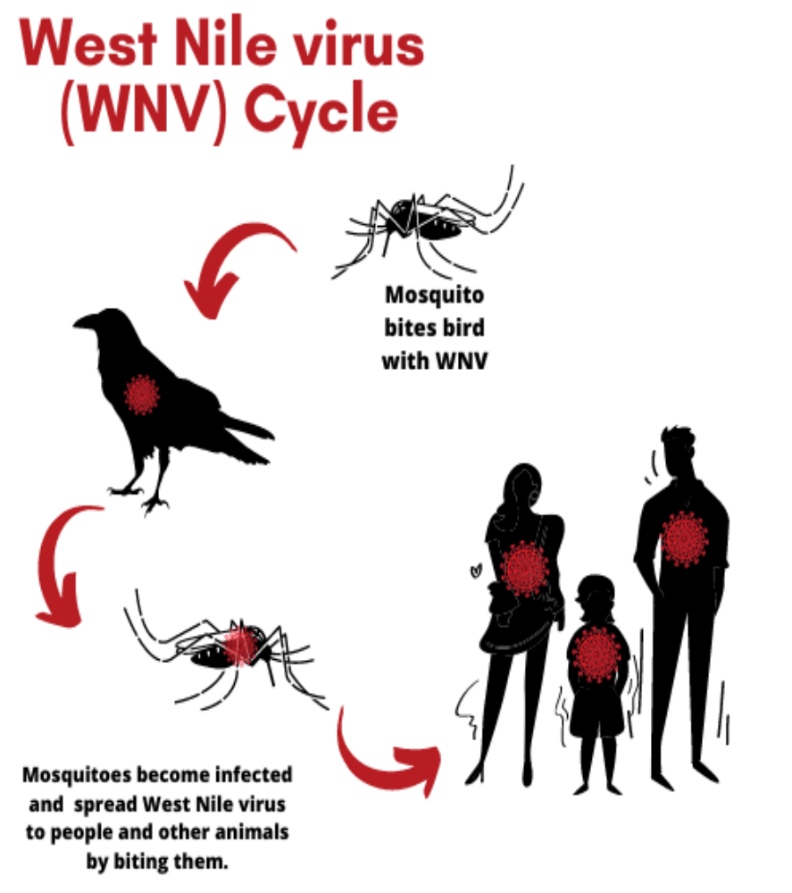
The Effects of Landscape Characteristics and Climate Variables on West Nile Virus Infection Risk in Massachusetts

Weizhi (Vanchy) Li, Florencia Sangermano

Department of Geography, Clark University

Introduction

West Nile virus (WNV) is a virus transmitted primarily by Culex mosquitoes, posing a persistent public health concern in North America (Allan, et al., 2009). Symptoms of WNV infection can range from fever to encephalitis and death. The intensity of transmission and the risk of endemic circulation of WNV depend on the abundance and distribution of infected mosquitoes, ecological conditions, and animal behaviors. The ecological conditions and landscape structure play a critical role in mosquito reproduction and activity (Boesing, et al., 2017). It is widely accepted that the species suitability area is closely associated with landscape characteristics (Pradier, et al., 2008) and environmental variables.



West Nile Virus Transmission Cycle
(source: orange county mosquito and vector control district)

The objective of this study is to identify areas suitable for infected mosquitoes and investigate how landscape structure and climate factors influence their distribution. The project aims to collect information that can aid in the development of intact ecosystems with reduced populations of WNV-infected mosquitoes, predict future mosquito distribution, and assist in controlling and mitigating the spread of WNV.

The study focuses on Massachusetts, where few studies have been conducted on WNV. Geographic and climatic factors such as landscape metrics at class and landscape level, NDVI, Land Surface Temperature, and Precipitation will be investigated at the town level in Massachusetts.

Data and Methods

Study Area and Dataset:

Massachusetts is located in the northeastern region of the United States and comprises 351 towns. Data on cases of infected mosquitoes were collected at the town level by the Massachusetts Department of Public Health for the years 2014 to 2020 (Fig.1).

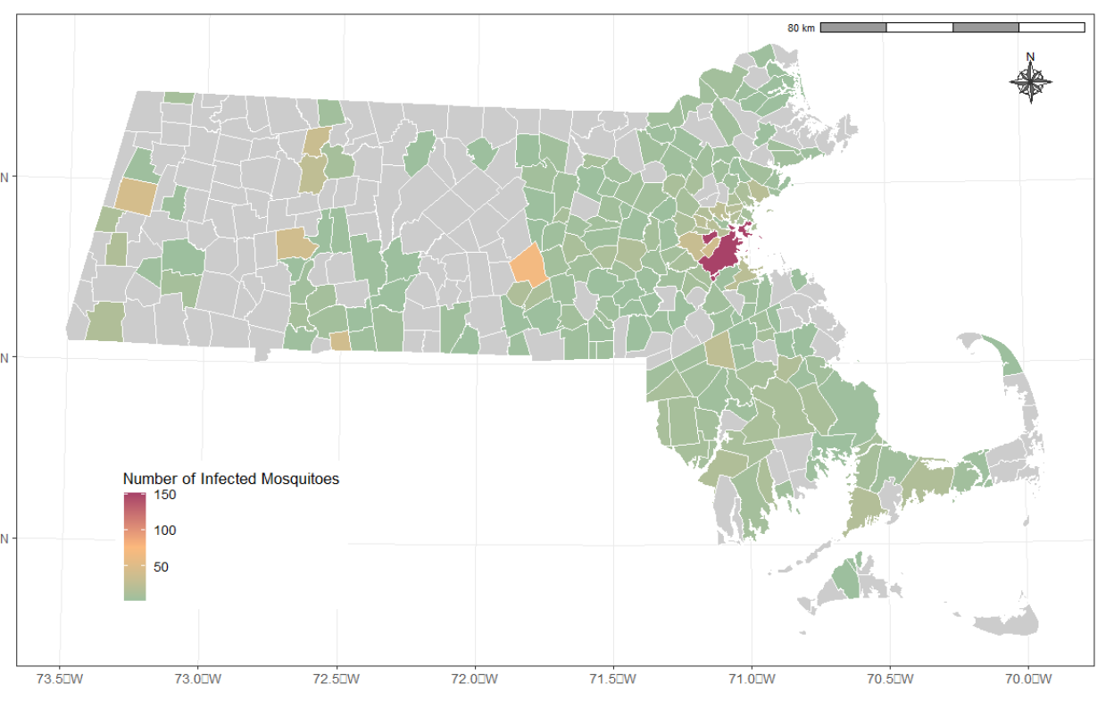


Fig.1. Massachusetts Town-Level WNV Infected Mosquito Cases, 2014-2020

The National Land Cover Database (NLCD) for the year 2019, was reclassified into 9 land cover types, including water, urban areas, forest, grassland, agricultural areas, and wetlands. This data was used to generate landscape metrics. Other environmental data included satellite images of land surface temperature (LST), precipitation, and NDVI (Fig.2), which are extracted during the summertime from 2014 to 2020.

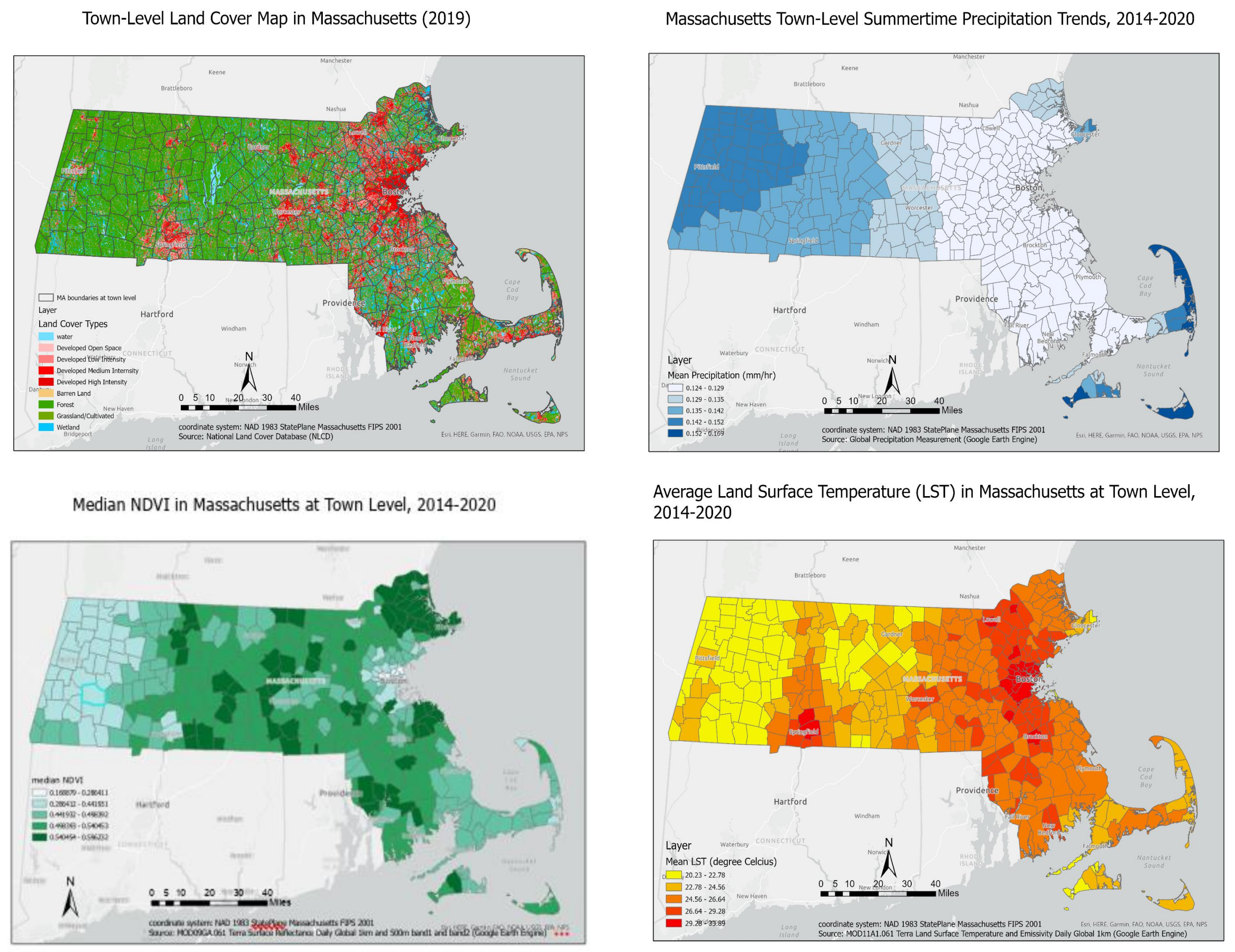


Fig.2. the land cover map in 2019 (upper left), the mean precipitation (upper right), median NDVI (lower left), and mean land surface temperature (LST) (lower right) during summertime in MA (2014-2020).

Data and Methods (Continued)

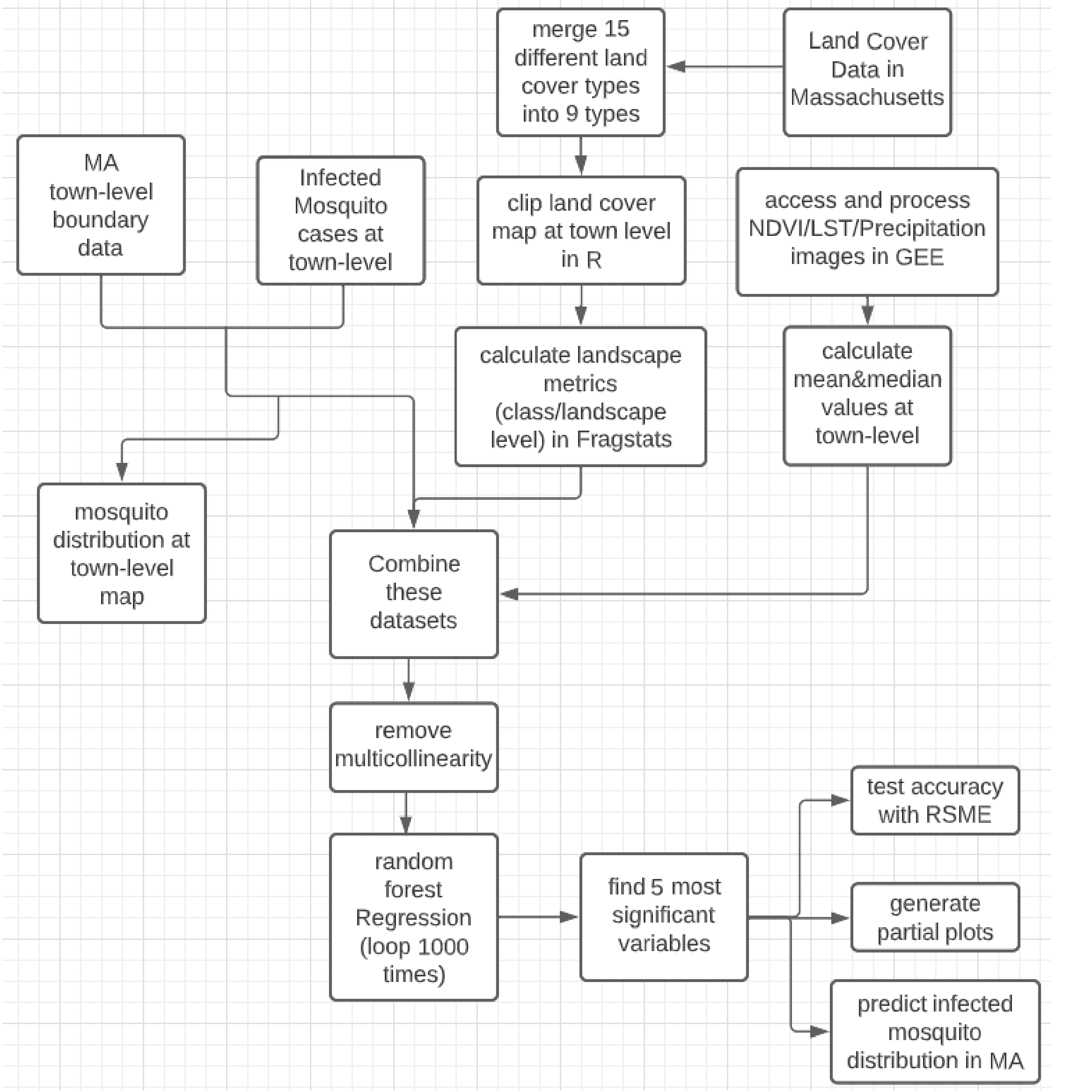


Fig.3. Conceptual Diagram of the project

Methods:

The Random Forest Regression Model was applied to identify the most important features that influence the distribution of infected mosquitoes in Massachusetts. The response variable was the number of infected mosquitos in each town, while the predictors included landscape and class metrics from the NLCD land cover, calculated using Fragstats software (McGarigal, K. 1995). Furthermore, mean and median NDVI, LST, and precipitation values were obtained for each town using Google Earth Engine.

These three datasets were combined into each town as variables for the Random Forest regression model, resulting in 210 variables. Multicollinearity was removed from highly correlated features, leaving 78 variables for the model. The Random Forest model was trained using a 70-30 training-testing split and trained 1000 times to identify the top 5 most significant and stable features for predicting infected mosquito cases at the town level.

Root mean square error (RMSE), which measures how much difference between the real value with the predicted value, was used to calculate the accuracy of the model, and partial dependence plots of the top 5 important features were generated to depict their relationships with the number of infected mosquitoes.

The final step is to plot the map of predicted infected WNV mosquito distribution at the town level in Massachusetts based on the random forest model generated by the 5 most important factors.

The goal is to identify the most significant factors that influence the distribution of infected mosquitoes and help inform mosquito control efforts in the region.

Results

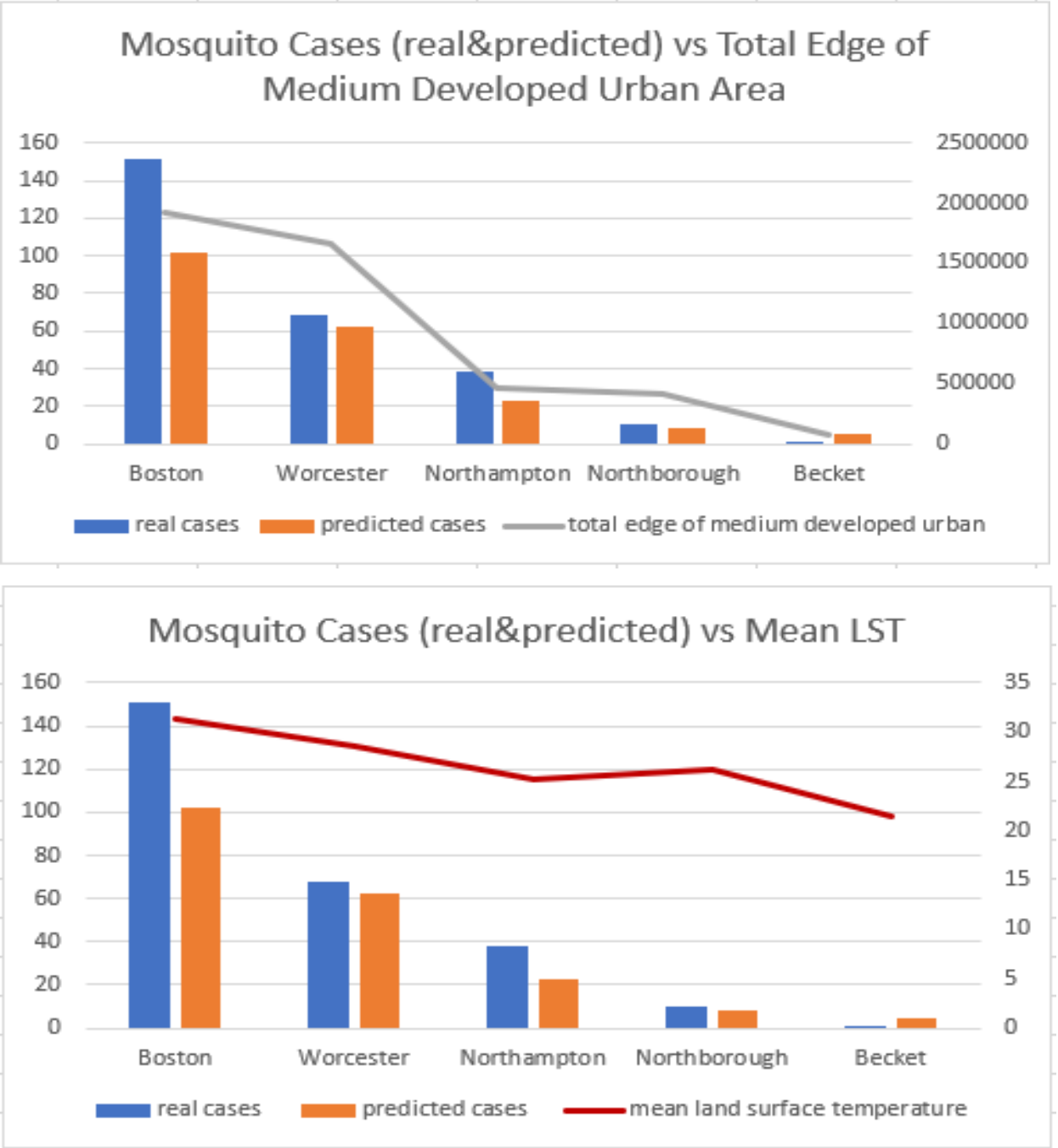


Fig.5(a). TE_MedUrb and AVG_LST vs. the Number of Mosquitoes (real and predicted) in 5 selected towns

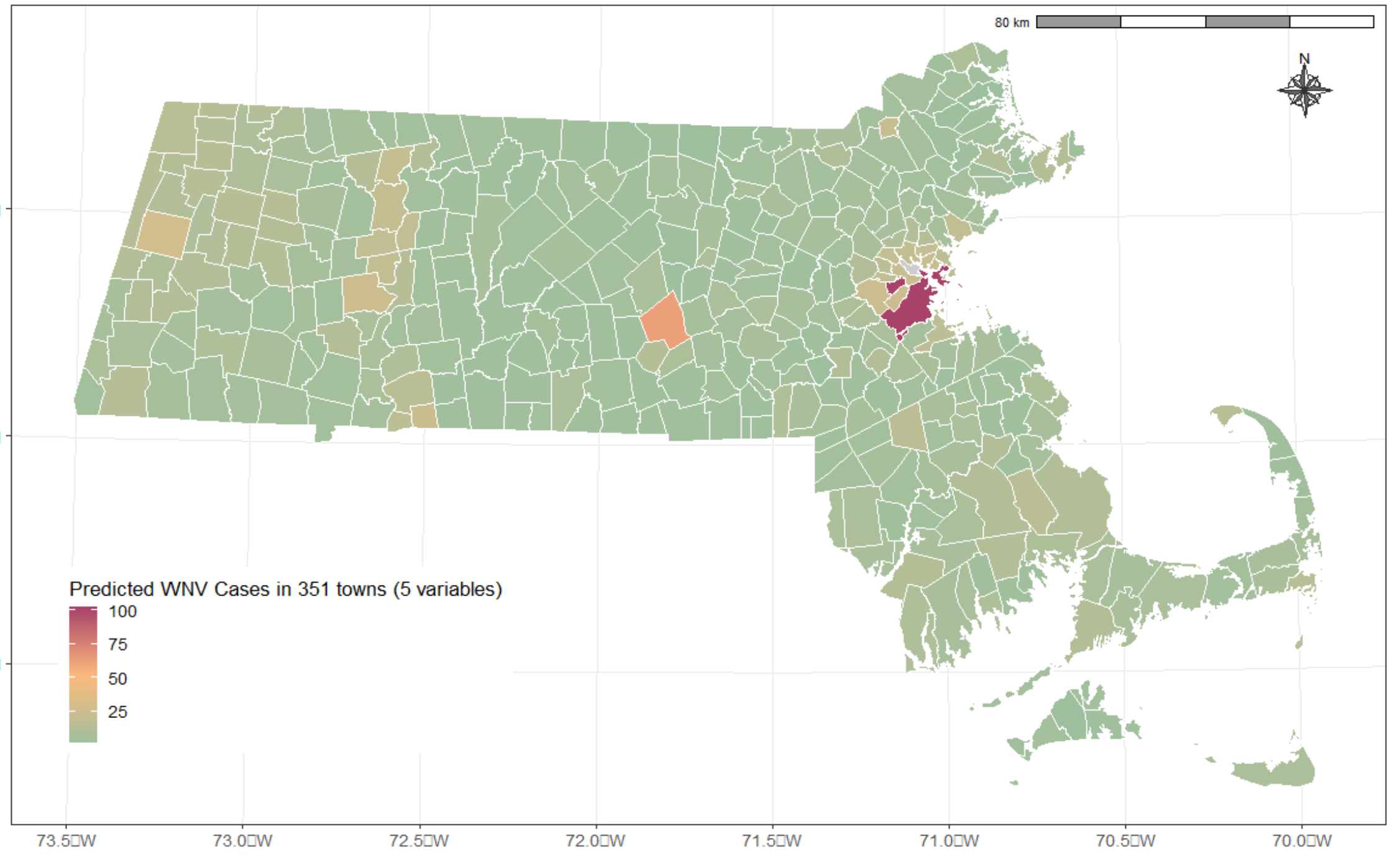


Fig.4. Predicted Infected Mosquitoes in MA at Town Level

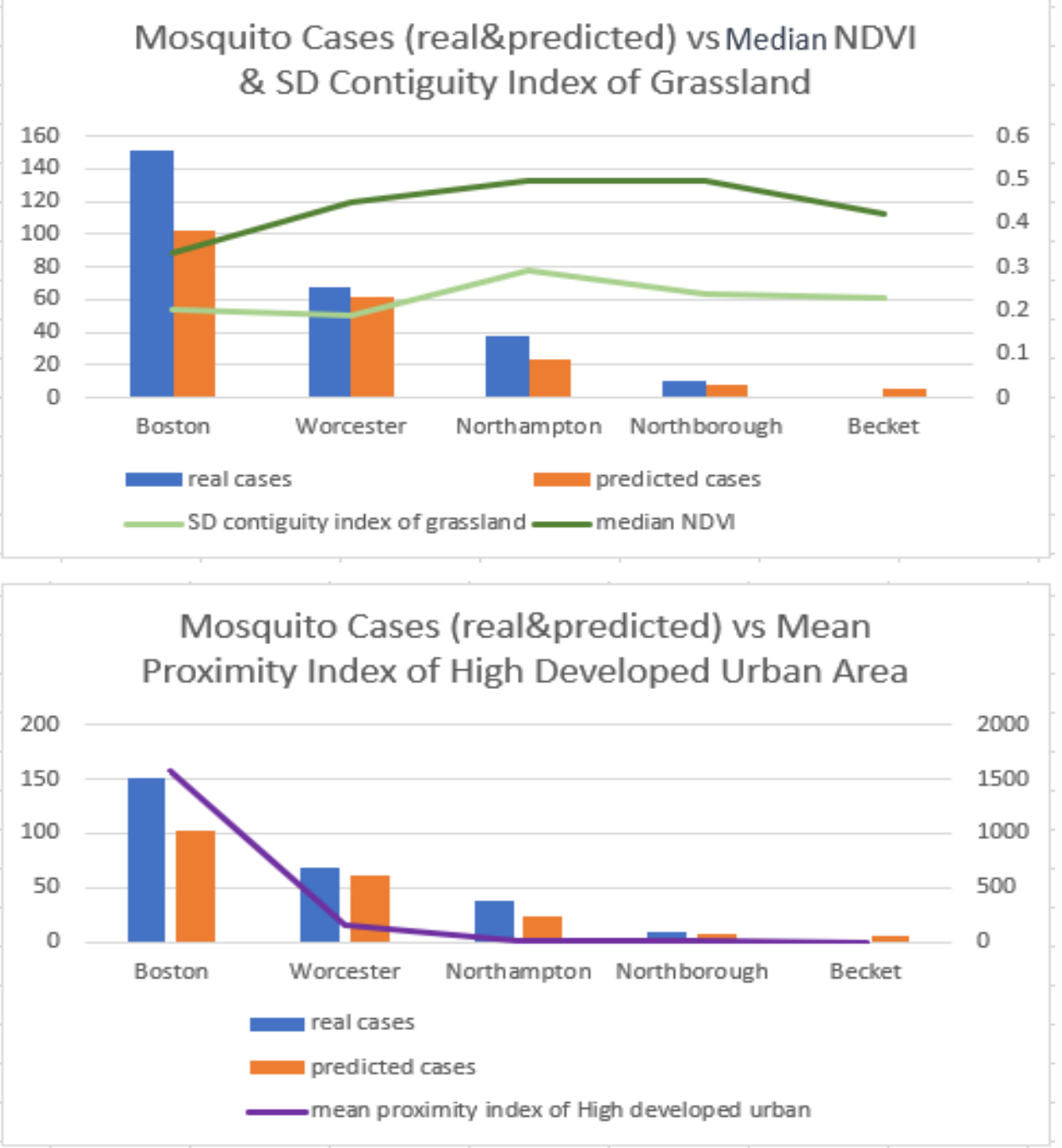
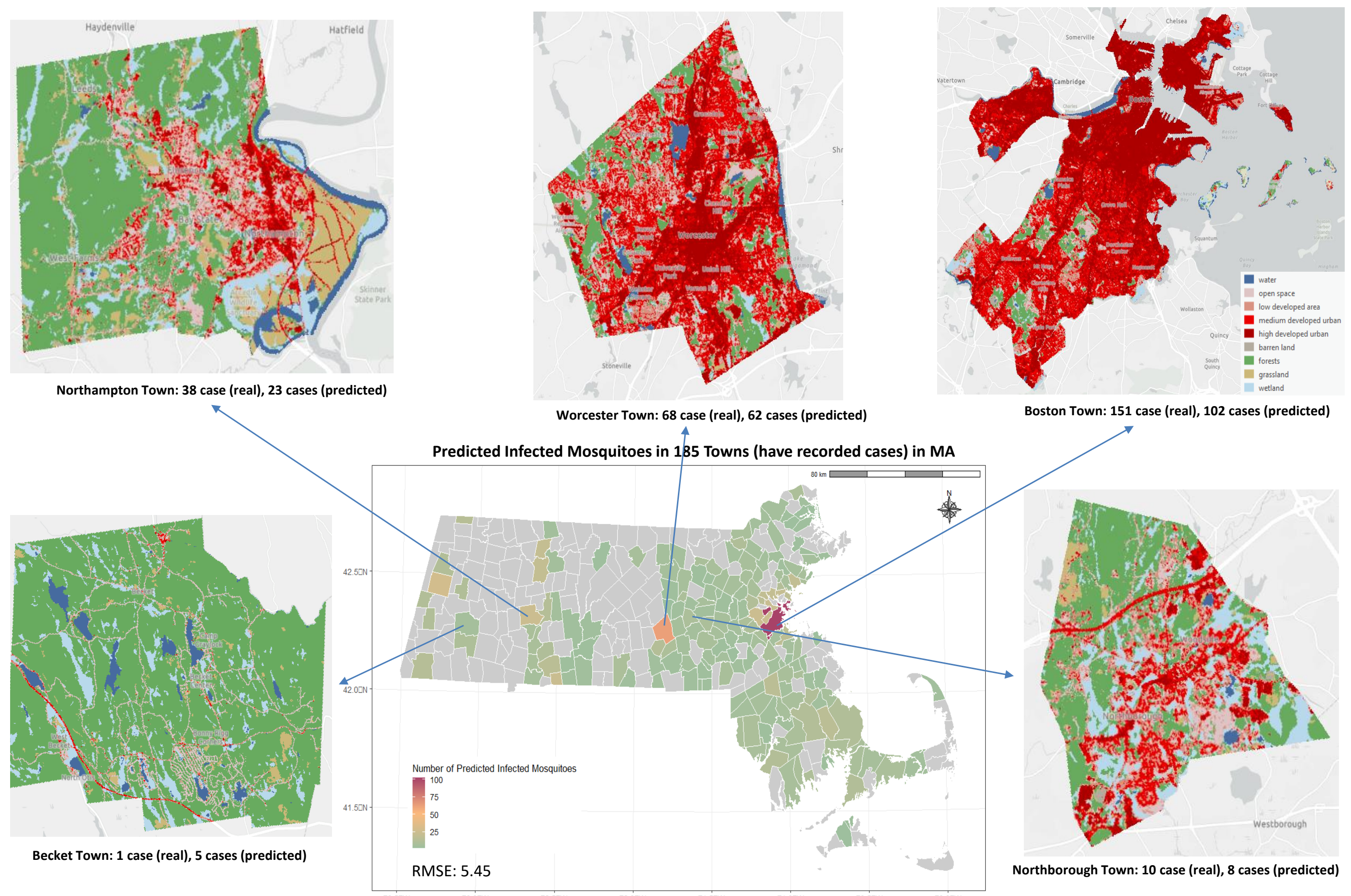


Fig.5(b). MED_NDVI, SD_CI_Grass, and AVG_PI_HUrb vs. the Number of Mosquitoes (real and predicted) in 5 selected towns

The original mosquito distribution map was limited to data from only 185 towns. Thus, the Random Forest model generated by 5 the most important features is used to create a predicted mosquito distribution map for all 351 towns in *Massachusetts (fig.4)*.

In the predicted map, Boston had the highest infection rate, with over 100 cases, represented in red. Surrounding cities and major towns in each county are shown in yellow, while towns with predicted low numbers of infected mosquitoes are shown in green. To better illustrate how landscape characteristics influence the number of infected mosquitoes, we extracted land cover maps from five towns with varying ranges of cases (fig.6).



Northampton Town: 38 case (real), 23 cases (predicted)

Worcester Town: 68 case (real), 62 cases (predicted)

Boston Town: 151 case (real), 102 cases (predicted)

Becket Town: 1 case (real), 5 cases (predicted)

Northborough Town: 10 case (real), 8 cases (predicted)

Fig.6. Predicted Infected Mosquitoes in 185 towns and the land cover maps of 5 towns (Northampton, Worcester, Boston, Becket, and Northborough)

The plots (fig.6) generated for each town provide a clear visualization of the relationship between the landscape and mosquito infection cases.

Towns with higher levels of urbanization, such as Boston and Worcester, are likely to have more infected mosquitoes, while those with more forest cover, like Northampton and Becket, may have fewer. In addition, grasslands are also a significant factor in mosquito detection. For instance, Northampton and Northborough have similar urbanization levels, but Northampton has a more varied size of grassland patches, which could explain its higher rate of infected mosquitoes compared to Northborough.

Fig. 6 shows the predicted distribution of infected mosquitoes in 185 towns as well, which closely matches the distribution observed in the map of real cases. The Root Mean Squared Error (RMSE) of the predicted distribution is 5.45.

Results (Continued)

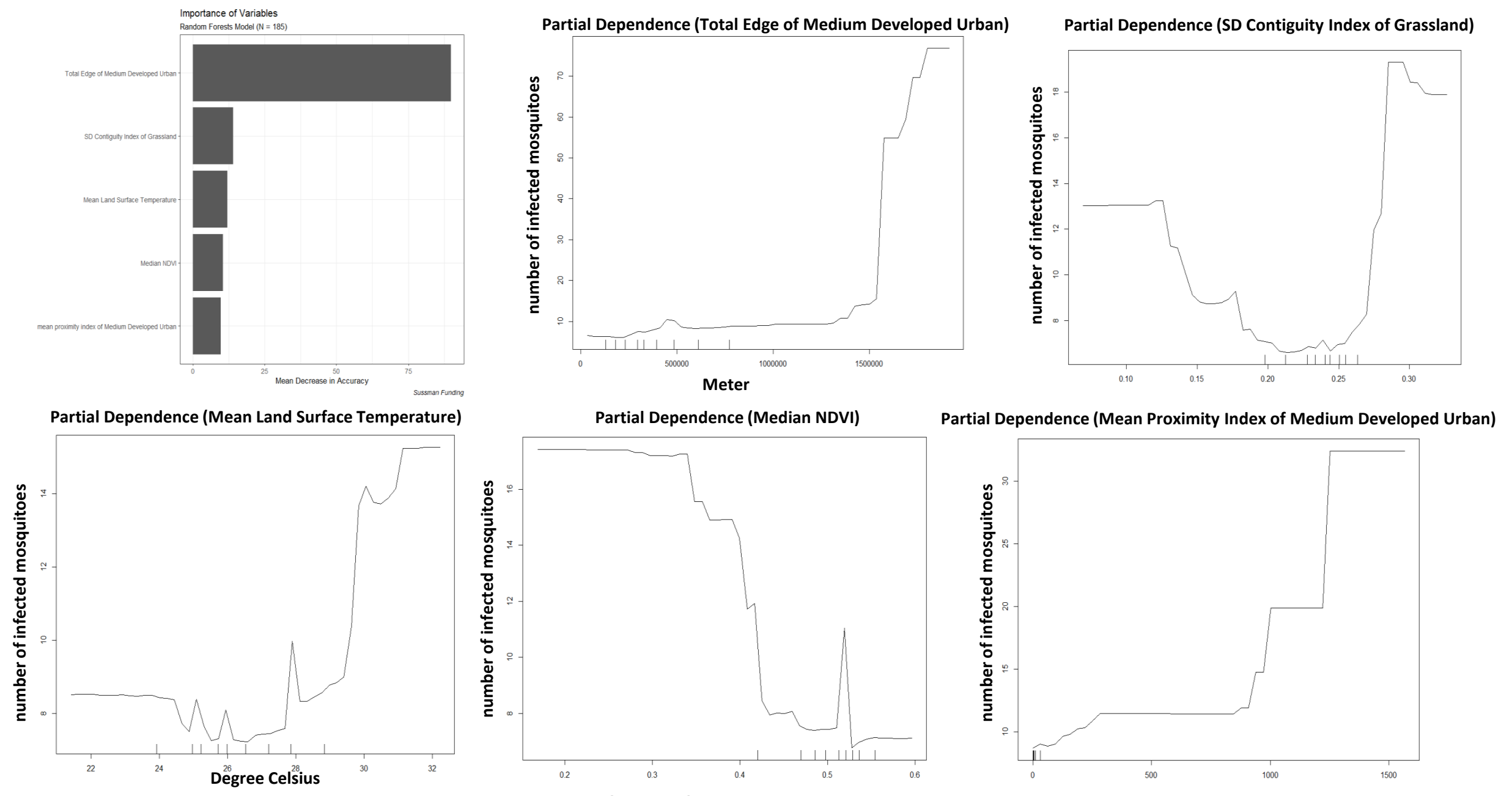


Fig.7. The Rank of Important Variables (left 1) and Partial Plots of 5 Topmost Important Features

Using random forest, the top 5 important environmental features were identified. They were the total edge of developed medium-intensity urban area (TE_MedUrb), the standard deviation of contiguity index of crops and grasslands (SD_CI_Grass), the mean land surface temperature (AVG_LST), the median NDVI (MED_NDVI), and the mean proximity index of developed high-intensity urban area (AVG_PI_HUrb).

TE_MedUrb is an absolute measure of the total edge length of the developed medium-intensity areas. SD_CI_Grass describes the variability in the connectivity of grassland patches. AVG_LST measures the mean LST temperature during the summertime from 2014 to 2020, which is obtained from MODIS Terra dataset. MED_NDVI is used to quantify the presence of living green vegetation in each town. AVG_PI_HUrb reflects the proximity of all developed high-intensity urban area patches whose edges are within a specific research radius of the focal patches (Gustafson, et al., 1992), and in our case the radius is 60m.

Discussion and Conclusions

Discussion:

According to the plots in fig.5 and partial plots of fig.7, the following findings were observed:

- TE_MedUrb: There is a positive correlation between the total edge of medium-intensity urban areas and the number of infected mosquitoes, especially in Boston and Worcester, where urban patches are more than other areas.
- SD_CI_Grass: The partial plot shows that as the variability in the connectivity of grassland patches increases, there are more infected mosquitoes. The dip in the relationship may be due to a lack of data on infected mosquitos at those SD values.
- AVG_LST: As the LST increases, the number of infected mosquitoes increases. When the temperature reaches 30 degree Celsius, the number of infected mosquitos reach to the top. This finding is consistent with the urban heat island effect (Paz, et al., 2008).
- MED_NDVI: There is a negative relationship between the median NDVI and mosquito cases (fig.7). Mosquito cases were high when NDVI was lower than 0.4 but decreased rapidly thereafter. This suggests that an increase in forest areas is related to a decrease in the number of infected mosquitoes.
- AVG_PI_HUrb: A positive relationship exists between the proximity index of developed high-intensity areas and the number of infected mosquitoes. More concentrated urbanization is associated with a higher number of infected mosquitoes, especially in areas near Boston.

Conclusion:

In this study, it was observed that mosquitoes infected with West Nile Virus are more commonly found in highly developed areas with irregular shapes, such as Boston. Furthermore, the closer together the high-intensity urban areas are, the higher the likelihood of having more infected mosquitoes. Additionally, it was found that infected mosquitoes tend to thrive in areas with higher temperatures, as indicated by LST. On the other hand, forests were found to play a role in reducing the spread of infected mosquitoes.

Ultimately, we believe that our project can help decision-makers to develop more targeted and effective interventions to reduce mosquito populations and prevent the spread of mosquito-borne diseases in affected areas.

Reference

Allan, B. F., Langerhans, R. B., Ryberg, W. A., Landesman, W. J., Griffin, N. W., Katz, R. S., ... & Chase, J. M. 2009. Ecological correlates of risk and incidence of West Nile virus in the United States. Oecologia, 158(4), 699-708.

Boesing, A. L., Nichols, E., & Metzger, J. P. (2017). Effects of landscape structure on avian-mediated insect pest control services: a review. Landscape Ecology, 32, 931-944.

Gökyer, E. 2013. Understanding landscape structure using landscape metrics. In Advances in landscape architecture. IntechOpen.

Gustarfson, E. J., G. R. Parker. 1992. Relationships between landcover proportion and indices of landscape spatial pattern. Landscape Ecology 7: 101-110

McGarigal, K. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure (Vol. 351). US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Paz, S., & Albersheim, I. (2008). Influence of warming tendency on Culex pipiens population abundance and on the probability of West Nile Fever outbreaks (Israeli case study: 2001–2005). EcoHealth, 5, 40-48.

Pradier, S., Leblond, A., & Durand, B. (2008). Land cover, landscape structure, and West Nile virus circulation in southern France. Vector-Borne and Zoonotic Diseases, 8(2), 253-264.

Acknowledgment

The Edna Bailey Sussman Foundation funded this project. Professor Florencia Sangermano is the faculty sponsor for this project. Dr. Paula Ribeiro Prist from the EcoHealth Alliance was the project supervisor. Both Dr. Sangermano and Dr. Prist provided helpful advice for this research. Dr. Karen Frey guided the application and reported the process of this internship.