

West-Nile Virus Prediction

Group 6: Stuart, Lester & Marianne



Problem Statement

We aim to produce a model which will **accurately identify the key contributing factors** that leads to the presence of the West Nile Virus (WNV).

The model and insights from the cost-benefit analysis could then be used by the Centers for Disease Control and Prevention (CDC) to help **predict future outbreaks of WNV** and **effectively allocate resources** to mitigate it.

Data Cleaning & Pre-Processing



Data Cleaning & Pre-Processing

Dropped

- Dropped redundant columns

Combined

- Combined Train and Weather data
- Spray data wasn't included

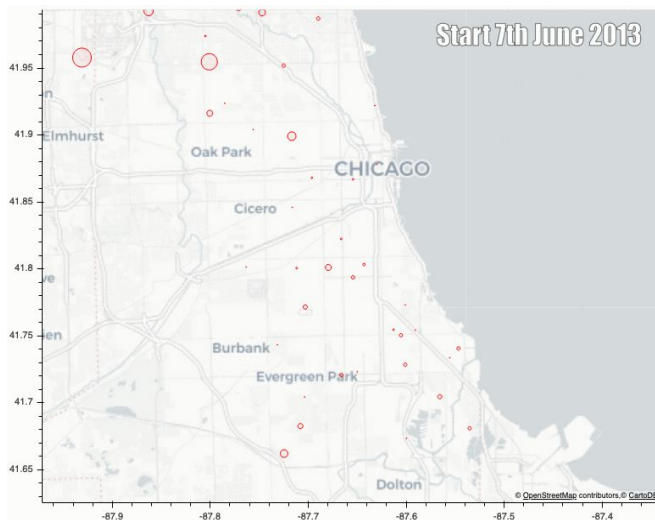
Feature Engineering

- One-hot encoded categorical features like species and month
- Added lagged weather features for rainfall and temperature

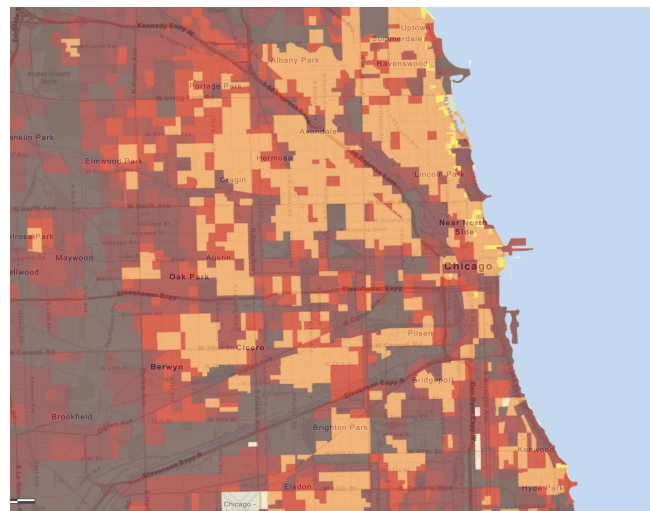
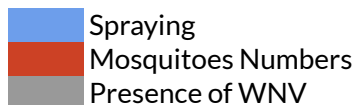
Exploratory Data Analysis



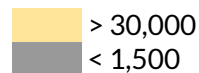
2013 Spray and Mosquito Count Visualisation



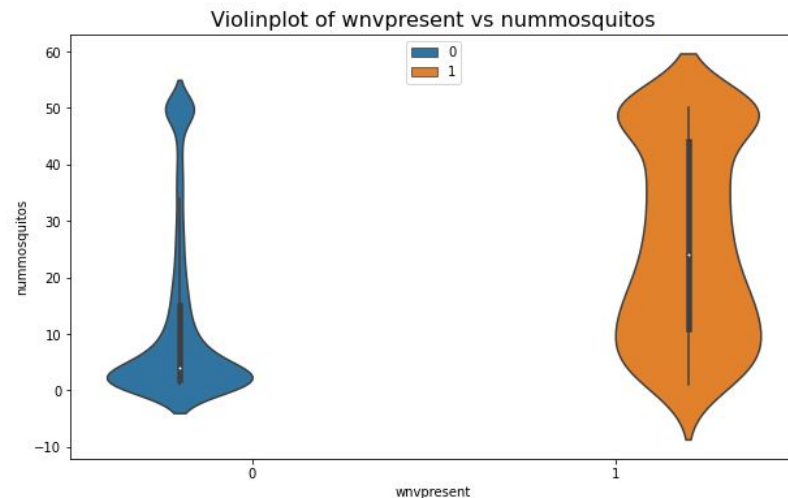
Time Frame: 7 Jun - 9 Sep 2013



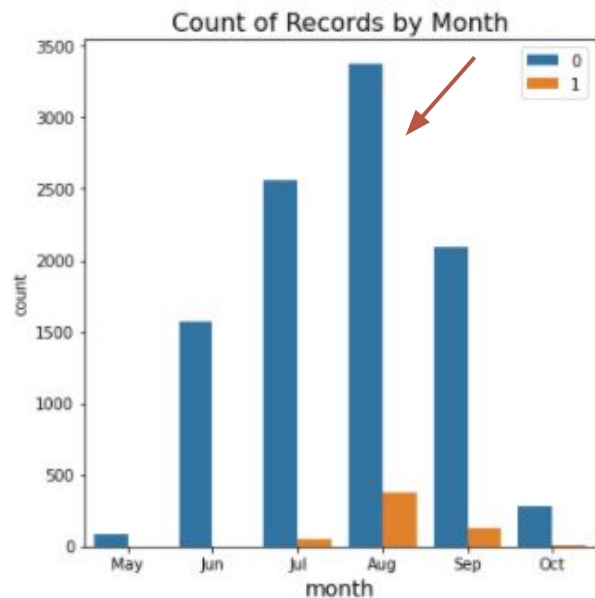
People per sq km



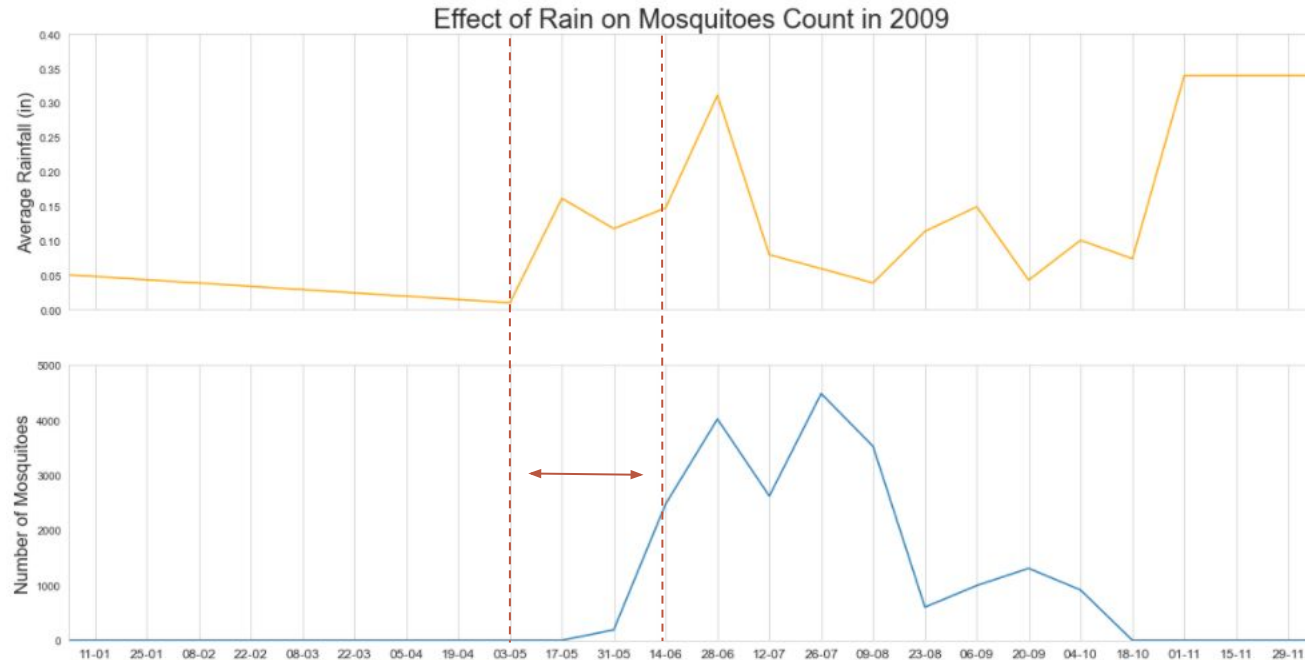
Finding #1:
The more mosquitos there
are in the trap, the more
likely WNV is present.



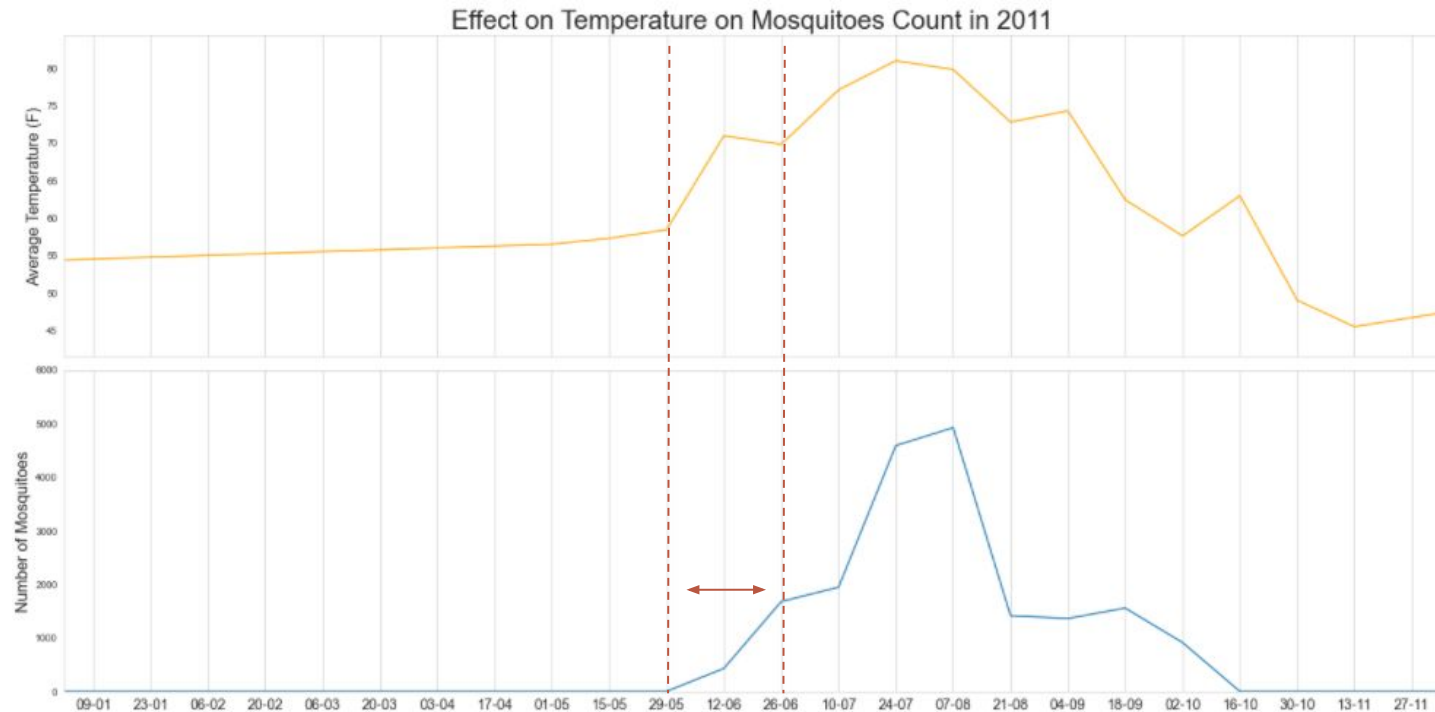
Finding #2:
Number of mosquitoes
peaks in August



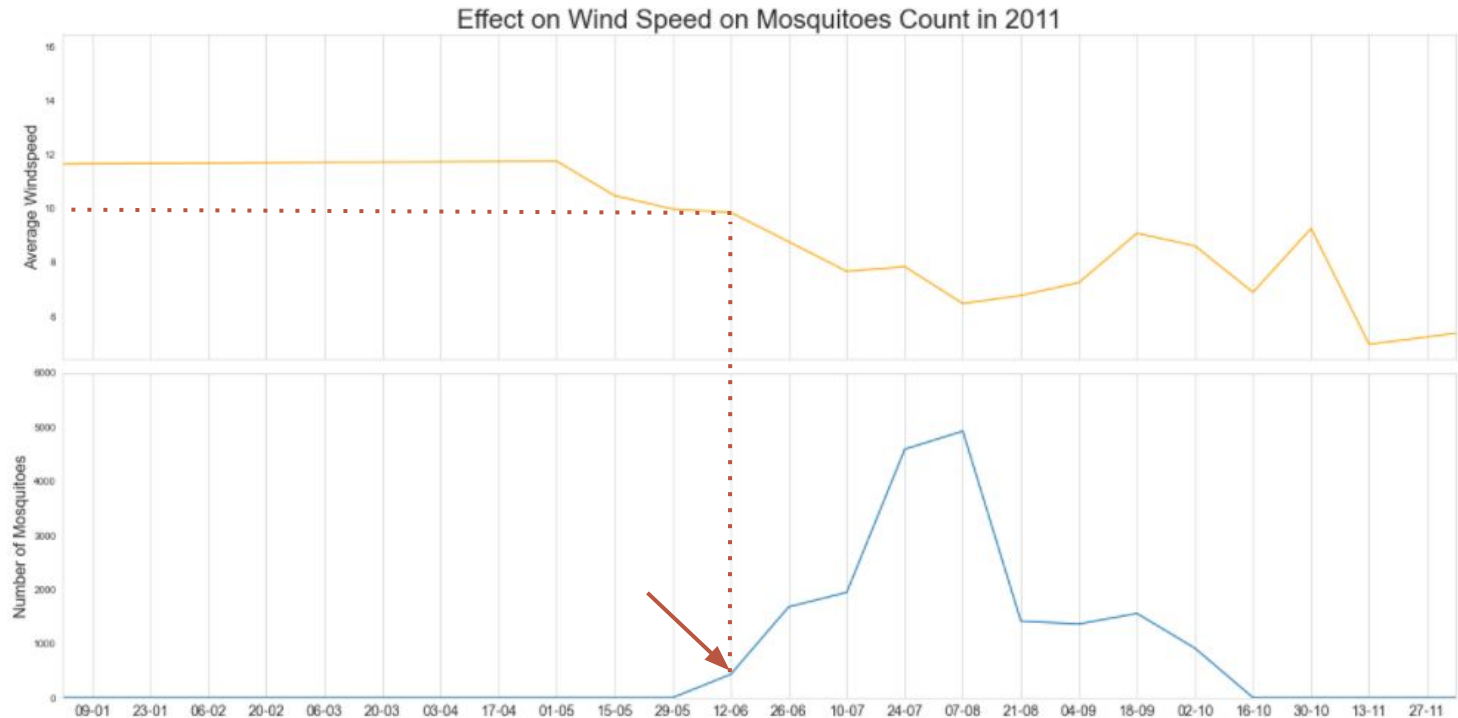
Finding #3: 2-4 week lag between rainfall and number of mosquitoes



Finding #4: 2 week lag between temperature and number of mosquitoes



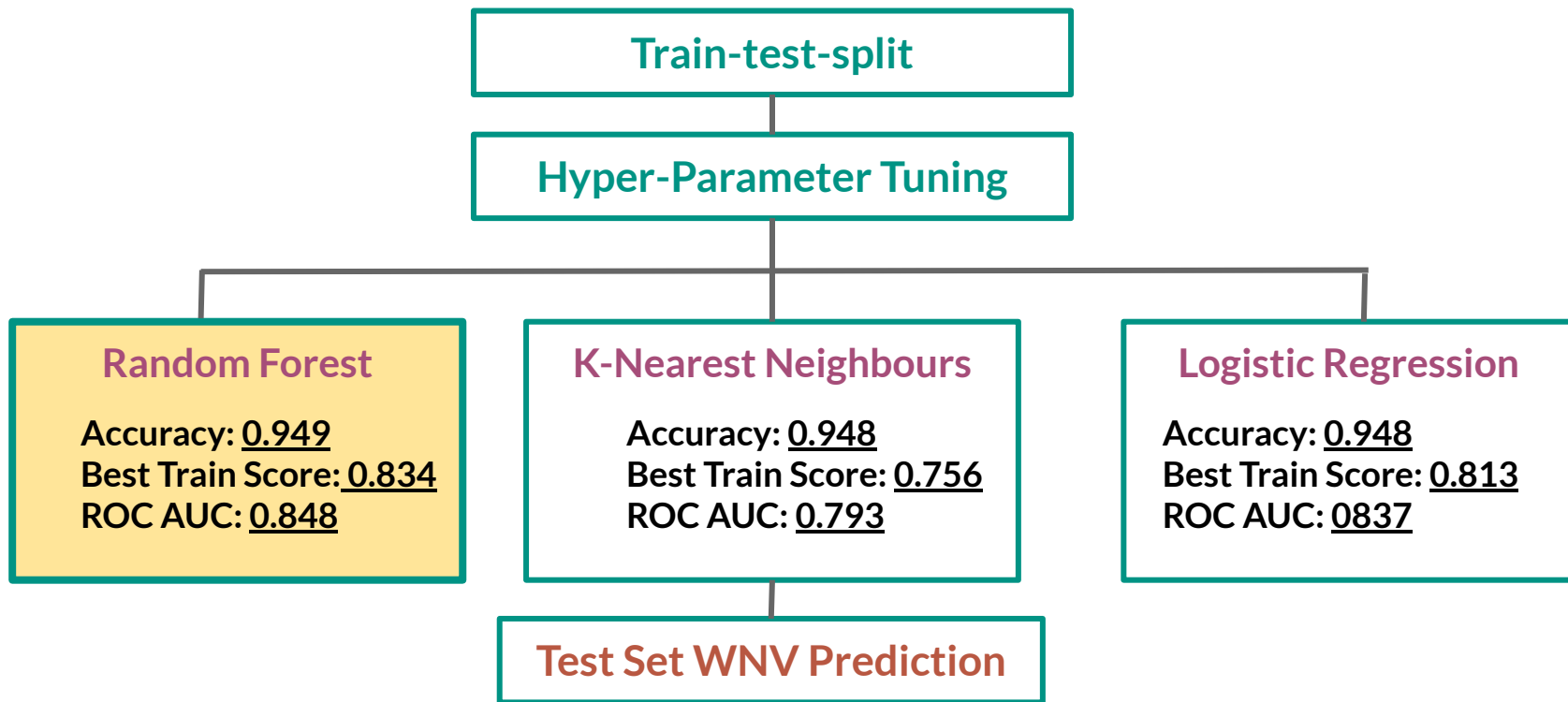
Finding #5: No lag between wind speed and number of mosquitoes



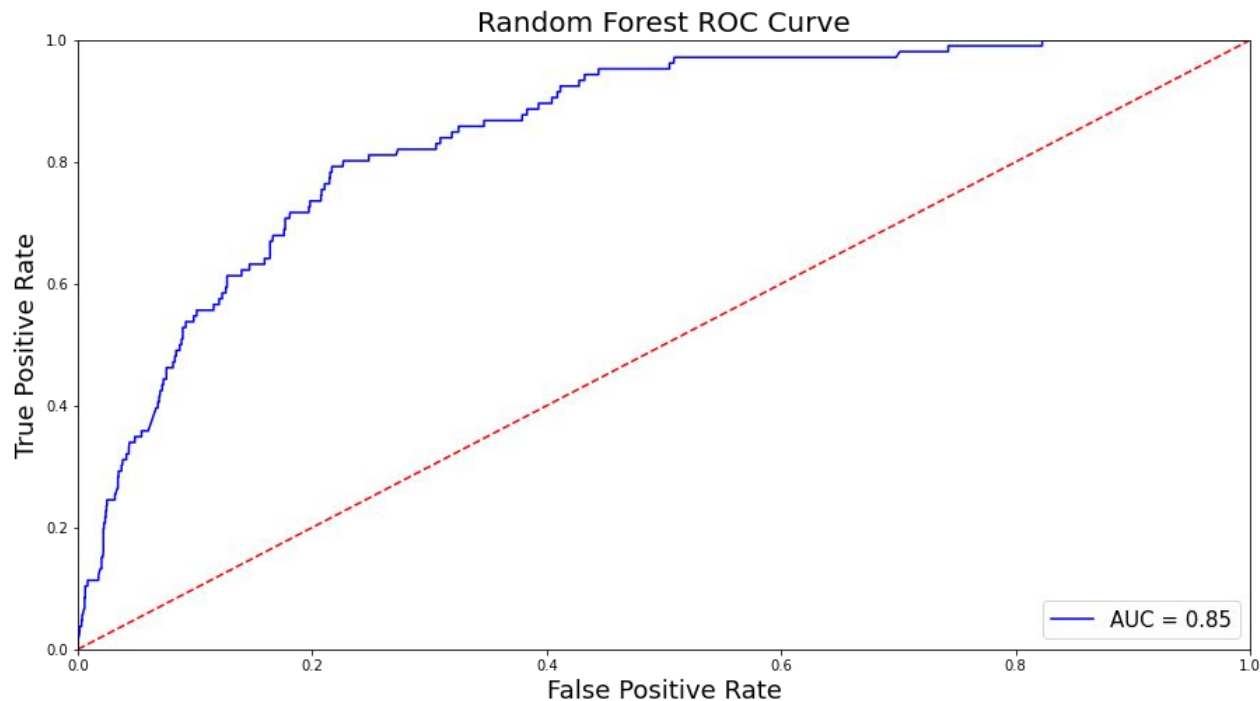
Modelling



Choosing the production model



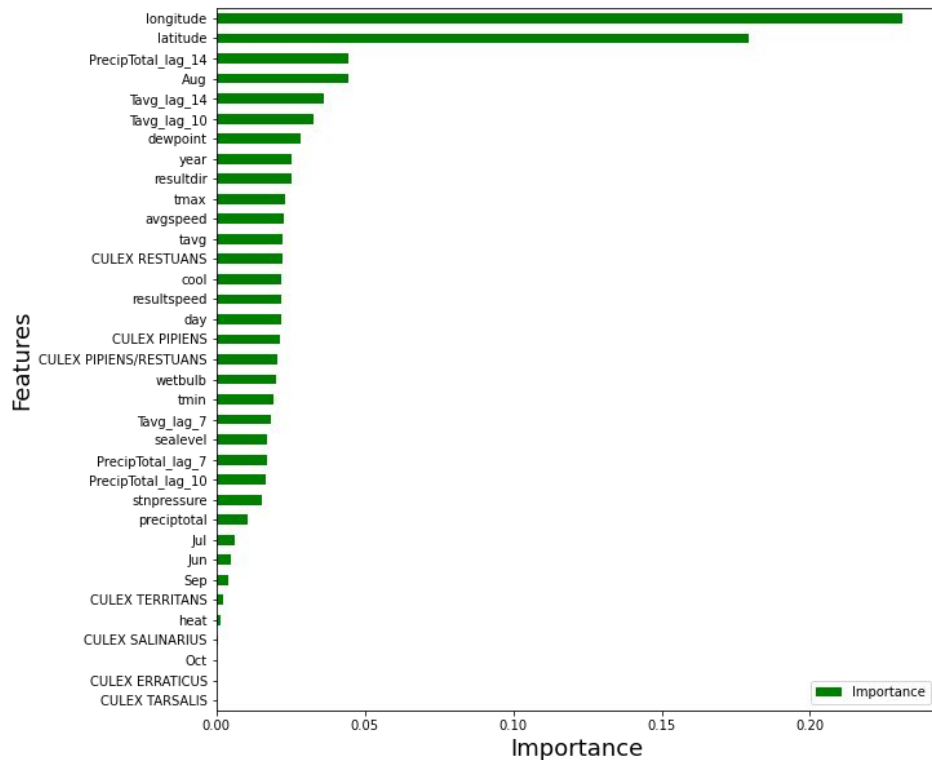
ROC Curve



	precision	recall	f1-score	support
0	0.95	1.00	0.97	1934
1	1.00	0.02	0.04	106
accuracy			0.95	2040
macro avg	0.97	0.51	0.51	2040
weighted avg	0.95	0.95	0.93	2040

Kaggle Score :
0.64626

Features Importance



Features	Importance
Longitude	0.231175
Latitude	0.179395
PrecipTotal_lag-14	0.044457
Aug	0.044378
Tavg_lag_14	0.036325
Tavg_lag_10	0.032645
dewpoint	0.028224
year	0.025359
resultdir	0.025208

Cost Benefit Analysis & Recommendations



Cost Benefit Analysis

How significant is the Cost?

- WNV hospitalizations cost \$778 million in U.S. 1999 - 2014
- Abatement areas have enough to contain mild outbreaks only
- Reduced surveillance activities due to lack of lab funding.

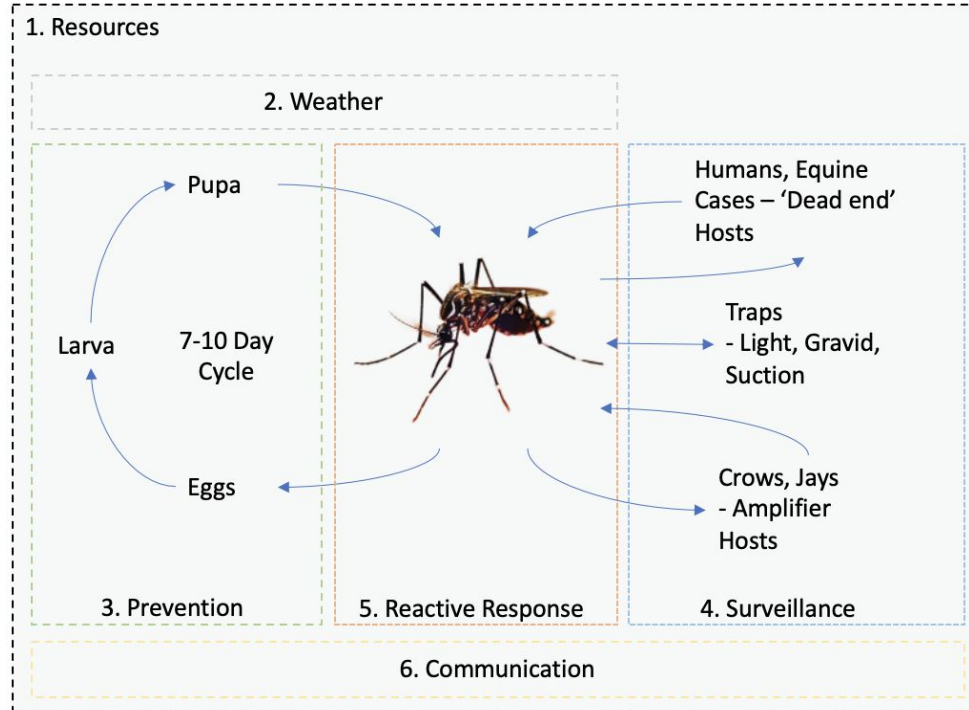
Spray Costs

- \$450,000 a truck spray session in 2002
- Adulticiding has an impact of only 1-2 days

Environmental & Health Cost

- Mixed reports on human and environmental impact even if its EPA approved
- At minimum fogging causes irritation to eyes

Strategy Recommendations



Further Improvement For the Model

Surveillance Data

- Dead Birds
- Wild Birds
- Human and Equine cases

Spray Data

- More spray and larvicidal data to determine effectiveness

Questions



References

J. E. Staples, M. Shankar, J. J. Sejvar, M. I. Meltzer, M. Fischer. Initial and Long-Term Costs of Patients Hospitalized with West Nile Virus Disease. *American Journal of Tropical Medicine and Hygiene*, 2014; DOI: [10.4269/ajtmh.13-0206](https://doi.org/10.4269/ajtmh.13-0206)

S, Hamill, <https://www.chicagotribune.com/news/ct-xpm-2002-09-09-0209090156-story.html>, Chicago Tribune, Viewed 10/7/2020