Put Title Here

Anne Marie Saunders
Imperial College London
MSc Computational Methods in Ecology and Evolution

August 27th, 2020

Declaration

This is the declaration.

Abstract

This is the abstract

Contents

Introduction

Introduce

At the end of the 20th century, the world experienced a global resurgence of vector-borne disease (Gubler, 2001). Diseases that had been well-controlled in the early-to-mid 1900s surged under complacent public health policies and insufficient research funding (Gubler, 1998). Today, vector-borne diseases represent 17% of the total global infectious disease burden and cause millions of deaths annually. Mosquito-borne diseases, such as malaria, dengue, Zika, yellow fever, West Nile virus, and chikungunya, singularly infect more than an estimated 200 million individuals worldwide every year (WHO, 2017).

Surveillance of mosquito populations is a successful method to control the public health impacts of vector-borne disease (Vazquez-Prokopec et al., 2010). Intervention in growing populations through chemical control measures can effectively reduce disease incidence (Tomerini et al., 2011). Sampling methods are, however, often limited by resource constraints (Sedda et al., 2019). Many studies have attempted to develop early warning models of disease incidence through prediction, rather than surveillance, of mosquito abundance (Beck-Johnson et al., 2013; Li et al., 2019; Poh et al., 2019). Because localised arbovirus disease data is often lacking in quality, mosquito abundance models can offer an alternative method to estimate disease risk (Lowe et al., 2013). Predictive models of mosquito abundance could also offer a cost-effective strategy with which to plan control measures (Yang et al., 2009).

Mosquito abundances are affected by many factors, including land-use, elevation, and vegetation cover, but meteorological variables, such as temperature and precipitation, are particularly predictive of population dynamics and commonly used in abundance models (Trawinski and MacKay, 2008; Li et al., 2019; Wang et al., 2011; Yoo et al., 2016). The effects of temperature on mosquitoes are well characterised. As ectotherms, mosquito life history traits such as development rate, biting rate, fecundity, and survival, are temperature dependent and vary with changing temperatures (Mordecai et al., 2019). With increasing temperature, trait performance will increase until an optimal temperature is reached, after which, trait performance decreases according to physiological constraints (Amarasekare and Savage, 2012). Because these traits shape reproductive output, trait variation determines abundance of mosquito populations (Cator et al., 2020). Precipitation affects mosquitoes in more complex ways. Rainfall can create larval habitats in man-made containers or expand natural pooled breeding habitats (Keith, 2005; Koenraadt and Harrington, 2008). Inter-seasonal variability in larval carrying capacity, which is dependent on rainfall for habitat creation, has been shown to be a main driver of mosquito abundances (Marini et al., 2016). Heavy rainfall, however, can flush immature mosquitoes from aquatic habitats, but the extent of this effect varies among species (Koenraadt and Harrington, 2008; Paaijmans et al., 2007). The effect of droughts on abundance is mixed; in some cases, drought is thought to benefit mosquito abundance by eliminating predators of larvae in drying water bodies (Chase and Knight, 2003). Other research have found that droughts simply increase sample collection, rather than true abundance of mosquitoes (Shaman et al., 2002). In light of projected increases in extreme weather events such as increased heavy rainfall frequency, droughts, and warming temperatures in this century due to climate change, it is vital that the effects of temperature and precipitation on vector abundances are understood (Seneviratne et al., 2012).

Statistical models are appropriate for understanding understanding explanatory power of disease drivers, but mechanistic models are more likely to be applicable across geographic ranges (Mordecai et al., 2019).

Cover: - temperature effects - lags - GAMS - non-linear relationship between climate and mosquito abundance (Roiz et al., 2014)

Gaps

- many species/locations - use of autoregressive term or not

- different temporal resolutions (so far restricted) - lag determination across species/location

Novelty of Research

Justify choice of Florida

More sustainable systems are needed for vector-borne disease surveillance (Vazquez-Prokopec et al., 2010)

Many species, many locations

assessing practical application by comparing predictive power of autoregressive vs non autoregressive model

Clear-cut Questions

1. **Which level of temporal resolution of temperature and precipitation data is best able to predict mosquito abundances?**

2. **How consistently can temporal lags and meteorological variable significance be characterised for species specific abundance models across multiple locations?**

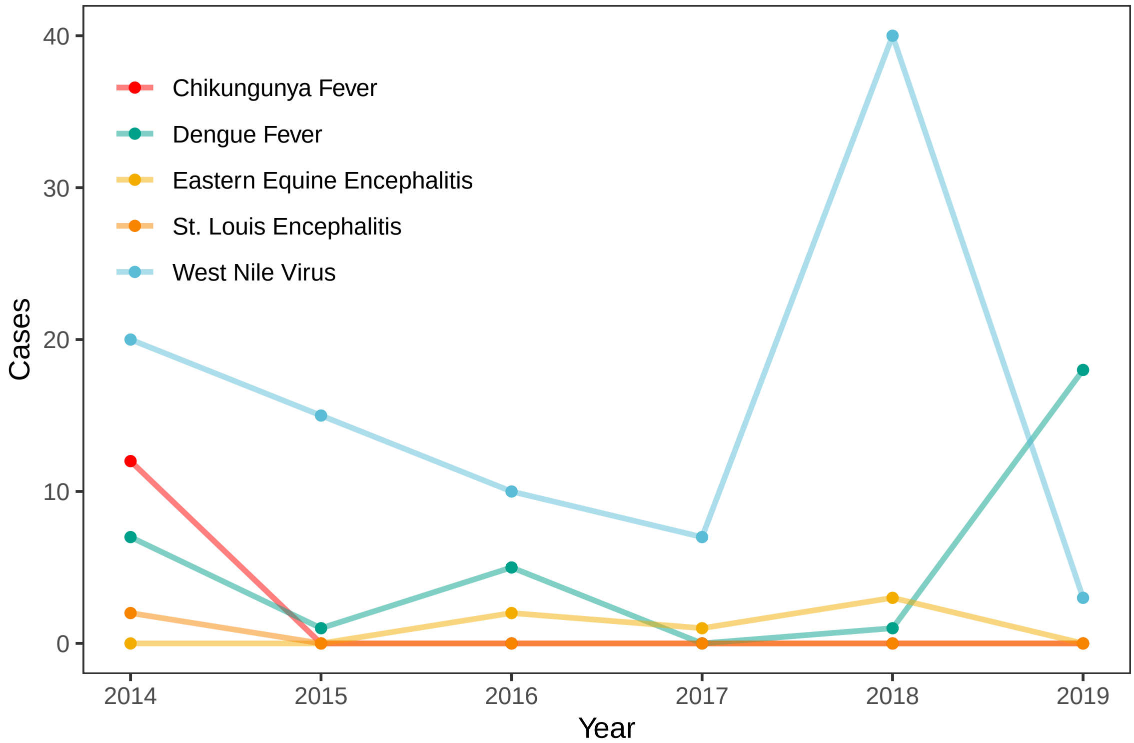


Fig. 1. Number of locally transmitted cases of mosquito-borne disease in humans in Florida between 2014 and 2019, as reported by the Florida Department of Health in 2020. Zika was also locally transmitted in 2016 (300 confirmed cases) and 2017 (2 confirmed cases) but are not shown due to the limited scale of the y axis. Imported cases, where infection was contracted elsewhere but reported in Florida, are not included as these were not contracted from Florida mosquitoes.

3. **How does incorporation of an auto-regressive term affect the predictive ability of temperature and precipitation driven models of mosquito abundance?**

Methods

Mosquito Abundance Data

Mosquito count data was obtained from the VectDyn database (**FIXME CITE**). VectDyn is a global database containing spatially and temporally explicit abundance data of mosquitoes and other arthropod vectors from published data and surveillance program records. Mosquito abundance from 205 global locations was narrowed down to seven data-rich counties in Florida, U.S.A. I decided to use data from five of these counties which had multi-year surveillance records and nearly year-round sampling from which fairly continuous time series of mosquito abundances could be formed.

Each location contained multiple trap sites. A variety of trap types were used, including BG-Sentinel traps, CDC light traps, animal-baited traps, and CDC gravid traps. Trap type can affect the efficacy of trapping for different mosquito species (Li et al., 2016). At locations with multiple traps, proportionality of each trap type was inconsistent across the time series. This would complicate conclusions of species abundances aggregated across multiple trap types. Thus, for locations with multiple types of traps, I identified the most common trap type and removed observations from all other trap types.

I averaged species-specific count data in each location and at weekly, biweekly, and monthly temporal resolutions. This transformed integer count values to averaged indicators of county-level abundance and allowed me to account for frequent variation in the number of traps deployed at each location. I aggregated species that are morphologically indistinguishable from one another according to the **ontology used by VectorBase- FIXME CITE/REFERENCE** (**list in SI**). I then removed datasets where less than 10% of the data points were greater than zero. Species counts that were identified only to the genus or family level were also removed. After this processing, I had 161 datasets of unique species and location combinations aggregated at weekly, biweekly, and monthly resolutions). Table 1 summarises the results of this processing at my five focal locations.

County	Trap Type	Years of Data	Number of Species
Lee	CDC Light Trap	2008 - 2017	18
Manatee	CDC Light Trap	2012 - 2016	48
Orange	CDC Light Trap	2012 - 2017	31
Saint Johns	baited light trap	2004 - 2015	38
Walton	New Jersey Trap	2014 - 2017	28

Table 1. Locations of abundance data and associated length of data collection, trap type, and number of species at each location. Individual species in these locations may have abundance records that are shorter than the overall years of data of each location.

Meteorological Data

Temperature and precipitation datasets were obtained from the NOAA Climate Data Online database as global NetCDF raster files at a spatial resolution of 0.50 degrees latitude and 0.50 degrees longitude. Maximum temperature in Celsius and total daily precipitation in millimetres were used based on availability of data. Rasters were rotated 180 degrees to match coordinate rotation of trap locations. Maximum daily temperature and total precipitation values were then extracted by taking the mean of the bilinear interpolation of the 4 closest raster cells to each trap location. Each raster file contained one year of meteorological data, and so this procedure was repeated for each year of the surveillance period. I then mapped the extracted daily maximum temperature and total daily precipitation to corresponding mosquito abundances by date and trap location.

Data Pre-Processing

I spatially and temporally aggregated maximum temperature data by averaging both at county-level and at weekly, biweekly, and monthly time scales. Consequently, temperature in this study refers to the average maximum daily temperature across respective temporal scales. Multiple temperature indices, including minimum, maxium, and average temperatures have all been shown to be significant drivers of mosquito abundances (Chaves et al., 2013; Roiz et al., 2014; Xu et al., 2017). Maximum temperature was chosen to capture the maximum extreme of ambient temperatures mosquitoes were exposed to as high temperatures can have detrimental effects on adult survival (Marini et al., 2016)

Number of days of rainfall has been shown to be a more effective predictor of mosquito-borne disease incidence than cumulative precipitation (Xu et al., 2017). This is likely due to the maintenance of humid conditions over time with frequent rainfall. Humidity has been independently assessed as a significant predictor of abundance dynamics (Trawinski and MacKay, 2008), and so this representation of precipitation may capture both humidity and precipitation effects. With this in mind, I aggregated precipitation to weekly, biweekly, and monthly scales by summing the number of days in each temporal period with non-zero cumulative precipitation. Consequently precipitation as I will further refer to it refers to a discrete number of days of rainfall, with a maximum of 7, 16, or 31 days dependent on temporal scale.

In order to investigate the effect of autoregressive predictors of abundance, I created an explanatory variable of abundance by shifting the continuous time series of response abundance by a lag of one time period; this was one week, bimonth, or month in each respective aggregation. Finally, I removed rows of data with missing values in any response or explanatory variables. By the end of this pre-processing, I had obtained three different temporally aggregated groups of equivalent species and location specific datasets. Each dataset contained explanatory variables of temperature, precipitation, and lagged abundance, as well as response variable abundance, at weekly, bimonthly, and monthly time scales.

Model Structure

I used univariate generalized additive models for each dataset to determine the best-fit temporal lags between abundance and temperature as well as between abundance and precipitation. Because aggregated abundance values are positive, non-integer, and non-normally distributed, I used a Gamma family distribution with a log-link function. These models had the form:

$$ln(V_t + 1) = a_0 + f_1(T_{t-l_T}) + \epsilon_t \quad (1)$$

$$ln(V_t + 1) = a_0 + f_1(P_{t-l_P}) + \epsilon_t \quad (2)$$

V_t is abundance at time t, a_0 is the intercept, T_{t-l_T} and P_{t-l_P} are the temperature and precipitation at l_T and l_P, respectively, time periods prior to time t. One is added to abundance at time t in order to prevent undefined values from the logarithm of zero abundance values. f_1 is a smooth function comprised of cubic polynomial basis functions. Each smooth function was allowed up to nine basis functions to account for flexibility in the model fit. The penalisation process in GAM fitting reduces the number of basis functions to an optimised number for each smooth function. If a smooth function had fewer than ten unique data points, such as with weekly precipitation datasets with values between zero and seven, this upper limit needed to be lowered. The maximum number of basis functions in these cases was set to one less than the number of unique values for that smooth function.

The use of cubic polynomial basis functions is an effective way to avoid the underestimation of standard error caused by concurvity with non-parametric basis functions (Dominici et al., 2002; Ramsay et al., 2003). Concurvity is the extent to which each smooth function can be approximated by other smooth function predictors and is analogous to multicollinearity in linear models. Because of the way variances are estimated in GAMs, concurvity can cause cause underestimation of p-values and thus lead to Type I errors (Ramsay et al., 2003). I used shrinkage methods to allow for full penalisation of spline complexity and REML (restricted maximum likelihood) for optimising smooth parameter estimates. Both of these smooth function estimation methodologies have been shown to maximise the predictive ability of GAMs (Marra and Wood, 2011).

Each dataset was also fit with two multivariate models, each incorporating temperature and precipitation, and one of which incorporating an autoregressive term.

$$ln(V_t + 1) = a_0 + f_1(T_{t-l_T}) + f_2(P_{t-l_P}) + \epsilon_t \tag{3}$$

$$ln(V_t + 1) = a_0 + f_1(T_{t-l_T}) + f_2(P_{t-l_P}) + f_3(V_{t-1} + 1) + \epsilon_t \tag{4}$$

In Equation 4, V_{t-1} is the abundance one time step prior to abundance at time t.

Model Selection and Evaluation

Models 1 and 2 were used to determine the best-fit lags of temperature and precipitation for each dataset at each temporal aggregation. At the weekly scale, lags of zero to twelve weeks were considered. At the bimonthly and monthly scales, lags were considered between zero, one, and two months. Because, on average, zero to two months is approximately the same number of days as zero to twelve weeks, this allowed equivalent time periods to be considered at each temporal scale. This wide range of lags were tested in order to avoid making strict assumptions about the relationship of meteorological variables with abundance and to account for wide variation in mosquito development time (Barrera and Medialdea, 1996; Beck-Johnson et al., 2013) Best-fit lags for each temporal scale were chosen by finding the minimum corrected Akaike's Information Criterion (AIC) value among models fit with each possible lag length.

$$AIC = -2ln[L(\hat{\theta}_p|y)] + 2p \tag{5}$$

AIC (Equation 5) determines the likelihood of a given model while penalizing model complexity (Johnson and Omland, 2004). Models that maximise both fit and simplicity are selected by choosing the model with the lowest AIC score. Corrected AIC fulfils the same goals as AIC, but uses likelihoods relevant to GAM models and corrects the over-selection of complex models that is typical of AIC methods for GAMs (Greven and Kneib, 2010; Wood et al., 2016). The Akaike weight of each best fit model was calculated in order to indicate the confidence in this lag as compared to other univariate lagged models. Once best-fit lags of meteorological variables for each dataset and temporal scale were determined, these lags were incorporated into multivariate models 3 and 4.

In order to compare equivalent datasets across resolutions, datasets were removed from each temporal aggregation that lacked matching datasets at other temporal resolutions due to filtering for at least 10% non-zero values. Mean Absolute Error (MAE, Equation 6) and deviance explained were used to compare the performance of lagged models at different temporal resolutions. Deviance explained is a relative measure that compares the likelihood of the proposed model to the likelihood of a saturated model with a parameter for each term and has a functionality equivalent to R^2 for non-Gaussian distribution families, such as the Gamma distribution that was used in this study (Nelder and Wedderburn, 1972). MAE (Mean Absolute Error) is an absolute measure that can be interpreted as the average error in the number of mosquitoes predicted for each dataset. Because it is non-directional, I also calculated mean bias (MB, Equation 7) to understand the overall under and overestimation of the data by the fitted model. MAE and mean bias for each dataset were found through cross validation. In equations 6 and 7, N is the number of data points and O_i and P_i are observed and predicted points, respectively.

$$MAE = \frac{1}{N}\sum |O_i - P_i| \tag{6}$$

$$MB = \frac{1}{N}\sum (O_i - P_i) \tag{7}$$

After finding the temporal resolution that best described the data, datasets at this temporal resolution was used for further analysis on comparison of variable significance and comparison of autoregressive and non-autoregressive models.

Significance of temperature and precipitation in multivariate models of mosquito abundance (Equation 3) were compared across all datasets, between genus groups, and within species. Covariates were considered significant if smooth functions had p values of less than 0.05.

Corrected AIC and deviance explained were used to compare the performance of the autoregressive and non-autoregressive models for each dataset. Relative change in deviance explained between the two multivariate models was be found through differencing. Comparison of model fit was assessed also by finding the minimum corrected AIC between the two models. Models with differences in AIC of less than two were considered to be equally best fit (Johnson and Omland, 2004).

Results

Each temporal resolution contained 161 datasets of unique species and location combinations in each temporal resolution. 50 weekly, 42 bimonthly, and 32 monthly datasets were removed that where less than 10% of the data was non-zero. Time series of temperature, precipitation, and abundance data for two sample datasets can be found in Fig 3.

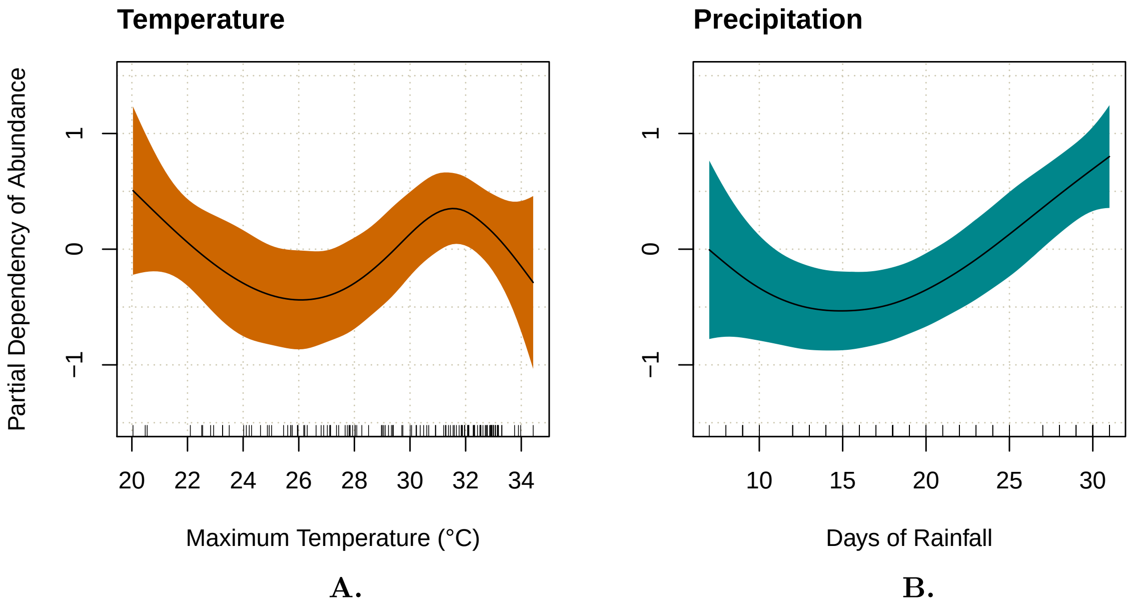


Fig. 2. Sample of multivariate model fitting showing the partial dependency of *Anopheles quadrimaculatus* abundance in Lee County, Florida on temperature and precipitation at the monthly temporal aggregation. Both temperature and precipitation are significant and best fit at a lag of two months.

Which level of temporal resolution of temperature and precipitation data is best able to predict mosquito abundances?

The fit of multivariate models (Fig 2) of temperature and abundance at best fit lags were compared across weekly, bimonthly, and monthly resolutions (Table 2) using MAE and deviance explained. 111 out of the original 161 datasets were used for this comparison. The fifty datasets that were removed had fewer than 10% non-zero data points or did not converge to a multivariate model due to a lack of sufficient data in at least one of the temporal resolutions. In all locations besides St. Johns, monthly aggregated datasets had a higher median deviance explained than the median deviance explained at weekly and bimonthly resolutions. Three out of five locations had the lowest median MAE at the monthly resolution, indicating that monthly datasets had tended to have less prediction error than other resolutions. St Johns datasets had equal median MAE at every resolution, while Lee datasets had the lowest median error at the weekly resolution. Because monthly datasets could explain the

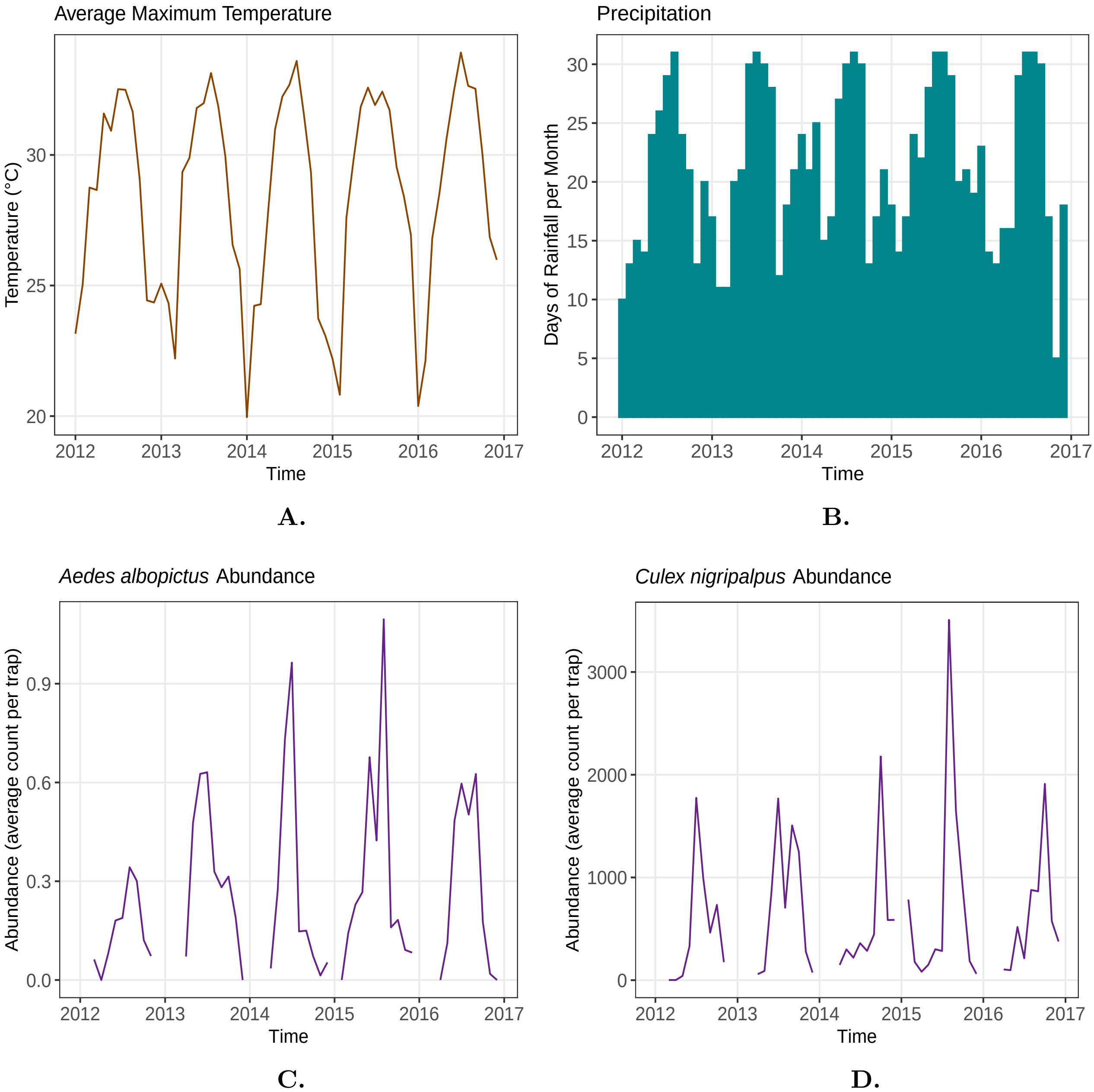


Fig. 3. Time series data of temperature, precipitation, and two vector species from Manatee County, Florida, 2012-2016. Data shown here are aggregated at the monthly scale. In **A.**, average daily maximum temperature per week shows regular seasonality with a range of about 15 °C. In **B.**, precipitation is very frequent during during a summer rainy season. This is common in all locations. **C.** and **D.** show patterns of abundance for two vector species, *Aedes albopictus* and *Culex nigripalpus*. Abundance datasets were non-continuous over winter periods where abundances are assumed to be low so traps are not employed. These missing data points were removed from datasets prior to analysis.

most deviance in four out of five locations, and had the lowest median MAE in three out of five locations, this resolution was used for the rest of the analysis.

Median Deviance Explained and Median MAE at Various Temporal Resolutions			
County	Weekly	Bimonthly	Monthly
Lee	37.7% (3.60)	45.8% (9.66)	52.2% (11.87)
Manatee	30.2% (1.04)	35.5% (0.92)	40.1% (0.90)
Orange	26.2% (0.50)	29.0% (0.44)	35.0% (0.43)
St. Johns	13.0% (0.20)	*25.9% (0.20)	12.6% (0.20)
Walton	31.7% (0.26)	40.2% (0.22)	49.9% (0.19)
All Counties	28.6% (0.76)	33.0% (0.83)	37.9% (0.78)

Table 2. Deviance explained and MAE by the best fit multivariate model of temperature and precipitation for each dataset summarised by the median value at each location and temporal resolution. MAE is in parentheses. Median was used because the distribution of deviance explained and MAE was left skewed.

How consistent are best fit temporal lags and the influence of temperature and precipitation on mosquito abundance across locations?

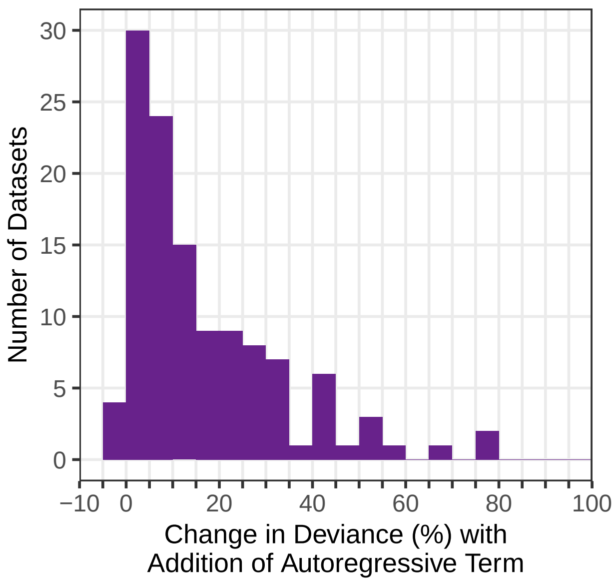
I used corrected AIC selection in univariate models of temperature and precipitation to find the best fit lags of temperature and precipitation for each dataset. Temperature in the contemporary time period ($l_T = 0$) was most frequently the best fit lag across all temporal resolutions. For precipitation models (Equation 2), lags of one and two weeks at the weekly scale were nearly equally most frequent (selected in 14 and 13 datasets, respectively). At the bimonthly scale, best fit lags of zero lags were most frequent but were favoured in only one more dataset than precipitation with one half month lag (31 and 30 datasets, respectively). At the monthly scale, the contemporary month was most frequently the best fit lag.

In order to assess the consistency of best-fit meteorological model characteristics in individual species across locations, I narrowed my monthly dataset to eleven species that occurred in all five locations and examined the frequency of best-fit lags and significant variables in each species. Nine out of eleven species (81.8%) had single temperature lag that was preferred in a majority of locations (Fig 5A.). Only one species, *Psorophora ciliata*, had consistent best fit lags for temperature across all locations. Eight of the eleven species had a precipitation lag that was best fit for a majority of the locations, but only 1 species, *Aedes infirmatus*, had a single lag length consistently chosen across all locations (Fig 5B.).

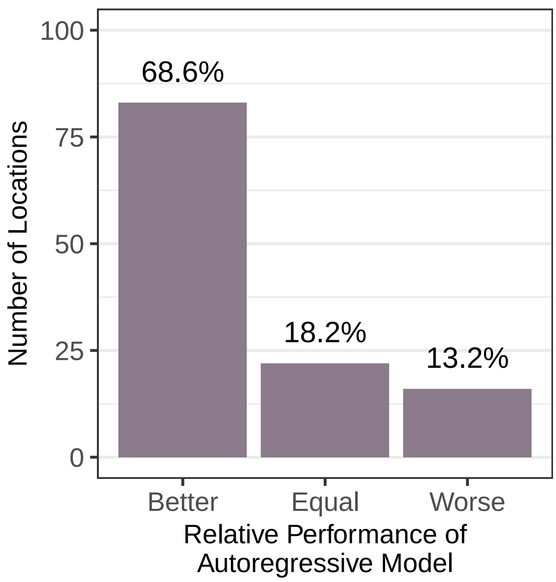
Across 128 monthly datasets with sufficient non-zero values and convergence in a multivariate model of abundance, models with significance in both temperature and precipitation was slightly more common than other combinations of variables. Variable categories were sorted by number of datasets where only temperature (25.8%), only precipitation (25.0%), both variables (30.5%), or neither variable (18.8%) was significant (Fig 5C.). In the focal eleven species, there were no species with a single category of significant predictors consistent across locations. Eight out of eleven species (72.7%), however, had a single category of significant predictors in at least a majority of locations.

How does incorporation of an autoregressive term affect the predictive ability of temperature and precipitation driven models of mosquito abundance?

I used relative comparison of deviance explained and corrected AIC to understand the change in multivariate model fits with the addition of an autoregressive term of lagged abundance (Equation 4). From 129 monthly datasets, 8 were removed that failed to converge in either multivariate model. As expected, the autoregressive term tended to improve model fit with a mean change in deviance explained of +16.7% and median change in deviance explained of +10.9% across all locations. Four datasets had a decrease in deviance explained with incorporation of an autoregressive term (range of [-2.1%, -0.2%]). Surprisingly, according to selection through corrected AIC, almost a third of datasets (31.4%) were equally or better fit by the model with only meteorological predictors.



A.



B.

Fig. 4. Comparison of the fit of autoregressive versus non-autoregressive meteorological models of mosquito abundance for each monthly dataset. After removing 8 datasets where the autoregressive model did not converge, 121 datasets were compared. **A.** is a histogram of the relative change in deviance explained with the incorporation of an autoregressive term to the multivariate model for each dataset. Four datasets had a decrease (range of [-2.1%, -0.2%]) in deviance explained with incorporation of an autoregressive term. In **B.**, the number of datasets where the autoregressive model was a better, equal, and worse fit than the non-autoregressive model is shown.

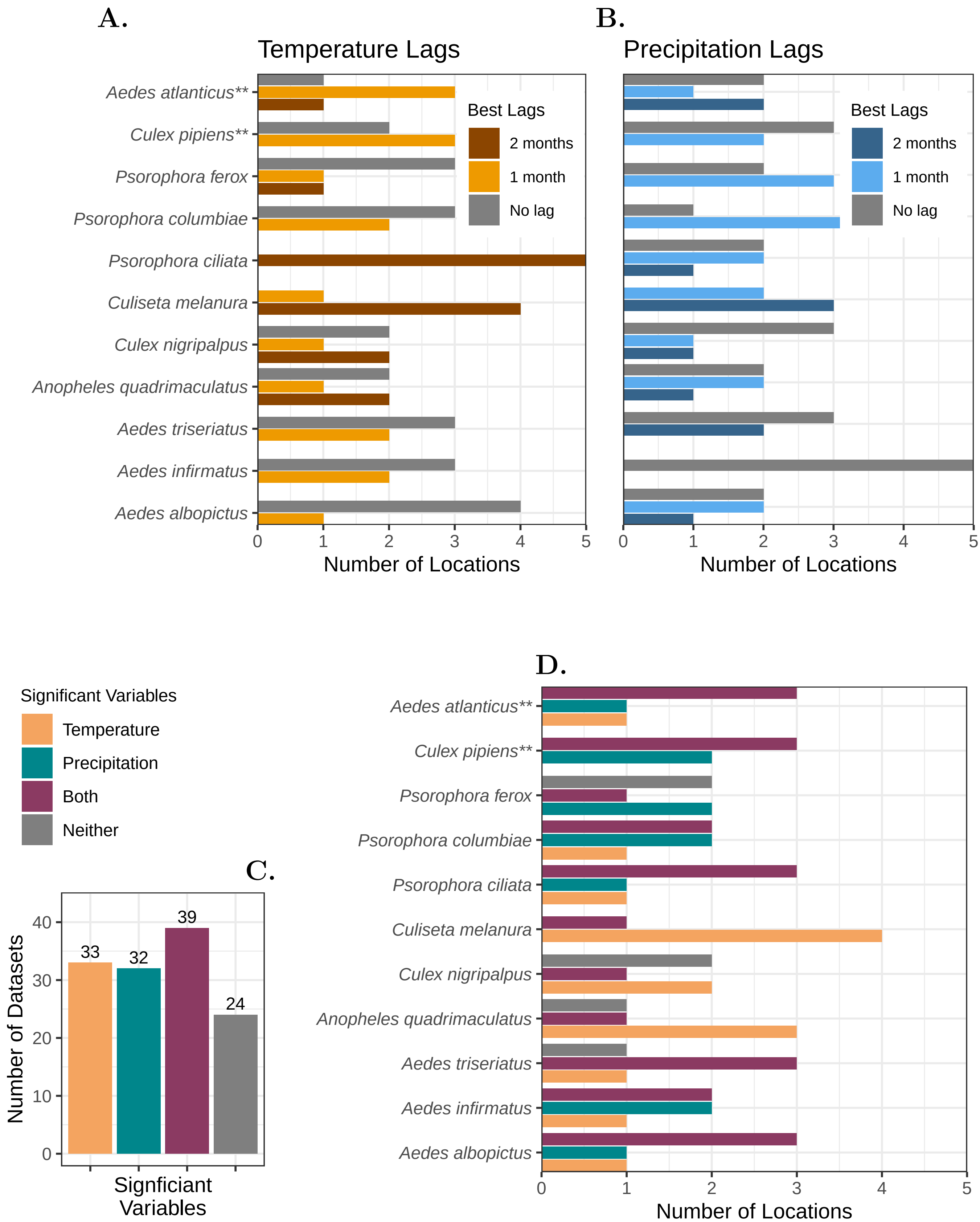


Fig. 5. Frequency of best fit lags and significant variables across locations for monthly datasets. Morphological groups are marked with asterisks. In **A.** and **B.**, frequency of best fit temperature and precipitation lags chosen by corrected AIC from univariate models of abundance are shown for eleven species that were present in all locations. **C.** shows the frequency of significant predictors for 128 abundance datasets fit with multivariate models of temperature and precipitation. In **D.**, the eleven species are shown with the frequency of significant variables in datasets across all locations.

Discussion

Mosquito-borne disease transmission is dependent on both natural and human environmental factors, such as climate, land use, health infrastructure, host demographics- using only climate will overestimate impact of climate change on populations (Parham et al., 2015)

Aggregation of precipitation will lose the importance of quick torrential rains likely to flush immature mosquitoes (Koenraadt and Harrington, 2008)

Even if monthly aggregated data captures more of the variation, we're talking aggregation not surveillance protocol. Studies have shown lower sampling error when doing more frequent (1/week) sampling versus less frequent but more sampling (2/fortnight) (Magbity and Lines, 2002)

Bibliography

(2017). Global vector control response 2017–2030. Technical report, World Health Organization, Geneva.

(2020). Mosquito-Borne Disease Surveillance.

Amarasekare, P. and Savage, V. (2012). A framework for elucidating the temperature dependence of fitness. *American Naturalist*, 179(2):178–191.

Barrera, R. and Medialdea, V. (1996). Development time and resistance to starvation of mosquito larvae. *Journal of Natural History*, 30(3):447–458.

Beck-Johnson, L. M., Nelson, W. A., Paaijmans, K. P., Read, A. F., Thomas, M. B., and Bjørnstad, O. N. (2013). The effect of temperature on Anopheles mosquito population dynamics and the potential for malaria transmission. *PLoS ONE*, 8(11).

Cator, L. J., Johnson, L. R., Mordecai, E. A., El Moustaid, F., Smallwood, T. R., LaDeau, S. L., Johansson, M. A., Hudson, P. J., Boots, M., Thomas, M. B., Power, A. G., and Pawar, S. (2020). The Role of Vector Trait Variation in Vector-Borne Disease Dynamics. *Frontiers in Ecology and Evolution*, 8(July).

Chase, J. M. and Knight, T. M. (2003). Drought-induced mosquito outbreaks in wetlands. *Ecology Letters*, 6(11):1017–1024.

Chaves, L. F., Higa, Y., Lee, S. H., Jeong, J. Y., Heo, S. T., Kim, M., Minakawa, N., and Lee, K. H. (2013). Environmental Forcing Shapes Regional House Mosquito Synchrony in a Warming Temperate Island. *Environmental Entomology*, 42(4):605–613.

Dominici, F., McDermott, A., Zeger, S. L., and Samet, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology*, 156(3):193–203.

Greven, S. and Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, 97(4):773–789.

Gubler, D. J. (1998). Resurgent vector-borne diseases as a global health problem. *Emerging Infectious Diseases*, 4(3):442–450.

Gubler, D. J. (2001). Prevention and control of tropical diseases in the 21st century: Back to the field. *American Journal of Tropical Medicine and Hygiene*, 65(1).

Johnson, J. B. and Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19(2):101–108.

Keith, J. (2005). The importance of agricultural tire habitats for mosquitoes of public health importance in New York State. *Source: Journal of the American Mosquito Control Association*, 21(2):171–176.

Koenraadt, C. and Harrington, L. C. (2008). Flushing Effect of Rain on Container-Inhabiting Mosquitoes Aedes aegypti and Culex pipiens (Diptera: Culicidae) . *Journal of Medical Entomology*, 45(1):28–35.

Li, R., Xu, L., Bjørnstad, O. N., Liu, K., Song, T., Chen, A., Xu, B., Liu, Q., and Stenseth, N. C. (2019). Climate-driven variation in mosquito density predicts the spatiotemporal dynamics of dengue. *Proceedings of the National Academy of Sciences of the United States of America*, 116(9):3624–3629.

Li, Y., Su, X., Zhou, G., Zhang, H., Puthiyakunnon, S., Shuai, S., Cai, S., Gu, J., Zhou, X., Yan, G., and Chen, X.-G. (2016). Comparative evaluation of the efficiency of the BG-Sentinel trap, CDC light trap and Mosquito-oviposition trap for the surveillance of vector mosquitoes. *Parasites & Vectors*, 9(1):446.

Lowe, R., Bailey, T. C., Stephenson, D. B., Jupp, T. E., Graham, R. J., Barcellos, C., and Carvalho, M. S. (2013). The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statistics in Medicine*, 32(5):864–883.

Magbity, E. and Lines, J. (2002). Spatial and temporal distribution of Anopheles gambiae s.l. (Diptera: Culicidae) in two Tanzanian villages: implication for designing mosquito sampling routines. *Bulletin of Entomological Research*, 92(6):483–488.

Marini, G., Poletti, P., Giacobini, M., Pugliese, A., Merler, S., and Rosà, R. (2016). The role of climatic and density dependent factors in shaping mosquito population dynamics:The case of culex pipiens in northwestern Italy. *PLoS ONE*, 11(4):1–15.

Marra, G. and Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics and Data Analysis*, 55(7):2372–2387.

Mordecai, E. A., Caldwell, J. M., Grossman, M. K., Lippi, C. A., Johnson, L. R., Neira, M., Rohr, J. R., Ryan, S. J., Savage, V., Shocket, M. S., Sippy, R., Stewart Ibarra, A. M., Thomas, M. B., and Villena, O. (2019). Thermal biology of mosquito-borne disease. *Ecology Letters*, 22(10):1690–1708.

Nelder, J. and Wedderburn, R. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, 135(3):370–384.

Paaijmans, K. P., Wandago, M. O., Githeko, A. K., and Takken, W. (2007). Unexpected High Losses of Anopheles gambiae Larvae Due to Rainfall. *PLoS ONE*, 2(11):e1146.

Parham, P. E., Waldock, J., Christophides, G. K., Hemming, D., Agusto, F., Evans, K. J., Fefferman, N., Gaff, H., Gumel, A., Ladeau, S., Lenhart, S., Mickens, R. E., Naumova, E. N., Ostfeld, R. S., Ready, P. D., Thomas, M. B., Velasco-Hernandez, J., and Michael, E. (2015). Climate, environmental and socio-economic change: Weighing up the balance in vector-borne disease transmission. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1665):1–17.

Poh, K. C., Chaves, L. F., Reyna-Nava, M., Roberts, C. M., Fredregill, C., Bueno, R., Debboun, M., and Hamer, G. L. (2019). The influence of weather and weather variability on mosquito abundance and infection with West Nile virus in Harris County, Texas, USA. *Science of the Total Environment*, 675:260–272.

Ramsay, T. O., Burnett, R. T., and Krewski, D. (2003). The Effect of Concurvity in Generalized Additive Models Linking Mortality to Ambient Particulate Matter. *Epidemiology*, 14(1):18–23.

Roiz, D., Ruiz, S., Soriguer, R., and Figuerola, J. (2014). Climatic effects on mosquito abundance in Mediterranean wetlands. *Parasites and Vectors*, 7(1):1–13.

Sedda, L., Lucas, E. R., Djogbénou, L. S., Edi, A. V. C., Egyir-Yawson, A., Kabula, B. I., Midega, J., Ochomo, E., Weetman, D., and Donnelly, M. J. (2019). Improved spatial ecological sampling using open data and standardization: an example from malaria mosquito surveillance. *Journal of The Royal Society Interface*, 16(153):20180941.

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., Mc Innes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., Zhang, X., Rusticucci, M., Semenov, V., Alexander, L. V., Allen, S., Benito, G., Cavazos, T., Clague, J., Conway, D., Della-Marta, P. M., Gerber, M., Gong, S., Goswami, B. N., Hemer, M., Huggel, C., Van den Hurk, B., Kharin, V. V., Kitoh, A., Klein Tank, A. M., Li, G., Mason, S., Mc Guire, W., Van Oldenborgh, G. J., Orlowsky, B., Smith, S., Thiaw, W., Velegrakis, A., Yiou, P., Zhang, T., Zhou, T., and Zwiers, F. W. (2012). Changes in climate extremes and their impacts on the natural physical environment. In Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen,M. Tignor and Midgley, P., editors, *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*, chapter 3, pages 109–230. Cambridge University Press, Cambridge, UK, and New York, NY, USA.

Shaman, J., Day, J. F., and Stieglitz, M. (2002). Drought-induced amplification of Saint Louis encephalitis virus, Florida. *Emerging Infectious Diseases*, 8(6):575–580.

Tomerini, D. M., Dale, P. E., and Sipe, N. (2011). Does mosquito control have an effect on mosquito-borne disease? the case of ross river virus disease and mosquito management in Queensland, Australia. *Journal of the American Mosquito Control Association*, 27(1):39–44.

Trawinski, P. R. and MacKay, D. S. (2008). Meteorologically conditioned time-series predictions of West Nile virus vector mosquitoes. *Vector-Borne and Zoonotic Diseases*, 8(4):505–521.

Vazquez-Prokopec, G. M., Chaves, L. F., Ritchie, S. A., Davis, J., and Kitron, U. (2010). Unforeseen Costs of Cutting Mosquito Surveillance Budgets. *PLoS Neglected Tropical Diseases*, 4(10):e858.

Wang, J., Ogden, N. H., and Zhu, H. (2011). The Impact of Weather Conditions on Culex pipiens and Culex restuans (Diptera: Culicidae) Abundance: A Case Study in Peel Region. *Journal of Medical Entomology*, 48(2):468–475.

Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, 111(516):1548–1563.

Xu, L., Stige, L. C., Chan, K. S., Zhou, J., Yang, J., Sang, S., Wang, M., Yang, Z., Yan, Z., Jiang, T., Lu, L., Yue, Y., Liu, X., Lin, H., Xu, J., Liu, Q., and Stenseth, N. C. (2017). Climate variation drives dengue dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 114(1):113–118.

Yang, G. J., Brook, B. W., and Bradshaw, C. J. (2009). Predicting the timing and magnitude of tropical mosquito population peaks for maximizing control efficiency. *PLoS Neglected Tropical Diseases*, 3(2):1–9.

Yoo, E. H., Chen, D., and Diao, C. (2016). The effects of weather and environmental factors on west nile virus mosquito abundance in greater toronto area. *Earth Interactions*, 20(3).