Class Exercise: Factors Influencing Abundance & Virus Isolation in Mosquitoes

Which environmental, ecological, and temporal factors impact mosquito populations

Andrew Ruiz

2024-02-02

# 1. Summary/Abstract

Eastern equine encephalitis (EEE) is caused by an Alphavirus transmitted to humans by the bite of an infected mosquito. Human infections are rare, but serious. About 30-50% of symptomatic cases lead to death and those who survive are left with life-long disability. The transmission cycle for EEE includes two distinct cycles: an amplifying enzootic cycle in which the virus is transmitted between the Culiseta melanura mosquito and birds and an epizootic cycle in which bridge vector mosquitoes transmit the EEE virus (EEEv) from birds to humans and other mammals. Because C. melanura feeds almost exclusively on birds, it is not considered a direct human threat. However, EEEv isolations in C. melanura are an early warning sign that EEEv is circulating in the ecosystem. Isolations in bridge vector mosquito species indicate heightened transmission risk to humans.

This project will lay the foundation for my dissertation research. **The goal for this phase of the project is to answer which environmental, ecological, and temporal factors influence the transmission cycles for EEE in southeastern Massachusetts.**

# 2. Introduction

## 2.1 General Background Information

Humans and other mammals are dead-end hosts for EEE. About 94% of human EEE infections are asymptomatic. While symptomatic cases are rare, the consequences can be severe. Mortality rates associated with symptomatic infections range from 30 to 50%. Those who survive often experience lifelong disability that includes debilitating neurological damage. On average, there are 11 cases reported in the US annually. Most occur in Massachusetts and Florida. Until recently, nearly all human EEE cases in Massachusetts occurred in two counties: Bristol and Plymouth.

Since 2000, there have been 45 human cases recorded in Massachusetts, resulting in 22 deaths. There are no human vaccines for EEE, and treatment consists of palliative care only. Prevention measures include personal behaviors to avoid mosquito bites and decreasing the mosquito population through pesticide use and environmental modifications like removing standing water. The Bristol County Mosquito Control Project (BCMCP) coordinates mosquito surveillance and testing in the county from June to October. BCMCP has used the same sentinel collection sites and trapping methods for over 20 years. Once trapped, mosquitoes are sorted by species, and vector species are submitted to the Massachusetts Department of Health’s (MDPH) State Lab for PCR virus testing. Results are available within 24 hours of submission. When rates of EEE mosquito infections are above a defined threshold, MDPH notifies the local boards of health and recommends preventive measures. Recommended measures include outreach and education to increase personal protective practices (Heymann, 2015). Additionally, MDPH may recommend that cities and towns in the affected areas cancel evening outdoor events and discourage outdoor activities when mosquito vectors are most active (dawn and dusk) (Massachusetts Department of Public Health, n.d.). When infection rates are high enough to indicate imminent human transmission, the Commonwealth of Massachusetts will recommend and fund aerial pesticide applications over the affected areas.

## 2.2 Description of data and data source

*The cornerstone of this research will be the mosquito collection data provided by a mosquito control project in SE Massachusetts. A formal data-sharing agreement will be in place by January 2024. The dataset encompasses nearly two decades of detailed mosquito surveillance. The datasets include spatial identifiers. Mosquito collection data will be integrated with other datasets using spatial or temporal attributes. These datasets will include biological and ecological data, environmental and spatial data, and demographic data, which are crucial in understanding the transmission dynamics of EEE. Potential datasets with their sources are outlined in the table below. Other datasets may be included based on availability and need.*

| Data | Description | Source |
| --- | --- | --- |
| Mosquito Surveillance Data | Records of mosquito species counts, locations, and dates of collection. Data on the presence of EEE in mosquito populations is particularly valuable. | XXX County Mosquito Control |
| Weather Data | Temperature, humidity, rainfall, and wind speed influence mosquito activity, population dynamics, and virus transmission. | NOAA |
| Bird Population Data | Since birds are a natural reservoir for EEE, information on bird populations and migration patterns could be relevant. | Crowd-sourced bird data |
| Human Case Reports | Data on confirmed human cases of EEE, including location, date of onset, and clinical outcomes. | MA Dept of Public Health publications |
| Veterinary Surveillance Data | Since horses and other mammals can also be affected by EEE, veterinary records could provide early warning signs of virus activity in an area. | MA Dept of Public Health |

### 2.2.1 Currently, I have collected 2 datasets that are described below:

#### 2.2.1.1 The first dataset is the mosquito collection dataset

dp1 = here("data", "raw-data", "MOSQ\_MADA.xlsx")  
  
mos\_data = read\_excel(dp1)  
  
head(mos\_data)

# A tibble: 6 × 12  
 MCD `Pool ID` `Species code` `Trap Set Date` `Collection Date` Town   
 <chr> <chr> <chr> <dttm> <dttm> <chr>  
1 Bristol BR07NS-0… NCR 2007-05-09 00:00:00 2007-05-10 00:00:00 Digh…  
2 Bristol BR07NS-0… NCR 2007-05-09 00:00:00 2007-05-10 00:00:00 Digh…  
3 Bristol BR07NS-0… NCR 2007-05-09 00:00:00 2007-05-10 00:00:00 Swan…  
4 State … SL07NS-0… MEL 2007-05-24 00:00:00 2007-05-25 00:00:00 New …  
5 State … SL07NS-0… RES 2007-05-24 00:00:00 2007-05-25 00:00:00 New …  
6 State … SL07NS-0… ABS 2007-05-24 00:00:00 2007-05-25 00:00:00 New …  
# ℹ 6 more variables: `Number of Traps` <dbl>, `Trap Type` <chr>,  
# `Pool Size` <dbl>, `Submitted for Testing` <chr>, Genus <chr>,  
# Species <chr>

summary(mos\_data)

MCD Pool ID Species code   
 Length:45393 Length:45393 Length:45393   
 Class :character Class :character Class :character   
 Mode :character Mode :character Mode :character   
   
   
   
   
 Trap Set Date Collection Date   
 Min. :2007-05-09 00:00:00.00 Min. :2007-05-10 00:00:00.00   
 1st Qu.:2011-09-18 00:00:00.00 1st Qu.:2011-09-19 00:00:00.00   
 Median :2016-07-18 00:00:00.00 Median :2016-07-19 00:00:00.00   
 Mean :2016-01-08 04:36:04.65 Mean :2016-01-09 07:29:04.68   
 3rd Qu.:2020-06-21 00:00:00.00 3rd Qu.:2020-06-22 00:00:00.00   
 Max. :2023-10-31 00:00:00.00 Max. :2023-11-01 00:00:00.00   
 NA's :4 NA's :4   
 Town Number of Traps Trap Type Pool Size   
 Length:45393 Min. : 1.000 Length:45393 Min. : 0.00   
 Class :character 1st Qu.: 1.000 Class :character 1st Qu.: 1.00   
 Mode :character Median : 1.000 Mode :character Median : 5.00   
 Mean : 1.681 Mean : 23.11   
 3rd Qu.: 2.000 3rd Qu.: 24.00   
 Max. :14.000 Max. :14367.00   
 NA's :4 NA's :13   
 Submitted for Testing Genus Species   
 Length:45393 Length:45393 Length:45393   
 Class :character Class :character Class :character   
 Mode :character Mode :character Mode :character

#### 2.2.1.2 The second dataset is the weather data spanning the same time period as the mosquito collection data

dp2 = here("data", "raw-data", "weather\_airport.csv")  
weather = read.csv(dp2, fileEncoding = "UTF-8")  
head(weather)

STATION NAME DATE AWND AWND\_ATTRIBUTES  
1 USW00054777 TAUNTON MUNICIPAL AIRPORT, MA US 1/1/07 6.49 ,,W  
2 USW00054777 TAUNTON MUNICIPAL AIRPORT, MA US 1/2/07 8.95 ,,W  
3 USW00054777 TAUNTON MUNICIPAL AIRPORT, MA US 1/3/07 7.61 ,,W  
4 USW00054777 TAUNTON MUNICIPAL AIRPORT, MA US 1/4/07 6.71 ,,W  
5 USW00054777 TAUNTON MUNICIPAL AIRPORT, MA US 1/5/07 8.28 ,,W  
6 USW00054777 TAUNTON MUNICIPAL AIRPORT, MA US 1/6/07 10.51 ,,W  
 PRCP PRCP\_ATTRIBUTES TMAX TMAX\_ATTRIBUTES TMIN TMIN\_ATTRIBUTES WT01  
1 1.17 ,,W 57 ,,W 26 ,,W 1  
2 0.01 ,,W 48 ,,W 34 ,,W NA  
3 0.00 ,,W 50 ,,W 25 ,,W NA  
4 0.00 ,,W 54 ,,W 37 ,,W NA  
5 0.04 ,,W 60 ,,W 48 ,,W 1  
6 0.05 ,,W 63 ,,W 54 ,,W 1  
 WT01\_ATTRIBUTES WT02 WT02\_ATTRIBUTES WT03 WT03\_ATTRIBUTES WT08  
1 ,,W NA NA NA  
2 NA NA NA  
3 NA NA NA  
4 NA NA NA  
5 ,,W NA NA NA  
6 ,,W NA NA NA  
 WT08\_ATTRIBUTES  
1   
2   
3   
4   
5   
6

summary(weather)

STATION NAME DATE AWND   
 Length:5977 Length:5977 Length:5977 Min. : 0.000   
 Class :character Class :character Class :character 1st Qu.: 3.360   
 Mode :character Mode :character Mode :character Median : 4.920   
 Mean : 5.381   
 3rd Qu.: 6.930   
 Max. :18.570   
 NA's :104   
 AWND\_ATTRIBUTES PRCP PRCP\_ATTRIBUTES TMAX   
 Length:5977 Min. :0.0000 Length:5977 Min. : 10.00   
 Class :character 1st Qu.:0.0000 Class :character 1st Qu.: 48.00   
 Mode :character Median :0.0000 Mode :character Median : 64.00   
 Mean :0.1304 Mean : 62.63   
 3rd Qu.:0.0700 3rd Qu.: 78.00   
 Max. :4.2100 Max. :100.00   
 NA's :22 NA's :95   
 TMAX\_ATTRIBUTES TMIN TMIN\_ATTRIBUTES WT01   
 Length:5977 Min. :-16.00 Length:5977 Min. :1   
 Class :character 1st Qu.: 27.00 Class :character 1st Qu.:1   
 Mode :character Median : 41.00 Mode :character Median :1   
 Mean : 40.47 Mean :1   
 3rd Qu.: 54.00 3rd Qu.:1   
 Max. : 78.00 Max. :1   
 NA's :95 NA's :4533   
 WT01\_ATTRIBUTES WT02 WT02\_ATTRIBUTES WT03   
 Length:5977 Min. :1 Length:5977 Min. :1   
 Class :character 1st Qu.:1 Class :character 1st Qu.:1   
 Mode :character Median :1 Mode :character Median :1   
 Mean :1 Mean :1   
 3rd Qu.:1 3rd Qu.:1   
 Max. :1 Max. :1   
 NA's :5584 NA's :5825   
 WT03\_ATTRIBUTES WT08 WT08\_ATTRIBUTES   
 Length:5977 Min. :1 Length:5977   
 Class :character 1st Qu.:1 Class :character   
 Mode :character Median :1 Mode :character   
 Mean :1   
 3rd Qu.:1   
 Max. :1   
 NA's :5603

## 2.3 Questions/Hypotheses to be addressed

\*This project will lay the foundation for my dissertation research. **The goal for this phase of the project is to answer which environmental, ecological, and temporal factors influence the transmission cycles for EEE in southeastern Massachusetts.** While the following questions will not be addressed in this phase, this project will hopefully inform research that will answer the following questions:

1. Can an AI-driven model that leverages historic mosquito surveillance, ecological, and environmental data accurately quantify the risk of human EEEv infection in southeastern Massachusetts?

2. What is the potential of machine learning algorithms to identify early warning signals for EEE outbreaks, enabling timely public health interventions in Massachusetts to prevent human infections?

3. How can machine learning models use real-time data on mosquito activity, environmental factors, and weather to accurately predict when the risk of EEE transmission to humans is low enough to lift evening outdoor activity restrictions? \*

To cite other work (important everywhere, but likely happens first in introduction), make sure your references are in the bibtex file specified in the YAML header above (here dataanalysis\_template\_references.bib) and have the right bibtex key. Then you can include like this:

Examples of reproducible research projects can for instance be found in [@mckay2020; @mckay2020a].

# 3. Methods

*Specific methods will be determined when all data sources are confirmed.*

## 3.1 The process will involve creating a relational database in which all datasets are joined using spatial or temporal attributes. This will lead to a thorough examination of the datasets to understand the distributions, identify missing data and outliers, and uncover potential correlations between variables. The EDA will involve using visualizations, geographic information systems, and statistical tests to understand how biological, ecological, environmental, and demographic forces influence the EEE transmission cycles. The EDA will inform the preprocessing steps, feature engineering, and selection of appropriate AI models.

### 3.1.1 Schematic of workflow

*This section will be updated when all data sources are confirmed*

## 3.2 Data aquisition

With the exception of the mosquito collection data, all other datasets are publicly available for use without restrictions.

## 3.3 Data import and cleaning

*This step will be completed once all data sources are confirmed*

## 3.4 Statistical analysis

This exercise will serve as the beginning of the exploratory data analysis phase. One of the main objectives for of this phase is describe the data, understand its structure, clean data, and identify associations within the datasets.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

*These will be generated in later phases of the exercise*

## 4.2 Basic statistical analysis

## 4.3 Full analysis

# 5. Discussion

## 5.1 Strengths and Limitations

## 5.2 Conclusions

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper [@leek2015] discusses types of analyses.

These papers [@mckay2020; @mckay2020a] are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# 6. References