

Modelling mosquito population in Alachua County

Joe Brew, Anthony Dennis, Paul Myers

54% of the variation in mosquito population in Alachua County can be explained by a model which takes into account minimum daily temperature and rainfall over a range of prior days.

Introduction

Precipitation and temperature variation are known causal predictors for the size of mosquito populations, and the associated human risk of infection with mosquito-borne illness. Many researchers have constructed complex mathematical models to predict both the magnitude of an area's mosquito population¹ as well as the risk for specific arbovirus.^{2 3 4} Though these models are useful for establishing the parameters by which certain mosquito populations grow and infect, their external applicability at the local level is low due to a diversity in species and climate conditions across national, state, and even county lines.

Given that a predictive climate-based modelling for mosquito populations in **north** Florida⁵ has never been carried out (to the authors' knowledge)⁶, the following represents a first attempt at using precipitation and temperature to estimate the size of the local mosquito population at any given time. We simulate more than 65,000 linear and 65,000 log-linear models using 148 trap collection dates of 6-10 traps over six summers. Our best result yields a model which explains 53.8% of the variation in mosquito population in Alachua County traps.

Methods

Data were collected on three key variables for the period of March 1st, 2008 through December 1st, 2013:

Trapped mosquitoes were provided by Clarke Scientific, in contract with the Florida Department of Health in Alachua County. These were collected and reported weekly every summer.

Daily rainfall was "webscraped" from www.wunderground.com using R⁷

Daily minimum temperature was acquired using the "weatherData" package in R⁸

Both rainfall and temperature readings come from the Gainesville Regional Airport, whereas the 10 mosquito trap sites are scattered throughout Alachua County (see Figure 1).

¹Annelise Tran et al. A Rainfall- and Temperature-Driven Abundance Model for Aedes albopictus Populations. International Journal of Environmental Research and Public Health (Impact Factor: 2). 01/2013; 10(5):1698-719.

²Ruiz, MO, et al. Local impact of temperature and precipitation on West Nile virus infection in Culex species mosquitoes in northeast Illinois, USA. Parasites Vectors 2010, 3:19

³Hosen MB, Morse AP. A weather-driven model of malaria transmission. Malaria Journal 2004, 3:32

⁴Beck-Johnson, LM et al. The Effect of Temperature on Anopheles Mosquito Population Dynamics and the Potential for Malaria Transmission. PLoS ONE 8(11): e79276.

⁵The Florida Medical Entomology Laboratory carries out extensive mosquito population surveillance using groundwater and other independent predictors in south Florida: <http://mosquito.ifas.ufl.edu/Index.htm>

⁶Important disclaimer: "authors' knowledge" is admittedly very limited

⁷R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL www.R-project.org/. Full code available upon request: Joseph.Brew@FLHealth.gov

⁸Narasimhan, Ran. Package 'weatherData' version 0.3 (2014-01-31)

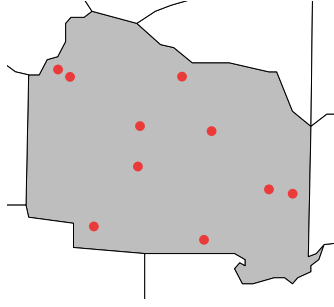


Figure 1: *Mosquito trap locations*

Model construction was approached in two distinct ways:

Linear: The first batch of simulations had the model parameters of a simple multivariable linear regression:

$$\hat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2$$

Where \hat{Y} = the number of mosquitoes per trap, $\beta_1 X_1$ = cumulative rainfall for the period of A through B, $\beta_2 X_2$ = the lowest minimum daily temperature for the period of C through D.

A, B, C, D (date ranges) were set as all combinations of 5 through 20 days prior to the date of trap collection. This produced 256 unique date ranges for both β_1 as well as β_2 . All potential combinations of β_1 and β_2 (256^2 or 65,536 unique models) were analyzed in a simple linear regression model, and retrospectively ranked for goodness of fit by their unadjusted R-squared values.

Log-Linear: The second batch of simulations was identical, but in this case the natural logarithm of the mean number of mosquitoes per trap was used as \hat{Y} instead of the simple mean number of mosquitoes per trap.

Results

Predictors

Linear

65,536 unique models were produced and compared for the linear approach, yielding R-squared values of 0.09 to 0.43 for the linear models (Figure 2).

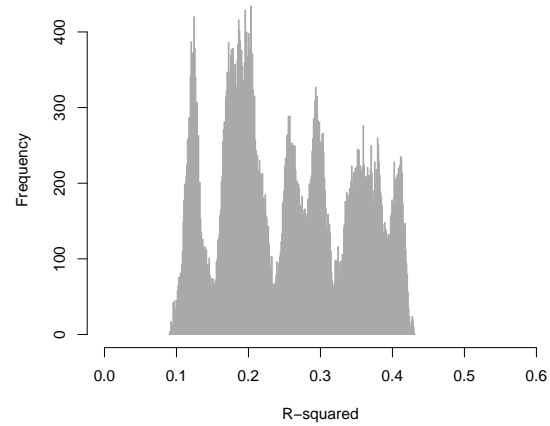


Figure 2: *Distribution of linear model R-squared values*

Of the linear models compared, the one with the best R-squared value (0.43) contained the following estimates:

β_1 = cumulative rainfall for the period 16-36 days prior to trap collection

β_2 = minimum temperature in the period 16-31 days prior to trap collection

The resulting complete formula was:

$$\hat{Y} = -234.105 + \beta_1(53.74) + \beta_2(3.496)$$

Cumulative rainfall for the simulation-determined best model date ranges exhibited a linear bivariate association with mosquitoes per trap.

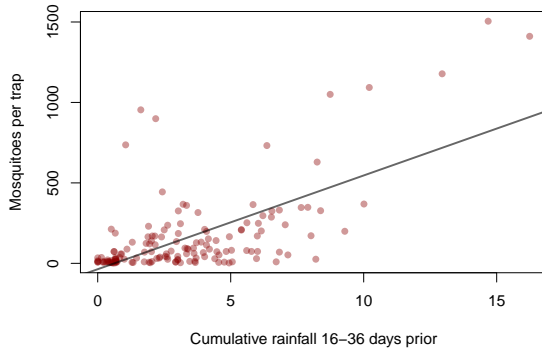


Figure 3: *Rainfall and mosquitoes per trap*

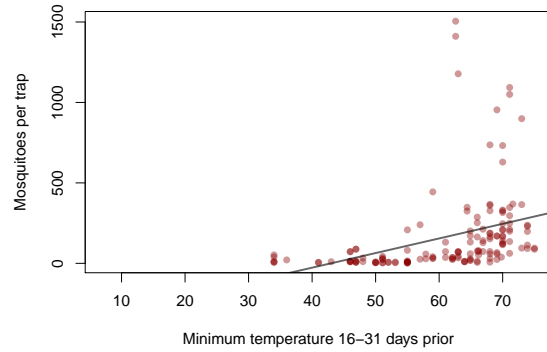


Figure 4: *Minimum temperature and mosquitoes per trap*

Though minimum temperature for the best model date range exhibited a more exponential shape, transformation yielded only marginally better results. On the other hand, tests yielded a significantly better fit when log-transforming the outcome variable (mosquitoes per trap) instead of the predictor.

The linear approach underestimated peaks and produced many nonsensical negative estimations, particularly in winter months. Though imperfect, this model does explain 43% of the variation in mosquito population:

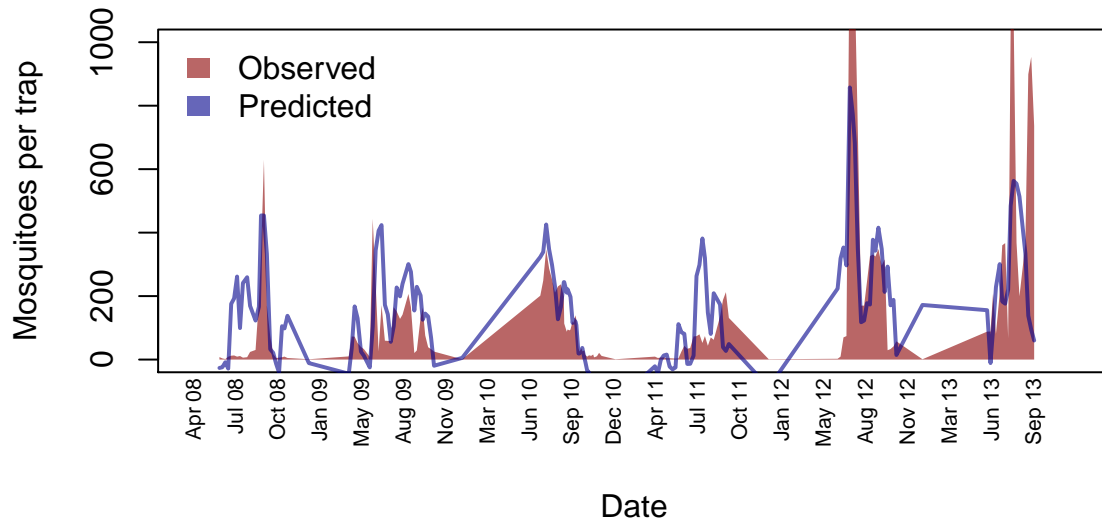


Figure 5: *Time series: predicted versus observed (best linear model)*

Log-Linear

65,536 unique models were also produced and compared for the log-linear approach, yielding R-squared values of 0.30 to 0.54 (Figure 3).

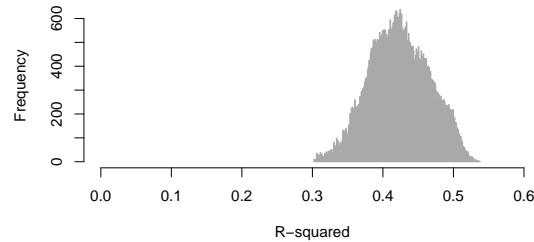


Figure 6: *Distribution of log-linear model R-squared values*

The best log-linear model (R-squared = 0.54) contained the following estimates:

β_1 = cumulative rainfall for the period 17-37 days prior to trap collection

β_2 = minimum temperature in the period 14-32 days prior to trap collection

The resulting complete formula was:

$$\hat{Y} = \exp(-1.121 + \beta_1(0.200) + \beta_2(0.074))$$

Cumulative rainfall for the simulation-determined best model date ranges exhibited a linear bivariate association with the natural logarithm of mosquitoes per trap, though it was heteroskedastic and could perhaps benefit from a log-log transformation.

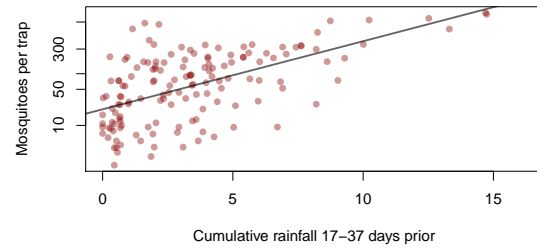


Figure 7: *Rainfall and mosquitoes per trap*

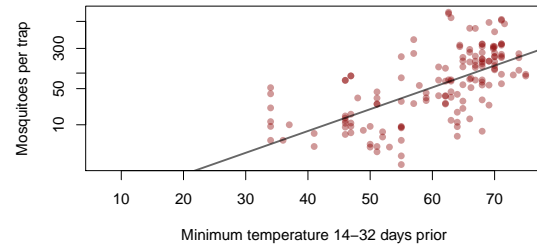


Figure 8: *Minimum temperature and mosquitoes per trap*

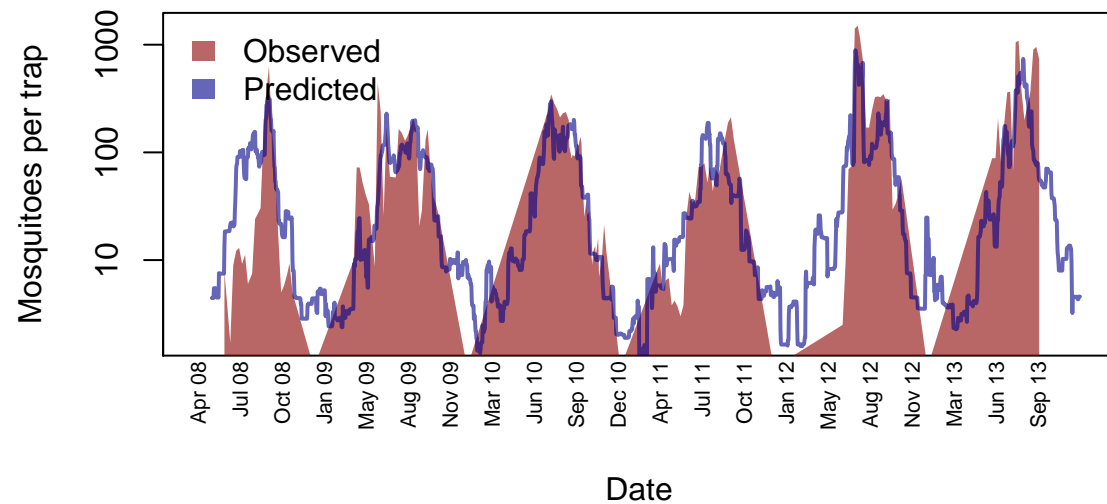


Figure 9: *Time series: predicted versus observed (best log-linear model)*

Discussion

Though not nearly as robust as more complex models, which take into account specific species' lifecycles and mating habits as a function of weather (and at times yield R-squared values as high as 0.8)⁹, simple linear modelling performed fairly well in terms of predicting general trends. It correctly predicted the large peaks in mosquito population in both 2012 and 2013 (with some slight underestimation), as well as the smaller peaks in other years.

A log-linear model represented a significant improvement, though further variable-specific transformation might yield even better results.

Of particular note, this model is valuable in that its independent variable range minimum values are 14 days prior to trap collection. In other words, mosquito population spikes can be predicted at least 14 days in advance, granting public health authorities time both to inform the population and take precautionary methods (spraying, etc.).

Further research

To fully understand the dynamics of mosquito populations in North Florida, further study is needed. The population numbers of vectors of specific emerging arboviruses (such as Dengue, Chikungunya and West Nile Fever) should be individually modelled so as to better understand the threat therein. Collaboration with entomologist and arbovirus specialists could lead to the inclusion and refinement of further potentially predictive variables. Finally, trap data should be shared among counties so as to ensure more precise modelling through larger sample size.

Conclusion

Given the arrival of new mosquito-borne illnesses in Florida, and the rapid changes in the epidemiology of arbovirus in the south, local governments should pay close attention to the threat posed by summer spikes in mosquito populations. Public health authorities in Florida can benefit from predictive modelling to assess and control the threat of spikes in mosquito population weeks in advance.

⁹Gong Honfei et al. A Climate Based Mosquito Population Model. Proceedings of the World Congress on Engineering and Computer Science 2007 WCECS 2007, October 24-26, 2007, San Francisco, USA