracy measures how much the models overall predictive power will decrease if that variable is removed. Mean decrease in Gini measures how each variable contributes on average to the homogeneity of each split in the decision tree. A larger Gini score implies that at any given node that variable is better at segregating the data between MME and non-MME points. Mean_Surf_Temp or Mean_Surf_Zscore measure the temperature or temperature deviance at the lake bottom (max and bottom temperatures were not included in this random forest due to collinearity). For any given month the difference from the normal temperature (Z-score) in that given month is not heavily correlated with mean or max temperature because the relative scale for Z-scores equalizes hotter and colder parts of the year. layer_dif is the mean difference between mean surface and mean temperatures for a lake at any given month.

It is not clear whether mean surface temperature or Z-score is a more accurate predictor of fish mortality events (see figure 2.11). The accuracy of this variable importance plot may be clouded by correlation between the two factors. Because adding them both to the final predictive model may be redundant, if the model suffers from overfitting only the more interpretable metric, mean surface temperature, should be used. Although random forests with limited predictor sampling at each node accommodate correlation to a degree, it is still better to exclude correlated variables that are not likely to add additional resolution to the model. For this reason all max temperature values will be excluded as their significance can be accounted for with means (as high maximums drive up the mean). For mean temperatures the difference between surface and bottom can be accounted for in the layer difference rather than including both surface and bottom temperatures in the model. Layer difference may be useful in our logistic regression as it has monthly predictive power and is not likely to correlate with temperature, although its degree of predictive power does not seem overwhelming (see figure 2.11). A quadratic term for mean surface temperature was considered to account for possible negative effect of temperature on winterkill MMEs but did not prove significant in the logistic regression and was never introduced in the random forests because when evaluating data sequentially interaction terms are not necessary. Month is not used in the final random forest model because it may suffer from reporting bias and is a high risk candidate for overfitting. Additionally, most information it encodes is captured by temperature differences.

There are other variables available in the datasets that are excluded from the final model for varying reasons. Although they are possibly significant factors, lake size, lake depth, lake stratification, and lake stratification depth cannot be used as many lakes are missing data on these features. Year cannot be used as by definition there is no overlap between the history being modeled and the future being forecasted.

Results and Discussion

Any predictive model applied to these data will perform with great accuracy by ignoring MMEs and consistently predicting zero. This is because the data used in this thesis is highly unbalanced with true positives comprising only 0.16% of the observations. For this reason it is impossible to say whether an MME will occur at any given lake during a specified time because probabilistically, it wont. We can however, aggregate the slim probabilities across all lakes and develop accurate estimates of the number of MMEs we would expect to see in total, and in any given year.

3.1   Estimating the Total Number of MMEs

To predict the number of MMEs we expect to see in Wisconsin three models will be considered, a logistic regression (GLM), a random forest with multiple variables (RF), and a simple classification forest with only one variable (CF).

The most accurate model is likely the logistic regression (GLM) as it performs the best when estimating the number of MMEs between 2004 and 2014 (see table 3.1). The classification forest (CF) built exclusively on temperature also makes feasible predictions on the lake history and lake future datasets. The complex random forest model (RF), although it predicts fairly accurately in the lake history dataset, predicts only 113 MMEs in the future dataset. This discrepancy in RFs ability to estimate future MMEs is strong evidence of overfitting. The random forest model is likely grossly underestimating the number of MMEs we will expect to see in the two periods 2042 - 2059 and 2082 - 2099.

There are many factors not considered by these analyses that may play into the likelihood of an MME occurring. Because of this, the individual lake-months predicted by this model may likely not experience MMEs, and, more probable given the conservative nature of MME reporting, many lake-months predicted to have no event may experience a die-off. However, there is no conceivable reason to believe that the GLM or CF model would be overestimating the occurrence of MMEs. On the contrary, there is only concern that they may, like the RF model, be underestimating the occurrence of MMEs by overfitting to similarities in the training and testing data that are not present in the future data. Assuming that the future temperature metrics are accurate (Winslow et al.,) and other contributing factors do not change in the next century, these two feasible models (GLM and CF) provide sophisticated estimates of the number of events we will see between the two periods 2042 - 2059 and 2082 - 2099.

	MMEs Observed	GLM Prediction	CF Prediction	RF Prediction
Lake History (2004-2014)	424	436	329	329
Lake Future (2042-2059 and 2082-2099)	*NA*	1,836	2,296	113

Table 3.1: Performance of multiple predictive models. A simple logistic regression model compared to two different tree-based classification methods. All models used the same random starting pattern (set.seed = 1234). For the simple logistic regression (GLM) variables Mean_Surf_Temp and layer_dif are used (misclassification rate = 0.2%). For the random forest model (RF) the variables used are Mean_Surf_Temp, Mean_Surf_Zscore, Peak_Temp, layer_dif and Schmidt (ntree = 500 , mtry = 2, OOB error = 0.18%). For the classification forest (CF) the variable used is Mean_Surf_Temp, (ntree = 500, OOB error = 0.28%). This table is examining both the predictive ability on training data as well as on future data (using metrics from the Commonwealth Scientific and Industrial Research Organization (CSIRO) and the Bureau of Meteorology (BOM)).

GLM estimates an average of 51 events per year and CF estimates an average of 64 MMEs per year. In the observational dataset (2004 - 2014) we observe an average of only 42 MMEs per year.

3.2 Forecasting MMEs at The Yearly Level

3.2.1 Logistic Regression

Forecasting with logistic regression does not provide accurate lake by lake predictions of where events are likely to occur. Instead, it randomly samples each lake based on its probability of having an event and uses this to build an estimation of the number of total affected lakes. At the monthly level, the sample size is large enough for reasonably confident estimations of the number of MMEs we should expect to see.

The same variables from the logistic regression in table 3.1 are adopted (mean surface temperature and layer difference) as these variables are safely non-correlated predictors. When training this model on a 1000 unit sample (set.seed = 1234) of a subset of the data and retesting it on the non-sampled subset of the data the misclassification rate was 0.2% which is a good indicator of its adequacy at extrapolating on the future dataset. Using this model 1,836 future mass mortality events are predicted.
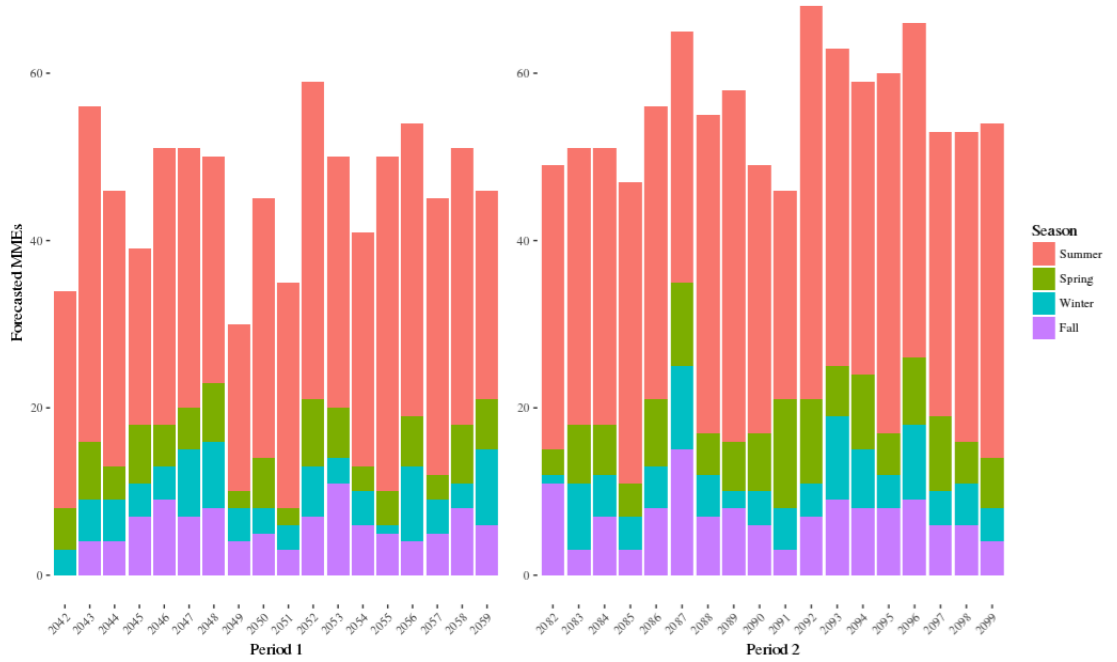
Figure 3.1: Logistic regression MME forecasting. The number of MMEs per year per season predicted by the logistic regression (set seed = 1234). This shows a moderate trend towards increasing MME frequency. Summer months (orange) are June through September, fall months (purple) are October and November, winter months (teal) are January through March, and spring months (green) are May and April. This model estimates 833 MMEs between 2042 and 2059 (49 per year), and 1003 MMEs between 2082 and 2099 (59 per year). This model forecasts a total number of 1836 MMEs (51 per year) and a maximum of 68 MMEs in 1992.

3.2.2   Classification Forest

The CF model is capable of predicting a feasibly realistic number of MMEs (see table 2.1). We can build a dataset of at risk lakes from this super conservative classification tree model (looking at mean surface temperature alone). Like the probability based estimates in the logistic regression forecasting, estimates with the classification forest rely on a high degree of random chance. This means that no lake can be categorized at the individual level.

This model is designed conservatively in order to minimize overfitting. However, even models with only one variable are capable of overfitting so to be confident in our result we must perform a misclassification test as we did with the logistic regression. Fotunately, the classification forest is not overfitting to any serious degree as the overall OOB error rate when practicing on the test subset of the data (0.32 %) does not differ greatly from the OOB error of the training subset of the data (0.28 %) (see table 3.2).

OOB (out of bag) estimation of error is only 0.28% for the model when classifying historic data, although it has a 95.99% chance of misclassifying an MME as a false

Table 3.2: Confusion matrix of classification forest. CF confusion matrix results
when separating historical data into training and testing segments. Variable used
is Mean_Surf_Temp, ntree = 500, set.seed = 1234. Rows are the true values in the
test dataset (1 for lake-months with an MME, 0 for those without). Columns are
the values predicted by the random forest based on the training data. The OOB
error rate is 0.32%, meaning that any given lake-month has a 99.68% chance of being
correctly identified.

MME predicted → MME occurred ↓	False	True	Class error
False	134,664	212	0.001571814
True	223	3	0.986725664

negative. It only has a 0.13% chance of misclassifying a non-event lake as a false
positive, however, due the drastic imbalance in the data the MMEs identified are
still very unlikely to actually occur. At higher levels of observation however, the
predictions become more compelling. The seasonal estimates of MMEs are more
accurate, the year by year estimates are even more probable, and the overall tally of
events is a sophisticated prediction. This model predicts 2,296 mass mortality events.
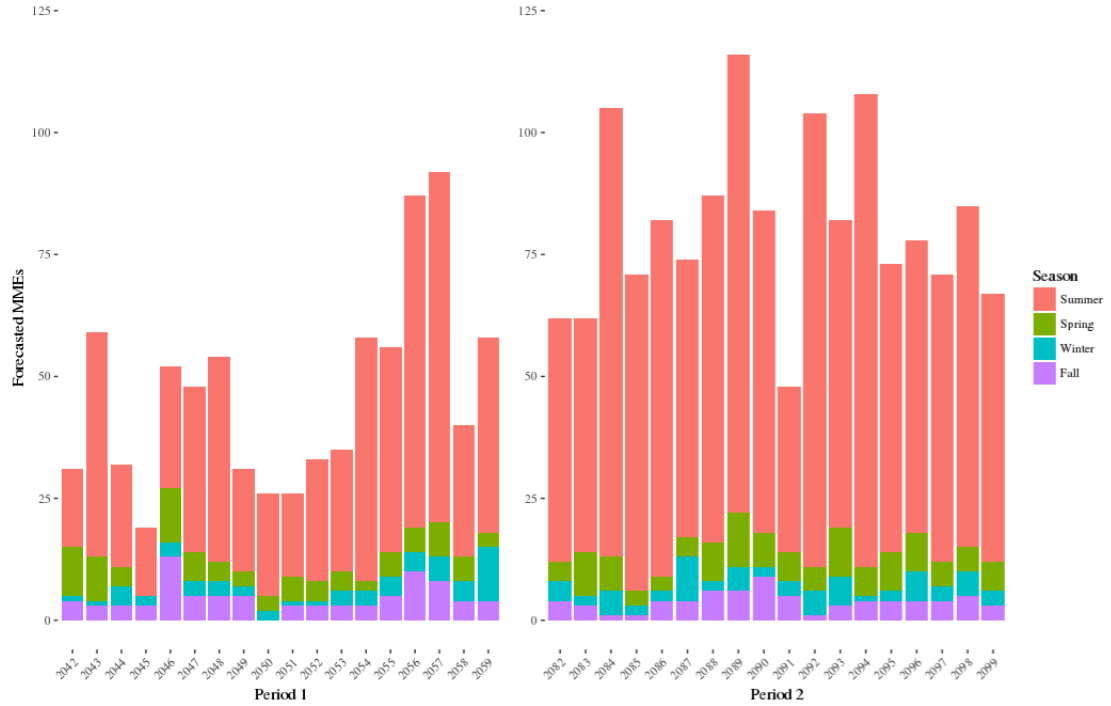
Figure 3.2: Classification forest MME forecasting. The number of MMEs forecasted by the classification forest. Summer months (orange) are June through September, fall months (purple) are October and November, winter months (teal) are January through March and spring months (green) are May and April. It is unclear whether this model is picking out the diverse, non-linear, non-quadratic effects of temperature or if predicted winterkills are merely a result of noise (as in the logistic model). The trend in this model is steeper and the year to year estimates are more variable than in the GLM model. For the first period the number of MMEs predicted are 837 (49 per year). For the second period the number of MMEs predicted are 1459 (86 per year). This model estimates a total of 2,296 MMEs (64 MMEs per year) and a maximum of 116 MMEs in 2089.

3.2.3   Comparing GLM Model and CF Model

Figure 3.3: Comparing GLM and CF forecasts. The relationship between year and MME probability (proportion of MMEs over proportion of non-events) for the two predictive models. The logistic regression for the classification forest (CF) has a beta coefficient = 0.013996 and p <0.001. The logistic regression for the logistic regression model (GLM) has a beta coefficient = 0.004801 and p <0.001.

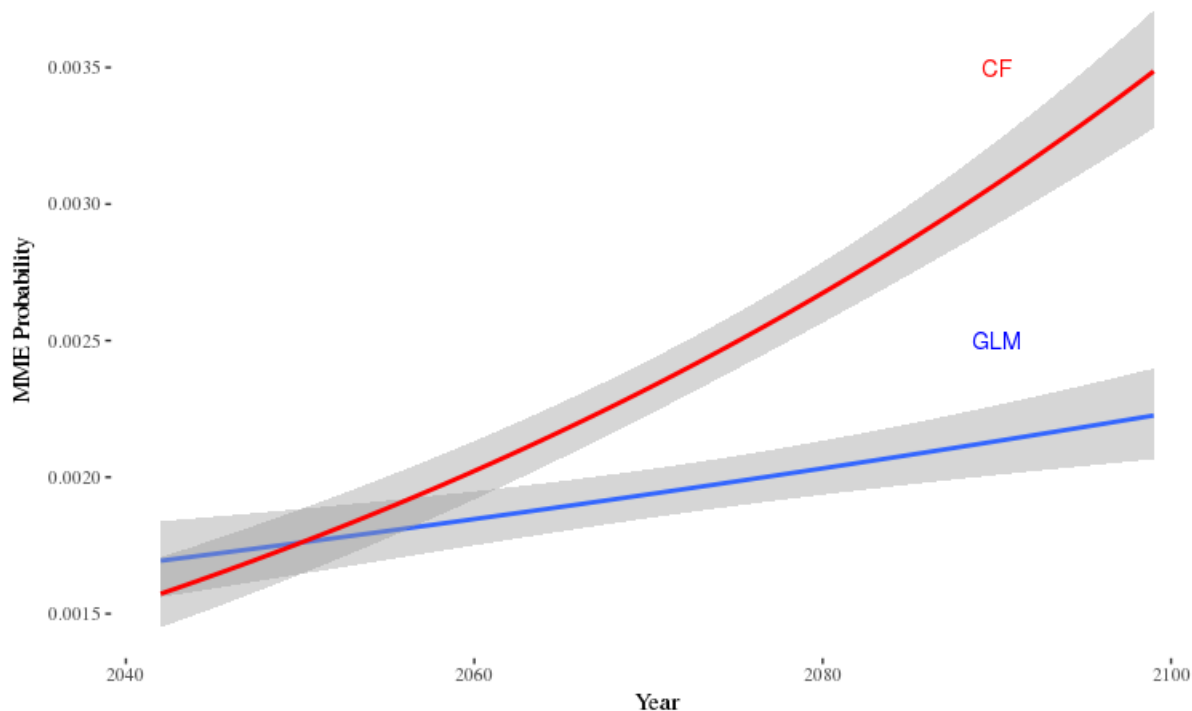Both models predict a significant increase in the probability of MME occurrence over the course of the next century. The classification forest shows a steeper trajectory for this increase (figure 3.3). It is not clear whether this discrepancy is an inherent result of non-linear classification.

Although the logistic model provides a better estimate of MMEs in the historic dataset (2004 - 2014), it is not necessarily a better model. The accuracy of the prediction depends on whether there are non-linear, non-quadratic temperature abnormalities that affect MME occurrence. If so, than the classification forest may actually be the more realistic performer.

As indicated by figure 3.4 both models are oversimplifications of reality. The logistic model primarily predicts a linear effect of temperature on probability of MME occurrence whereas the classification forest primarily classifies MMEs as at risk if they fall within a narrow temperature threshold (28 - 30). Nevertheless, although these models do not account for the plethora of other factors that may influence MMEs they do argue that rising temperatures will have a positive effect on their frequency. If waters continue to warm as projected, we will see more fish die-offs.
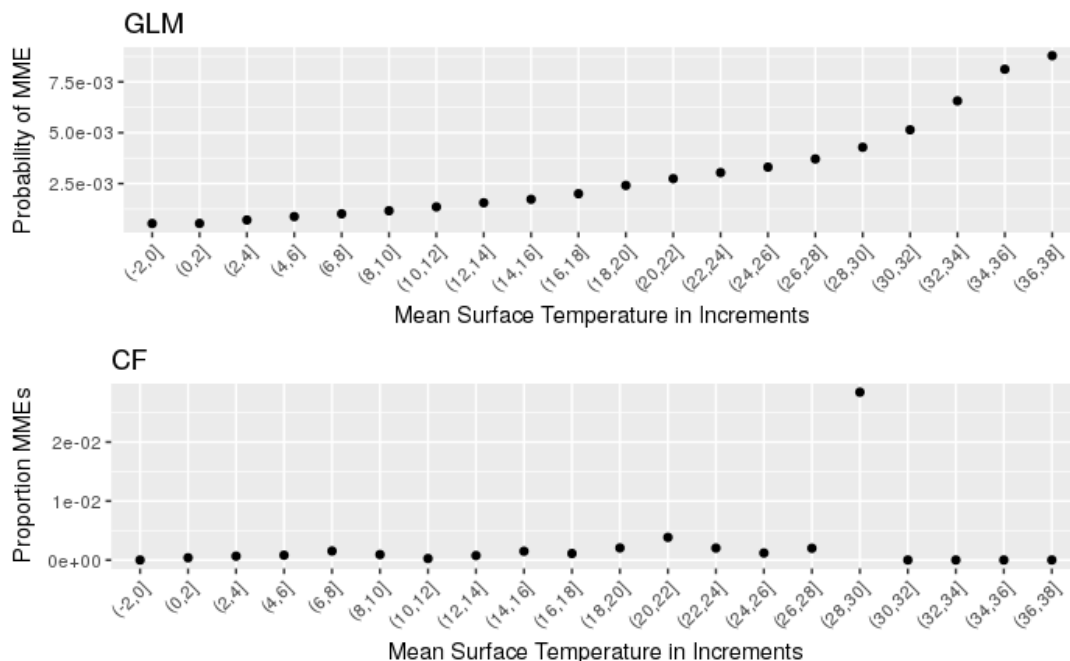
Figure 3.4: [Comparing GLM and CF methods of estimation. The probability of an MME occurring at different temperature intervals for the logistic regression predictions (top) and the proportion of lakes predicted to have an MME by the classification forest for different temperature intervals (bottom). This illustrates how the different approaches converge at relatively similar estimates. The logistic regression displays a linearly increasing effect of temperature on MME probability. The classification forest on the other hand, demonstrates that the majority of the binning occurs at the most common MME temperature interval in the training data (28 - 30). Other, relatively mild spikes in MME occurrence (6-8, 20-22), may reflect noise from the random sampling process or they may correlate with various types of MMEs (winterkill or infectious agents).

3.3 Consequences of Increasing MMEs

These analyses show that, following the worst case scenario estimate of emission output referred to as Representative Concentration Pathway 8.5 (Vuurren et al., 2011), the yearly number of MMEs experienced will continue to increase. For historic data (2004 - 2014), the average number of MMEs per year is 42. For future data (2052 - 2059 and 2082-2099), the average number of MMEs observed in a single year is estimated at 51 for the logistic regression and 64 for the classification forest. Is this within the tolerable frequency of Midwestern lakes or will we instead expect to see significant biological changes to these ecosystems?

Some consequences of MMEs are inherent and can be generalized to other ecosystems. For example, every Wisconsin lake that experiences an MME undergoes a bottleneck effect and loses some of its standing genetic variation. This may lead to a reduced ability to respond to non-thermal consequences of climate change. Addi-

tionally, if MMEs do become more frequent, than Wisconsin fish populations will also become more stochastic, and the likelihood of extirpation will increase (Lande 1993).

In incomplete MMEs, populations that experience selective extermination may become more resistant to similar shocks. Sometimes, species on the brink of extinction may be saved by key mutations that are selected heavily for by the primary mortality cause, this is known as evolutionary rescue (Carlson et al., 2014). However, it is possible that MMEs may occur at a rate too frequent for some populations to undergo evolutionary rescue.

One life cycle strategy that might be favored under increasingly stochastic environments is r-selected specialization. R-selected species maximize growth and early reproductive potential, which may be beneficial in highly volatile environments where the population's carrying capacity is never reach. Indeed, evidence shows that short generation time is heavily correlated with the frequency of catastrophic die-offs (Reed et al., 2003).

Some consequences of MMEs are not inherently generalizable but may be true for most freshwater systems. A case study in lake Mendota, Wisconsin, found that zooplankton were likely to experience explosive growth after a die-off of fish predators leading to zooplankton overgrazing and a serious reduction in lake phytoplankton biomass (Vanni et al., 1990). As healthy phytoplankton concentrations are key to maintaining dissolved oxygen levels in lakes, this is an example of how MMEs may increase the likelihood of subsequent MMEs.

Most ecological effects however, trophic and non-trophic, depend on the unique biotic composition of the affected waterbody. For instance, food web complexity plays an important role in determining the severity of an ecosystem's response to a perturbation (May 1973). Superficially more complex systems are less resilient to changes but this is not always the case as weak trophic interactions may play a more important role in dampening subsequent oscillations (McCann et al., 1998). Before these highly variable scenarios can be considered we must know what species are involved in the MMEs.

Enregisterment of these species is found in appendix A (see table A.1). Viewing this table, we can explore the unifying attributes that might shape certain responses to climate change. One thing that must be considered is the inherent carrying capacity of the species. Bluegill appear as the most frequently affected species likely in part due to the fact that they are such prolific fish, sometimes even considered overpopulated (Otis et al., 1998).

Despite this interference, many patterns emerge. For instance, multiple members of the *Centrarchidae* family (eg., bluegills, crappies, pumpkinseeds) appear highly susceptible to *Columnaris* outbreaks. Interestingly, bass, which also belong to the family, do not appear sensitive to the bacterial pathogen while bullheads, a species of catfish, demonstrate similar patterns of susceptibility. It is not clear what physiological trait leaves some species exposed to *Columnaris* while others remain immune. Identifying and controlling for this unifying feature is an important first step in further analysis of infectious agent kills. If it were possible to only consider lakes that contain species with the same probability of infection then the effect of temperature on infectious agent die-offs might be visible.

Table 4 (see appendix 1) is also useful in determining species that may be of less concern. American gizzard shad for instance, although the fourth most likely fish to be involved in an MME, are almost exclusively winterkill victims. The same is true for freshwater drum. This may reflect an element of the species natural life cycle. Many more gizzard shad MMEs were excluded from the dataset as they were specifically referred to as density dependent spawning stress. The proliferation of gizzard shad winterkills along with the naturalized stochasticity of their population could be evidence that winterkills are less concerning than they would appear.

One at risk species may be northern pike (*E. Lucius*), a member of the only remaining genus of the Esocidae family and a popular game fish. Although abundant, this species appears particularly susceptible to heat related events. It has the largest percentage of summer related mortality events of any major fishkill species, and if the area affected metric is an accurate proxy for kill magnitude than pike die-offs tend to be very large. The CT_{max} tolerance (the temperature at which pike stop consumption) appears to be in the middle of the range of thermal tolerances exhibited by 130 species (Beitlinger et al., 2000; Hanson et al., 1997). This CT_{max} may not reflect the limiting characteristic in its survivability. Alternatively, the high event participation may reflect the pikes prevalence and the skew towards summer related events may be a result of northern pikes superior ability to respond to winterkill, nearly quadrupling its ventilation capacity at low temperatures (Petrosky et Magnuson, 1973). Nevertheless, field studies should be conducted in the Midwest to determine if we should be worried about northern pike populations given their projected increase in mortality with climate change.

Anglers and managers of game populations may be relieved to see that three of the primary commercial species, bass, walleye, and trout, are primarily affected by winterkills which are not predicted to increase. Indeed, few people familiar with the biota of Midwestern lakes would be upset to know that bluegills are the most affected species. Is thinning the population of one of the most abundant and least desirable fish necessarily a bad thing?

3.4 Ethical Considerations

Given that MMEs are projected to increase in frequency, and given that human caused climate change is the reason for this, do we have an ethical responsibility to do something? What if it turns out only less desirable commercial species such as Bluegill will be affected by future MMEs, are we still beholden to conservation? There are two schools of thought around which we can focus this question.

3.4.1 The Reductionist Life School

There is an unspoken acceptance that living things, particularly non-human living things, can be deconstructed to and are only as valuable as the matter and energy that comprise them. Following this logic, there is a pragmatic argument that the importance of saving a species is to preserve the biodiversity of the gene pool and the

balance of ecological systems. The environment should be protected so that humans can continue to extract resources and/or knowledge from it. Following this train of rhetoric MMEs are only consequential if they exterminate or irreparable damage the affected species or its sympatrics. When examining bluegills under climate change, for the next century at least, this does not appear to be the case.

3.4.2   The Emergent Life School

A counter proposal is that there is some inherent, emergent property in living things that is worth preserving for its own sanctity. This concept of the intrinsic value of life was first proposed by Immanuel Kant and is commonly referred to as vitalism. Although there is no way to empirically confirm vitalism, most scientists likely operate under its assumption. If the last member of a species has been studied in every possible manner, and is contributing in no meaningful way to the ecosystem, and you believe exterminating it is still wrong, than you belong in this school of thought. However, the idea that all life is sacred is irreconcilable with the existence of a natural world that is fundamentally built around mortality. What distinguishes the killing of a fish population due to a chemical spill from the extermination of native grasses by invasive spotted knapweed from the allelopathy of a black walnut tree? In these levels of injustice there lies some arbitrary definition of what is natural. Through design or through selection there is some fairness to the natural world that humans somehow transcend. When human action causes an eliminating disruption in the typical, stochastic fluctuations of life and death, some unclearly defined moral threshold is crossed for vitalist environmentalists. By this logic, whether the expected surge in bluegill mortality matters or not is entirely dependent on whether that uptick is a consequence of our actions. In regards to climate change, it is.

3.4.3   Management Strategies

The objective of conservation must be clearly defined before creating management strategies. What is the goal of mitigating MMEs? Is preventing ecological chaos a means to maintaining a healthy sporting industry? If so than restocking game species may be a viable response strategy. In fact, in Midwestern American lakes a surge in summer mortality events may provide a competitive edge to the primary game fish.

If the objective of conservation is to minimize unnatural mortality than restocking only select species in commercially valuable lakes is not sufficient. Additionally, evidence suggests that restocking is less effective at establishing fish populations than natural recruitment (Thiem et al., 2017). One preventative option is to expand water aeration infrastructure in lakes likely to see summerkills. Like restocking however, this is an artificial fix with the potential for a cascade of ecological consequences.

The only truly meaningful response is to reduce our carbon footprint. Temperature estimations of the worst case emission scenario were used to forecast MMEs. Hopefully, as we become aware of the effect we have on our environment, conscientious energy consumption will be adopted at a global scale and the predictions of this thesis will be revealed as exaggerations.

Conclusion

Lakes around the world have begun a rapid but variable trend towards warmer temperatures (OReilly et al., 2015). This change is estimated for two periods of future years by evaluating a complex assortment of lake metrics (Winslow et al., 2017). I have shown that based on these changing lake temperatures, we will also see a rise in the number of summerkill die-off events. This influx of fish mortality will change the ecological dynamics of many Midwestern lakes and possibly put populations of summerkill susceptible fish, such as northern pike, at risk of extirpation. Mitigation efforts range from restocking to artificial lake aeration, but the best way to prevent catastrophic freshwater restructuring is to end global warming.

Appendix A

R code used in this thesis can be found at: https://github.com/AaronTill/Thesis_2018.git

(for table A.1 see page 44)

Table A.1: Species involved in MMEs. The type of fish mentioned in the qualitative assessments of the MMEs. Different fish (or amphibians or crustaceans) were referred to primarily by the common name for their species but also occasionally by family. Other terms such as gamefish or panfish were used to refer to affected species but were not included as these terms are ambiguous. A large portion of MMEs did not have any fish species mentioned, many had multiple. The number of MMEs identified in is only a rough estimate of the species true participation in MMEs. Total stream miles or lake acres affected is a proxy for total magnitude of the kills, although it is susceptible to a high degree of reporting bias. The CT_{max} values (max temperature at which consumption can occur) are taken from Fish Bioenergetics 3.0 (Hanson et al., 1997). CT_{max} variation reflects the range of juveniles to adults, or in the case of bass, the differences between striped, smallmouth and largemouth bass.

Classifier	Number of MMEs Identified In	Mean Stream Miles or Lake Acres Affected	CT_{max} (max temperature of consumption)	Percent of MMEs In June, July, August	Most Common MME Type Involved In
Bluegill	96	531	36-37°C	60.4%	Infectious Agents
Crappie	69	290	NA	63.8%	Infectious Agents
Bass	66	546	22.7-31.3°C	40.9%	Winterkill
Gizzard Shad	41	353	NA	9.8%	Winterkill
Bullhead	28	851	NA	60.8%	Infectious Agents
Perch	28	247	28-32°C	42.9%	Winterkill
Northern Pike	27	711	34°C	74.1%	Summerkill
Pumpkinseed	21	733	NA	61.9%	Infectious Agents
Freshwater Drum	19	12	NA	5.3 %	Infectious Agents
Suckerfish (Catostomidae)	14	287	NA	71.4%	Summerkill
Carp	14	332	NA	35.7%	Winterkill
Walleye	11	722	28°C	18.2%	Winterkill
Sunfish (Centrarchidae)	8	526	NA	100%	Infectious Agents
Minnow	5	57	NA	100%	Summerkill
Cisco	4	913	NA	75%	Summerkill
Trout	3	9	NA	33.3%	Winterkill
Muskellunge	3	1766	34°C	100%	Infectious Agents
Catfish	2	803	NA	0%	Winterkill
Bigmouth Buffalo	2	<1	NA	0%	Summerkill
Redhorse	2	15	NA	100%	Summerkill
Frogs	2	18	NA	0%	Winterkill/ Anthropogenic
Golden Shiner	2	61	NA	0%	Winterkill
Crayfish	1	4	NA	0%	Anthropogenic

References

Adrian, Rita, Catherine M. OReilly, Horacio Zagarese, Stephen B. Baines, Dag O. Hessen, Wendel Keller, David M. Livingstone, et al. 2009. Lakes as Sentinels of Climate Change. Limnology and Oceanography 54 (6part2): 228397.

Allison, Edward H., Allison L. Perry, Marie-Caroline Badjeck, W. Neil Adger, Katrina Brown, Declan Conway, Ashley S. Halls, et al. 2009. Vulnerability of National Economies to the Impacts of Climate Change on Fisheries. Fish and Fisheries 10 (2): 17396.

Anderson, Sean C., Trevor A. Branch, Andrew B. Cooper, and Nicholas K. Dulvy. 2017. Black-Swan Events in Animal Populations. Proceedings of the National Academy of Sciences of the United States of America 114 (12): 325257.

Barrett, Rowan D. H., and Dolph Schluter. 2008. Adaptation from Standing Genetic Variation. Trends in Ecology & Evolution 23 (1): 3844.

Beitinger, Thomas L., Wayne A. Bennett, and Robert W. McCauley. 2000. Temperature Tolerances of North American Freshwater Fishes Exposed to Dynamic Changes in Temperature. Environmental Biology of Fishes 58 (3). Kluwer Academic Publishers: 23775.

Brookes, Justin D., and Cayelan C. Carey. 2011. Resilience to Blooms. Science 334 (6052): 4647.

Bucher, Enrique H. 1992. The Causes of Extinction of the Passenger Pigeon. In Current Ornithology, edited by Dennis M. Power, 136. Boston, MA: Springer US.

Buckley, Lauren B., and Raymond B. Huey. 2016. Temperature Extremes: Geographic Patterns, Recent Changes, and Implications for Organismal Vulnerabilities. Global Change Biology 22 (12): 382942.

Carlson, Stephanie M., Curry J. Cunningham, and Peter A. H. Westley. 2014. Evolutionary Rescue in a Changing World. Trends in Ecology & Evolution 29 (9): 52130.

Decostere, A., F. Haesebrouck, J. F. Turnbull, and G. Charlier. 1999. Influence of

Water Quality and Temperature on Adhesion of High and Low Virulence Flavobacterium Columnare Strains to Isolated Gill Arches. Journal of Fish Diseases 22 (1): 111.

Fey, Samuel B., Adam M. Siepielski, Sbastien Nussl, Kristina Cervantes-Yoshida, Jason L. Hwan, Eric R. Huber, Maxfield J. Fey, Alessandro Catenazzi, and Stephanie M. Carlson. 2015. Recent Shifts in the Occurrence, Cause, and Magnitude of Animal Mass Mortality Events. Proceedings of the National Academy of Sciences of the United States of America 112 (4): 108388.

Greenbank, John. 1945. Limnological Conditions in Ice-Covered Lakes, Especially as Related to Winter-Kill of Fish. Ecological Monographs 15 (4): 34392.

Hanson, Paul C., Johnson, Timothy B., Schindler, Daniel E., and James F. Kitchell. 1997. Fish Bioenergetics 3.0. Center for Limnology, University of Wisconsin-Madison. WISCU_T_97-001.

Ho, Tin Kam. 1995. Random Decision Forests. In Proceedings of 3rd International Conference on Document Analysis and Recognition, 1:27882 vol.1.

Holt, R. A., J. E. Sanders, J. L. Zinn, J. L. Fryer, and K. S. Pilcher. 1975. Relation of Water Temperature to Flexibacter Columnaris Infection in Steelhead Trout (Salmo Gairdneri), Coho (Oncorhynchus Kisutch) and Chinook (O. Tshawytscha) Salmon. Journal of the Fisheries Research Board of Canada 32 (9): 155359.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, doi:10.1017/CBO9781107415324.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. An Introduction to Statistical Learning: With Applications in R. 1st ed. Springer Texts in Statistics. Springer-Verlag New York.

Kundzewicz, Z. W., L. J. Mata, N. W. Arnell, P. Dll, B. Jimenez, K. Miller, T. Oki, Z. en, and I. Shiklomanov. 2008. The Implications of Projected Climate Change for Freshwater Resources and Their Management. Hydrological Sciences Journal 53 (1): 310.

La, Van T., and Steven J. Cooke. 2011. Advancing the Science and Practice of Fish Kill Investigations. Reviews in Fisheries Science 19 (1): 2133.

Lande, R. 1993. Risks of Population Extinction from Demographic and Environmental Stochasticity and Random Catastrophes. The American Naturalist 142 (6): 91127.

Lovejoy, Thomas E., and Hannah Lee. 2005. Climate Change and Biodiversity. The Energy and Resources Institute (TERI).

Magnuson, John J., Annamarie L. Beckel, Ken Mills, and Stephen B. Brandt. 1985. Surviving Winter Hypoxia: Behavioral Adaptations of Fishes in a Northern Wisconsin Winterkill Lake. Environmental Biology of Fishes 14 (4): 24150.

May, Robert M. 1973. Qualitative Stability in Model Ecosystems. Ecology 54 (3): 63841.

McCann, Kevin, Alan Hastings, and Gary R. Huxel. 1998. Weak Trophic Interactions and the Balance of Nature. Nature 395 (October). Macmillan Magazines Ltd.: 794.

McNutt, Marcia. 2013. Climate Change Impacts. Science 341 (6145): 435.

Moore, Walter G. 1942. Field Studies on the Oxygen Requirements of Certain Fresh-Water Fishes. Ecology 23 (3). Ecological Society of America: 31929.

Mora, Camilo, Abby G. Frazier, Ryan J. Longman, Rachel S. Dacks, Maya M. Walton, Eric J. Tong, Joseph J. Sanchez, et al. 2013. The Projected Timing of Climate Departure from Recent Variability. Nature 502 (7470): nature12540.

Moreno, Juan, and Anders Pape Mller. 2011. Extreme Climatic Events in Relation to Global Change and Their Impact on Life Histories. Current Zoology 57 (3): 37589.

ilsson, Gran E., and Sjannie Lefevre. 2016. Physiological Challenges to Fishes in a Warmer and Acidified Future. Physiology 31 (6): 40917.

OReilly, Catherine M., Sapna Sharma, Derek K. Gray, Stephanie E. Hampton, Jordan S. Read, Rex J. Rowley, Philipp Schneider, et al. 2015. Rapid and Highly Variable Warming of Lake Surface Waters around the Globe. Geophysical Research Letters 42 (24): 2015GL066235.

Otis, Keith J., Randal R. Piette, John E. Keppler, and Paul W. Rasmussen. 1998. A Largemouth Bass Closed Fishery to Control an Overabundant Bluegill Population in a Wisconsin Lake. Journal of Freshwater Ecology 13 (4). Taylor & Francis: 391403.

Pecl, Gretta T., Miguel B. Arajo, Johann D. Bell, Julia Blanchard, Timothy C. Bonebrake, I-Ching Chen, Timothy D. Clark, et al. 2017. Biodiversity Redistribution under Climate Change: Impacts on Ecosystems and Human Well-Being. Science 355 (6332).

Petrosky, Bernard R., and John J. Magnuson. 1973. Behavioral Responses of North-

ern Pike, Yellow Perch and Bluegill to Oxygen Concentrations under Simulated Winterkill Conditions. Copeia 1973 (1). [American Society of Ichthyologists and Herpetologists (ASIH), Allen Press]: 12433.

Reed, D. H., J. J. OGrady, J. D. Ballou, and R. Frankham. 2003. The Frequency and Severity of Catastrophic Die-Offs in Vertebrates. Animal Conservation 6 (May): 10914.

Smith, Daniel Wilkins, and Raul H. Piedrahita. 1988. The Relation between Phytoplankton and Dissolved Oxygen in Fish Ponds. Aquaculture 68 (3): 24965.

Solomon, Susan, Gian-Kasper Plattner, Reto Knutti, and Pierre Friedlingstein. 2009. Irreversible Climate Change due to Carbon Dioxide Emissions. Proceedings of the National Academy of Sciences of the United States of America 106 (6): 17049.

Stephens, P. A., and W. J. Sutherland. 1999. Consequences of the Allee Effect for Behaviour, Ecology and Conservation. Trends in Ecology & Evolution 14 (10): 4015.

Thiem, Jason D., Ian J. Wooden, Lee J. Baumgartner, Gavin L. Butler, Jamin P. Forbes, and John Conallin. 2017. Recovery from a Fish Kill in a Semi-Arid Australian River: Can Stocking Augment Natural Recruitment Processes? Austral Ecology 42 (2): 21826.
Vanni, Michael J., Chris Luecke, James F. Kitchell, Yvonne Allen, Jo Temte, and John J. Magnuson. 1990. Effects on Lower Trophic Levels of Massive Fish Mortality. Nature 344 (March). Nature Publishing Group: 333.

Vuuren, Detlef P. van, Jae Edmonds, Mikiko Kainuma, Keywan Riahi, Allison Thomson, Kathy Hibbard, George C. Hurtt, et al. 2011. The Representative Concentration Pathways: An Overview. Climatic Change 109 (1): 5.

Wethey, David S., Sarah A. Woodin, Thomas J. Hilbish, Sierra J. Jones, Fernando P. Lima, and Pamela M. Brannock. 2011. Response of Intertidal Populations to Climate: Effects of Extreme Events versus Long Term Change. Journal of Experimental Marine Biology and Ecology, Global change in marine ecosystems, 400 (1): 13244.

Winslow, Luke A., Gretchen J. A. Hansen, Jordan S. Read, and Michael Notaro. 2017. Large-Scale Modeled Contemporary and Future Water Temperature Estimates for 10774 Midwestern U.S. Lakes. Scientific Data 4 (April): sdata201753.

Xenopoulos, Marguerite A., David M. Lodge, Joseph Alcamo, Michael Mrker, Kerstin Schulze, and Detlef P. Van Vuuren. 2005. Scenarios of Freshwater Fish Extinctions from Climate Change and Water Withdrawal. Global Change Biology 11 (10): 155764.