Predicting Future Mosquito Habitats Using Time Series Climate Forecasting and Deep Learning

Christopher Sun, Jay Nimbalkar, Ravnoor Bedi

Partner

ABSTRACT

Mosquito habitat ranges have expanded globally due to climate change, introducing mosquitoes to new ecosystems and altering their prevalence in existing locales. The objective of this investigation is to analyze the preferred ecological conditions of mosquito larvae and forecast these environmental factors to predict likely future mosquito habitats. Publicly accessible atmospheric records and citizen-science mosquito larvae observations are used to compile a data set that includes ecological features of interest, such as temperature, precipitation, and elevation. The target variable is mosquito larvae abundance, as identifying these counts allows for effective prevention of potential outbreaks before infestation. A deep neural network consisting of six dense layers is trained on the data set to predict larvae counts from ecological inputs, with the oldest records set aside as test data for validation by backcasting. Subsequently, climate forecasting is conducted for each state in the contiguous US. Long Short-Term Memory networks are employed to forecast temperature and precipitation until the year 2050. These climate projections are then fed back into the deep learning model to generate mosquito larvae abundance predictions. Our model projects that by 2050, mountainous locales will contain elevated levels of mosquito larvae, more so than coastal and prairie regions. These results motivated a regional analysis of Texas due to the state's high ecological and climatological diversity and its steep elevation gradient. The model projects that between 2030 and 2050, mosquito abundance will increase at the fastest rate in high-altitude western Texas. This proves that high-altitude ecosystems will become better suited for mosquito breeding in the future as temperatures rise. The results of this investigation support our hypotheses of regional ecosystem-driven changes in mosquito distribution, allowing for proper prevention of mosquito outbreaks and containment of vector-borne disease transmission.

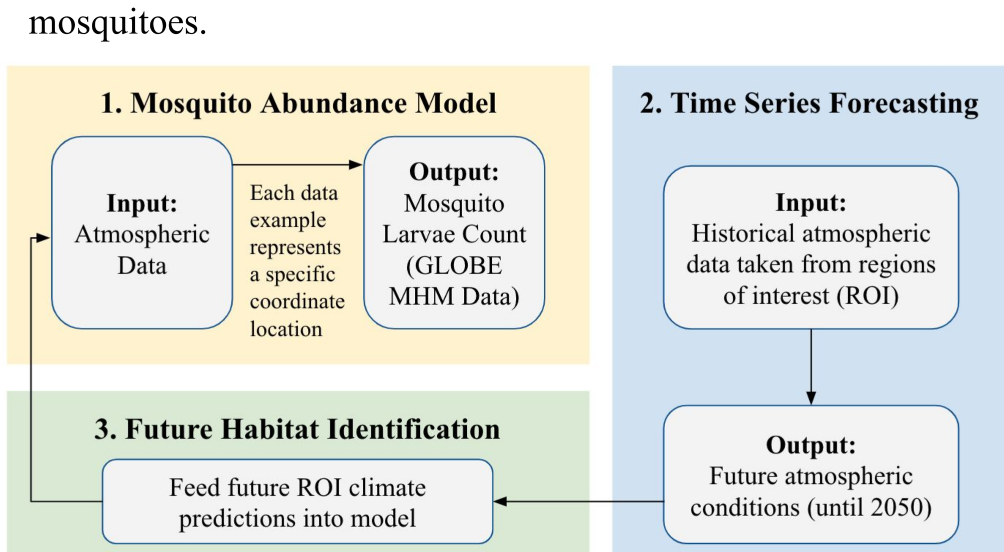
INTRODUCTION

Mosquito habitat and breeding ranges have increased globally. Mosquito habitat preferences are based on the interaction of several factors, including temperature, humidity, rainfall, elevation, and availability of hosts. Climate change has been identified as a key driving factor for the shifts in mosquito distribution over the past 70 years and is likely to continue to be the chief determinant of mosquito population spread. According to current trends, climate change will lead to major shifts in meteorological variables and land cover distributions, including an increase in average temperature, rising ocean levels, and increased severity of storms and droughts.

Using artificial intelligence, predictions can be scaled to adaptable national or global mosquito models to identify nuanced relationships between atmospheric variables and mosquito abundance. In this investigation, we harness deep learning to develop a mosquito abundance model and conduct time series climate forecasting in order to predict where future infestations of mosquito larvae may occur in the United States.

Research Overview

1. Gather meteorological data and mosquito larvae counts from various locations in the United States to create a predictive model for mosquito larvae abundance.
2. Extract time-series sequences of the said ecological variables from satellites for specific regions of interest, to allow for the forecasting of environmental conditions in these regions.
3. Pass these environmental predictions into the predictive model developed in (1) to obtain quantitative measurements of mosquito larvae abundance and to identify future breeding regions for mosquitoes.



Research Overview

METHODS

Data Collection

Sources:
- GLOBE Mosquito Habitat Mapper
- Weather Underground and Weather UX

Features:
- Average Daily Mean Temperature
- Average Daily Maximum Temperature
- Average Daily Minimum Temperature
- Monthly Days of Precipitation
- Average Daily Precipitation Amount
- Elevation
- Larvae Count (Target Variable)

Preprocessing:
- Filtering out artificial larvae observations
- Z score standardization
- Log_{10} transformation of larvae counts

1. Mosquito Larvae Abundance Model

A deep neural network was assembled for the prediction of mosquito larvae count (see below).

Architecture:
- 6 dense layers, each with 64 hidden nodes
- ReLU activation function
- Dropout regularization (20%)

Specifications:
- Adam optimization, Xavier initialization
- Mini-batch size = 8

Thirty-five of the oldest data examples were withheld as validation data to gauge whether the model could backcast previous mosquito larvae counts from historical ecological data.

2. Climate Forecasting

A Long Short-Term Memory network (see below) was trained on the historical temperature and precipitation sequences to conduct climate forecasting on selected regions of interest.

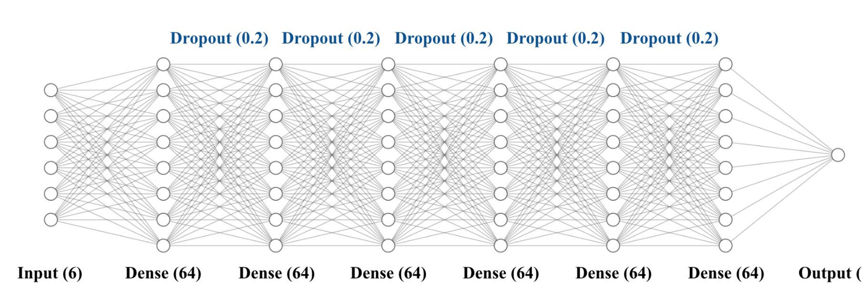
Architecture:
- 32 LSTM units
- Dense layer with 10 output nodes

Specifications:
- Lookback sequence: 20 years
- Prediction sequence: 10 years
- Same optimizer/initializer as Mosquito Larvae Abundance Model

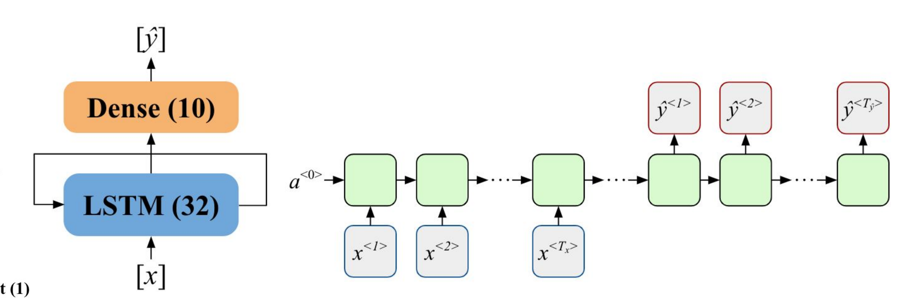
Algorithm 1 Climate Forecasting Process

Input: X ▷ sequence of m arrays of length l
Output: Y ▷ prediction sequence of m arrays of length $p \cdot t$

```
function FORECAST(X)
    for i ← 0 to m do
        x ← X_i
        Y_i ← []
        μ ← μ_x    σ ← σ_x
        x ← (x − μ)/σ
        for j ← 0 to t do
            y ← f(x)
            y ← y·σ + μ    x ← x·σ + μ
            Y_i ← Y_i || y
            x ← x[(p − l) : ] || y
            μ ← μ_x    σ ← σ_x
            x ← (x − μ)/σ
        end for
        Y ← Y||Y_i
    end for
    return Y
end function
```



Deep Neural Network Architecture



LSTM Architecture

Notation for Climate Forecasting

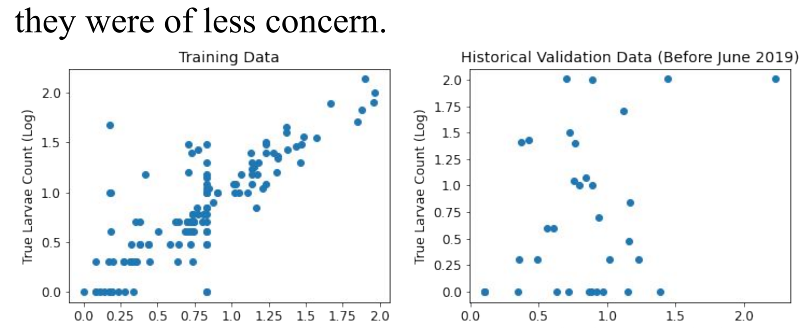
$m \leftarrow$ number of locations,
$l \leftarrow$ length of lookback sequence
$p \leftarrow$ length of prediction sequence
$t \leftarrow$ number of future sequences

$f: x \rightarrow y \leftarrow$ trained LSTM model
$X \leftarrow \{[X_{11}...X_{1l}], [X_{21}...X_{2l}], ...[X_{m1}...X_{ml}]\}$
$Y \leftarrow$ output array

RESULTS AND INTERPRETATION

1. Mosquito Larvae Abundance Model

The deep neural network was able to understand the intricacies of the training data but fell short when it came to generalization to unseen data. Though a moderate positive correlation existed between predicted larvae counts and ground-truth larvae counts on validation data, there were numerous instances of large negative residuals between these value, meaning the model tended to liberally flag locations as containing high mosquito abundances, when in truth they were of less concern.



	Train	Val
R	0.888	0.489
P	1.26E-45	1.44E-3

2. Climate Forecasting

Prior to using a Long Short-Term Memory network (LSTM) for time series forecasting, it was discovered that the trends in temperature and precipitation somewhat conformed to a pattern resembling the following periodic function, where T is the target atmospheric variable given the year t since the initial year t_0.

$$T(t) = \lambda t - e^{-\alpha t}\sin(\theta t)\gamma t^{\beta} + \phi. \quad (1)$$

The approximate parameters that can be used to estimate the trends are as follows:

$$\lambda \approx 0.01, \alpha \approx -0.01, \theta \approx 0.6, \gamma \approx 0.5, \beta \approx 0.03, \phi \approx T(t_0).$$

It was also discovered that minimum and maximum temperature shared a high correlation with mean temperature. In particular,

$$T_{min}(t) = T_{mean}(t) - k_{min}, \forall t \in \{t_i\}_{i=0}^{n} \quad (2)$$

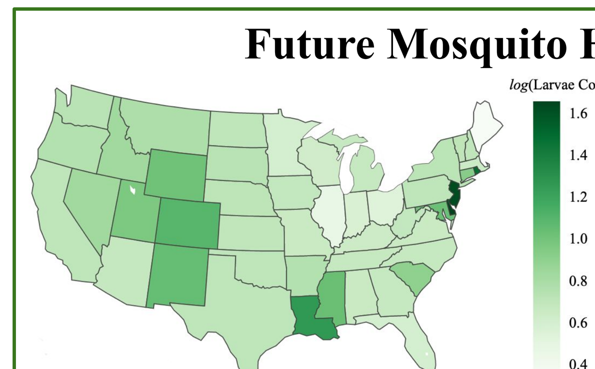
$$T_{max}(t) = T_{mean}(t) + k_{max}, \forall t \in \{t_i\}_{i=0}^{n}, \quad (3)$$

where k_{min} is a constant of adjustment between the minimum and mean temperature and k_{max} is a constant of adjustment between the maximum and mean temperature. Hence, the climate forecasting task for minimum and maximum temperatures was simplified into the following problem: Find k such that

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|T(t_i) - S_i|$$

is minimized, where $T(t_i)$ is the temperature predicted using the corresponding function above and S_i is the true temperature.

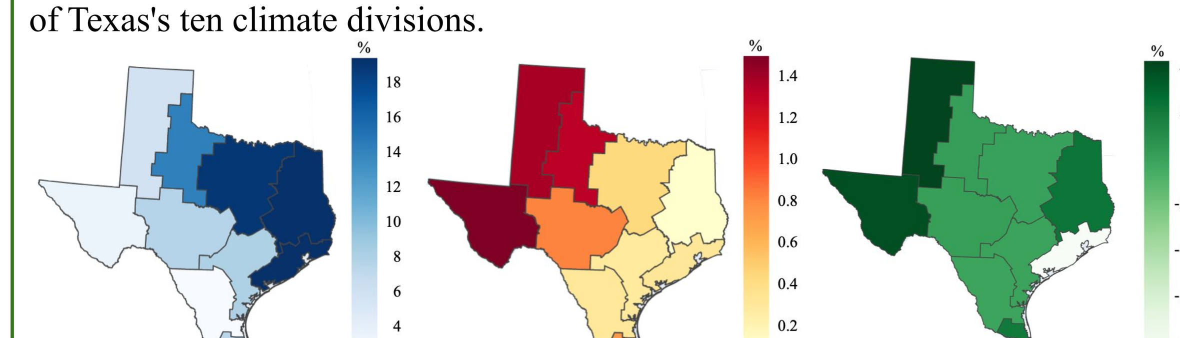
Future Mosquito Habitat Identification



Larvae Abundance by State in 2050

Our results show that by 2050, the US Mountain states will be likely locations of mosquito breeding. Namely, Colorado, Utah, Wyoming, and New Mexico will contain high larvae counts than neighboring Pacific Coastal states to the west as well as neighboring Great Plains states to the east.

These projections align with previous research findings that changing weather variability could shift mosquito species into higher elevations. Due to this observation, the state of Texas was selected for further analysis at the regional level, as Texas contains a steep longitudinal elevation gradient, ranging from the Rockies to the Gulf of Mexico coast. Climate forecasting was conducted for each of Texas's ten climate divisions.



Texas Meteorological and Larvae Abundance Changes (2030-2050)

Between 2030 and 2050, the highest increase in larvae abundance will occur in the high-altitude western divisions of Trans Pecos and High Plains. Precipitation is projected to increase more so in the eastern portion of the state compared to the southern and western portions. The opposite trend was observed for mean temperature. In fact, western Texas was projected to experience the highest rate of temperature increase. The likely relationship between these factors is that warmer temperatures in western Texas will intensify and prolong drought conditions, stalling precipitation over the next few decades, creating conditions more favorable for mosquito habitation.

CONCLUSIONS

This investigation aims to predict the abundance of mosquito larvae across the United States in the year 2050. To achieve this purpose, a data set consisting of citizen-collected mosquito larvae counts and several accompanying atmospheric and spatio temporal variables is compiled. Then, atmospheric variables are analyzed to identify the conditions most suited to the habitation of mosquito larvae using a deep learning framework. Next, these variables are forecasted using an LSTM model to project future climatic conditions. Finally, these atmospheric projections are inputted back into the original deep learning model to obtain the desired predictions.

The results from this experiment support the idea that mosquito spread is largely location and ecosystem-dependent, which points out the benefits of utilizing localized citizen-science observations and conducting regional examinations. One note of interest was that states along the Rocky Mountain Range, which contains some of the highest elevations around the country, were predicted to have the highest larvae abundance in 2050. This observation was further supported by the case study of Texas, which predicted the greatest change in larvae counts to occur in the high-altitude western region. These results clearly show that the greatest shifts in mosquito larvae abundances will occur in high-altitude locales, which is most likely occurring since the increase in temperature is rendering high-altitude regions warm enough for mosquito habitation for the first time. Precipitation and proximity to large bodies of water, however, did not appear to have a generalized correlation with larvae abundance, yielding varying larvae counts across the US. It is likely that unusually large larvae count observations in regions along the coastline were due to the presence of high population densities and levels of urbanization rather than due to the coastal location itself.

These findings point to the need for increased resource allocation to high-elevation areas to contain mosquito spread and vector-borne diseases since these locations are forecasted to become high-risk targets in the future. In addition, because there is only a meager presence of mosquitoes in high-altitude regions today, the awareness and containment protocols in these areas regarding mosquitoes are likely lacking, which may lead to greater future consequences if no actions are taken.

REFERENCES

[1] Colón-González, F. J., Sewe, M. O., Tompkins, A. M., Sjödin, H., Casallas, A., Rocklöv, J., ... & Lowe, R. (2021). Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multi-scenario intercomparison modelling study. The Lancet Planetary Health, 5(7), e404-e414.
[2] Environmental Protection Agency. EPA. Retrieved July 25, 2022, from https://www.epa.gov/climate-indicators/climate-change-indicators-west-nile-virus
[3] Amerasinghe, F. P., & Indrajith, N. G. (1994). Post-irrigation breeding patterns of surface water mosquitoes in the Mahaweli Project, Sri Lanka, and comparisons with preceding developmental phases. Journal of medical entomology, 31(4), 516-523.
[4] Joshi, A., & Miller, C. (2021). Review of machine learning techniques for mosquito control in urban environments. Ecological Informatics, 61, 101241.
[5] Elango, S., Ramachandran, N., & Low, R. (2022). Autonomous Mosquito Habitat Detection Using Satellite Imagery and Convolutional Neural Networks for Disease Risk Mapping. arXiv preprint arXiv:2203.04463.
[6] Low, R. (2018, December). Citizen Scientists as Community Agents of Change: GLOBE Observer Mosquito Habitat Mapper. In AGU Fall Meeting Abstracts (Vol. 2018, pp. IN22B-03).
[7] Local weather forecast, news and Conditions. Weather Underground. https://www.wunderground.com/
[8] WeatherWX.com, https://www.weatherwx.com/
[9] Huntington, J. L., Hegewisch, K. C., Daudert, B., Morton, C. G., Abatzoglou, J. T., McEvoy, D. J., & Erickson, T. (2017). Climate engine: Cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding. Bulletin of the American Meteorological Society, 98(11), 2397-2410.
[10] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, 18-27.
[11] Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. International Journal of Climatology, 33(1), 121-131.
[12] Kimura, M. (2021). Globe Data Dashboard. Retrieved July 19, 2022, https://piphi5-globe-data-dashboard-srcmain-8lo71r.streamlitapp.com/
[13] Chang, K., Chen, C. D., Shih, C. M., Lee, T. C., Wu, M. T., Wu, D. C., ... & Ho, C. K. (2016). Time-lagging interplay effect and excess risk of meteorological/mosquito parameters and petrochemical gas explosion on dengue incidence. Scientific reports, 6(1), 1-10.
[14] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
[15] Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 249-256). JMLR Workshop and Conference Proceedings.

ACKNOWLEDGMENTS

We would like to thank CSR, Texas Space Grant Consortium, NESEC, NASA, and SEES for providing funding for the program in which this research was conducted. We would also like to thank our mentor Uma Desai as well as Dr. Russanne Low and Matteo Kimura from the Earth Explorers Team.