THE WEST NILE VIRUS: WHEN AND WHERE TO SPRAY

Philip Bradfield
Parisa Yarandi
Ritika Bhasker

Agenda

What is the problem?

Background and literature review

Our approach

Exploring the data

Modelling the data

Results and Recommendation

What is the problem?

West Nile Virus

Vector : Mosquitoes

Chicago: Mosquito traps

Data: Trap location, weather

Provide where and when

Reference:

Predicting *Culex pipiens/restuans* population dynamics by interval lagged weather data

Karin Lebl,[1] Katharina Brugger,[1] and Franz Rubel[1]

Author information ▸ Article notes ▸ Copyright and License information ▸

Abstract

Go to: ⌄

Background

Culex pipiens/restuans mosquitoes are important vectors for a variety of arthropod borne viral infections. In this study, the associations between 20 years of mosquito capture data and the time lagged environmental quantities daytime length, temperature, precipitation, relative humidity and wind speed were used to generate a predictive model for the population dynamics of this vector species.

Methods

Mosquito population in the study area was represented by averaged time series of mosquitos counts captured at 6 sites in Cook County (Illinois, USA). Cross-correlation maps (CCMs) were compiled to investigate the association between mosquito abundances and environmental quantities. The results obtained from the CCMs were incorporated into a Poisson regression to generate a predictive model. To optimize the predictive model the time lags obtained from the CCMs were adjusted using a genetic

Parasit Vectors

Our approach:

Model used: XGBoost classifier

Variables for the model: precipitation, wind speed, species, length of day, trap location (long./lat.), temperature variance and maximum temperature.

Target: If trap has a WNV mosquito - yes or no.

Locations of mosquito traps across Chicago. The size of the circle relates to the number of WNV mosquitoes captured.

© OpenStreetMap contributors
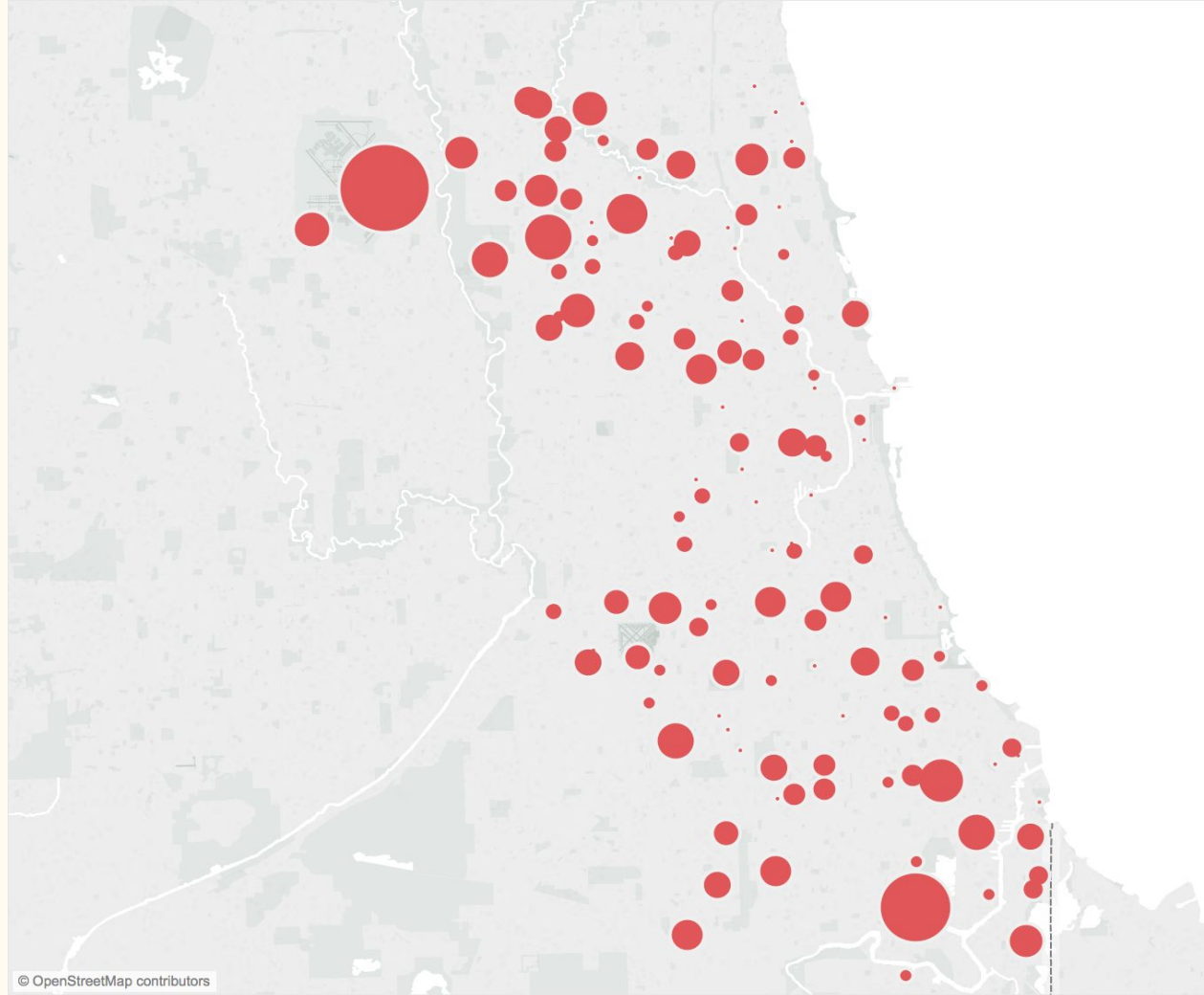
Data

Train

⟶ Data we used to make our model

Weather

Test ⟶ Data provided to test our model

Spray ⟶ Data we use for recommendation

Train Dataset

- The train dataset includes information about traps (address of the trap, species/number of mosquitoes and if West Nile Virus is present.)

Looking at Tableau:

Test Data

- Test dataset columns have also different Trap information includes (address of the Trap, Species of mosquitoes.
- Test dataset does not have any information for the number of mosquitoes.
- Test dataset has 8 type of species.

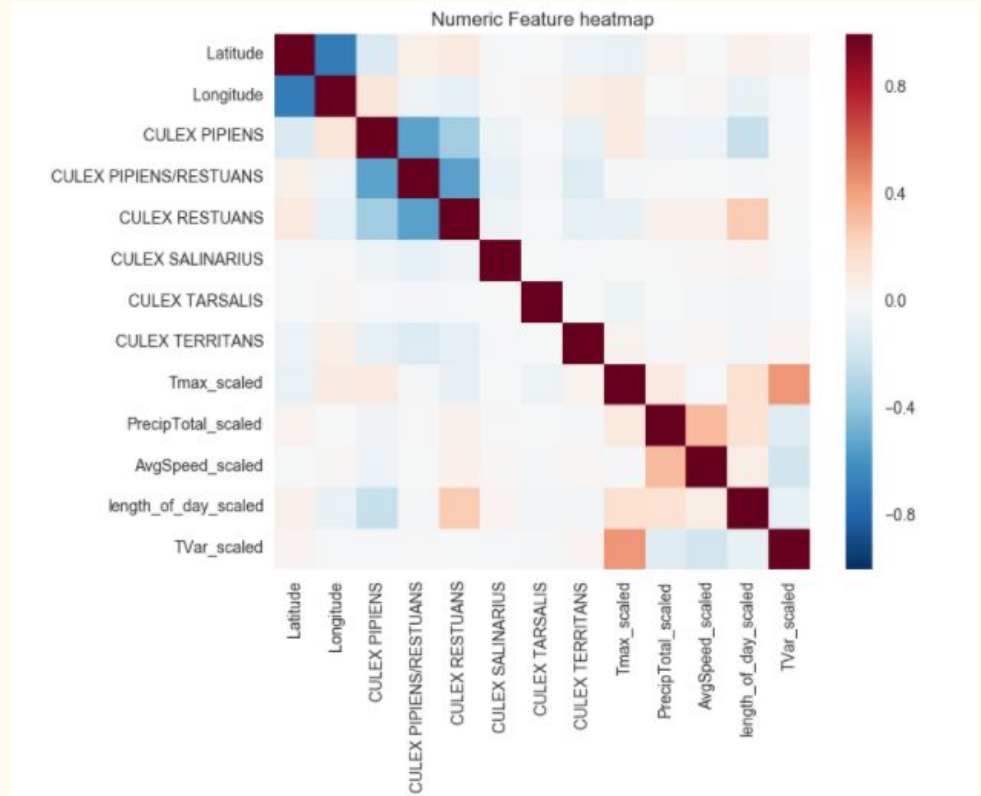| Species	
CULEX ERRATICUS	Abc
CULEX PIPIENS	Abc
CULEX PIPIENS/RESTUANS	Abc
CULEX RESTUANS	Abc
CULEX SALINARIUS	Abc
CULEX TARSALIS	Abc
CULEX TERRITANS	Abc
UNSPECIFIED CULEX	Abc

Data used to make a model

- ● Train + weather

Latitude, Longitude, Species, Max Temperature,

Precip, Wind Speed, Length of Day,

Temperature Variance

- ● Correlation matrix

Model selection

- Ran our training data through several classification models:
    - Logistic Regression, Random Forest, Gradient Boosting Classifiers, XGBoost
    - We split our training data into two sections: one to train the initial model, and another to test it (while we had a target we could test against).
- Found Gradient Boosting and XGBoost to be the best classifiers of when and where West Nile virus would be present.
- Decided to use both models.

Confusion Matrix

	Model predicted present	Model predicted not present
WNV Present	147	35
WNV Not Present	908	2377

Receiver operating characteristic for WNV detection
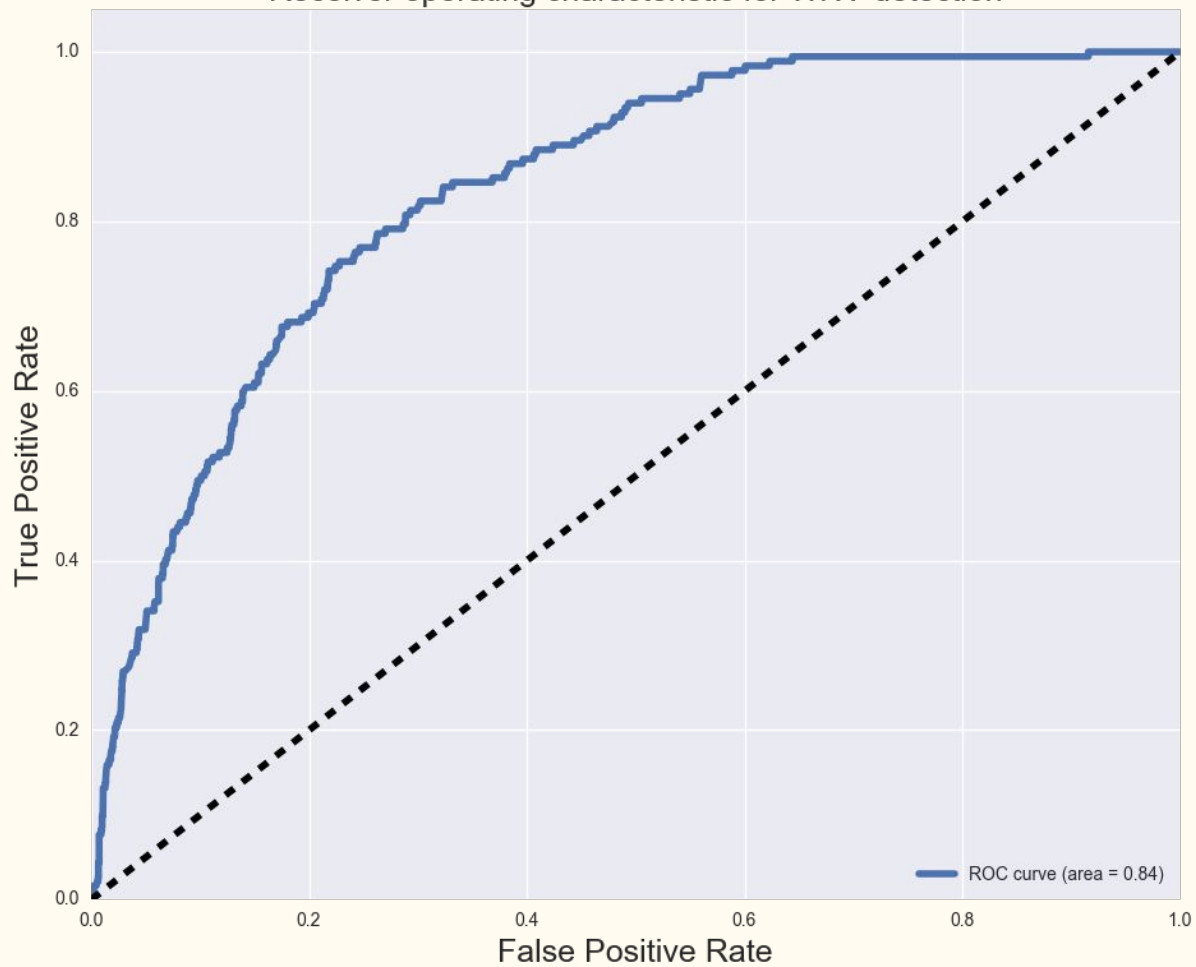
Feature Importance

- Length of day
- Location
- Average wind speed
- Maximum temperature

Recommendations

- Let's take a look at Tableau
- Future research could look into:
  - The effect of still water locations on mosquito populations (including lakes, ponds, sewers, etc.)
  - Number of mosquitoes in a given trap.
  - Looking at how spraying affected mosquito populations over time.
  - Looking at changing annual temperature and how it affects mosquito population density.

QUESTIONS?

How we did

577	↑156	comanda	0.72338	6	Wed, 17 Jun 2015 22:35:02 (-25.3h)
578	↑8	📊 LHY4	0.72336	2	Tue, 02 Jun 2015 07:01:34
579	↑65	klmno	0.72318	23	Wed, 17 Jun 2015 18:25:50 (-39.7h)
580	↑29	maveric	0.72294	12	Wed, 13 May 2015 06:19:36 (-5d)
-		R_B_	0.72243	-	Thu, 12 Jan 2017 22:20:56 **Post-Deadline**

Post-Deadline Entry
If you would have submitted this entry during the competition, you would have been around here on the leaderboard.

581	↓12	sidb	0.72241	7	Wed, 03 Jun 2015 16:59:56 (-12.1d)
582	↑23	Chinmay Jog	0.72240	28	Mon, 08 Jun 2015 10:38:14 (-3.2d)
583	↑38	Frank Schilder	0.72234	17	Tue, 16 Jun 2015 04:05:12 (-22.1d)
584	↓22	Ted Cybulski	0.72221	20	Thu, 14 May 2015 17:10:26 (-13.8h)
585	↓38	gooron	0.72210	36	Wed, 03 Jun 2015 21:14:28 (-47.4h)
586	↑37	Dipanjan Paul	0.72198	16	Wed, 17 Jun 2015 23:52:18 (-12.3d)