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 and cost-effective method to control the public health impacts of vector-borne disease (Vazquez-Prokopec et al., 2010). 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). Intervention in the growth of mosquito popualtions through chemical control measures is an effective public health strategy for reducing incidence of mosquito borne disease (Tomerini et al., 2011). Timing these intervention measures to interrupt peak mosquito abundances, whether predicted or surveilled, could inform cost-effective disease management solutions 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 (Yoo et al., 2016)

Precipitation could have mixed effects on mosquito abundance dynamics. Precipitation raises near-surface humidity, which increases adult mosquito activity and host-seeking behaviour (Shaman and Day, 2007). This increased reproductive activity would lead to lagged effects on mosquito abundance, but would also cause immediate increases in mosquito trap counts. Many trap types are designed to attract host-seeking or gravid females and are consequently also attractive to mate-seeking male mosquitoes (Li et al., 2016). Rainfall also affects many of the aquatic habitats important for early life stages (Shaman and Day, 2007). Container-breeding species that use man-made containers for oviposition may experience increased breeding sites when precipitation creates habitats in otherwise dry containers (Keith, 2005). Precipitation can also expand suitable habitats for mosquitoes breeding in natural water bodies (Koenraadt and Harrington, 2008). While some rainfall seems likely to have positive effects on mosquito abundance, the effect of heavy rainfall is less clear. Heavy rainfall can flush immature mosquitoes from aquatic habitats, but the extent of this effect varies among species (Koenraadt and Harrington, 2008; Paaijmans et al., 2007). For example, specialist container-breeding

Aedes aegypti has been found to have a stronger protective diving response to rainfall compared to generalist habitat breeding *Culex pipiens* (Koenraadt and Harrington, 2008).

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

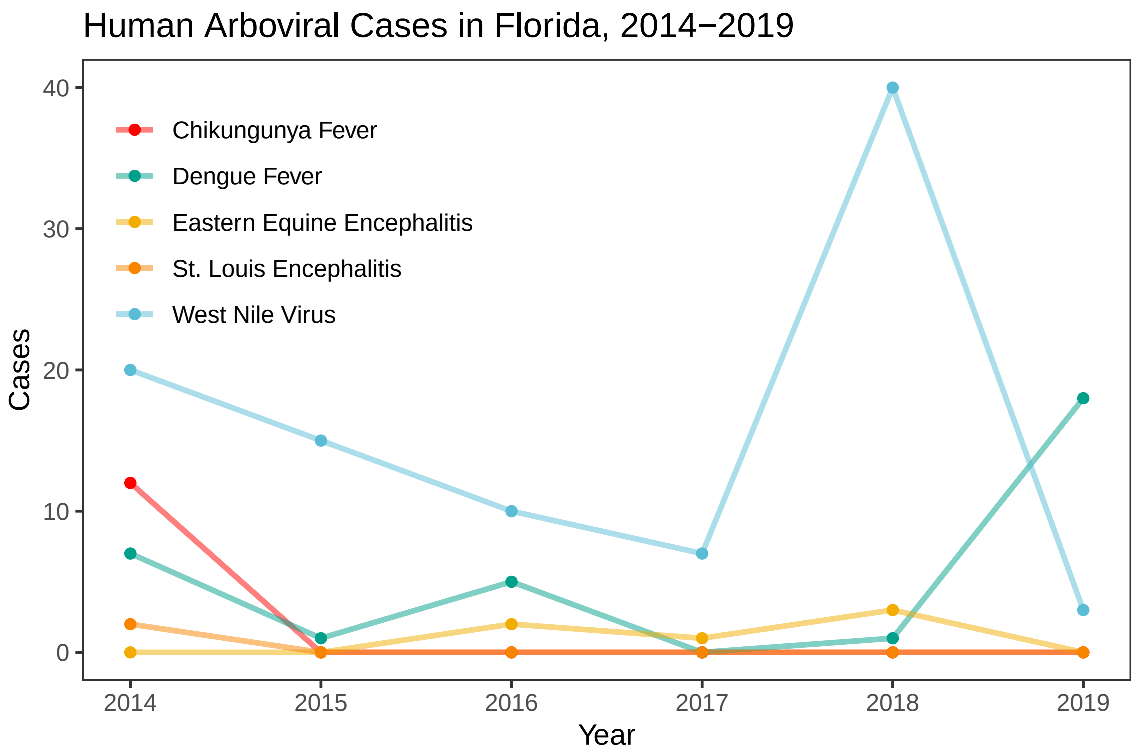


Fig. 1. Number of locally transmitted cases of mosquito-borne disease in humans in Florida between 2014 and 2019, as reported by the Florida Department of Health in 2020. Zika was also locally transmitted in 2016 (300 confirmed cases) and 2017 (2 confirmed cases) but are not shown due to the limited 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.

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. **To what extent can species-specific abundance models be predictive at multiple locations?**

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 for 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 species that had only zero or NA abundance values. Species counts that were identified only to the genus or family level were also removed. After this processing, I had 163 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	Observations	Number of Species
Lee	CDC Light Trap	11		18
Manatee	CDC Light Trap	5		48
Orange	CDC Light Trap	6		31
Saint Johns	baited light trap	13.5		38
Walton	New Jersey Trap	3		28

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

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. **FIXME explain why choice of max temp vs average, min, etc**

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, 14, or 31 dependent on temporal scale. Finally, I removed rows of data with missing values in any response or explanatory variables.

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:

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

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

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. The use of cubic polynomial basis functions has been shown as an effective way to avoid the bias introduced by concurvity with non-parametric basis functions(Dominici et al., 2002; Ramsay et al., 2003). Concurvity is the extent to which each predictive smooth function can be approximated by other smooth function predictors and is analogous to multicollinearity in linear models. This can cause unstable estimates of the response variable and cause underestimation of standard error of explanatory variables (Ramsay et al., 2003).

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 and biweekly scale, lags between zero and six weeks were considered. At the monthly scale, lags between zero and two months were considered. Best-fit lags for each temporal scale were chosen by finding the minimum 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 is a method for determining the likelihood of a given model while penalizing model complexity (Johnson and Omland, 2004).

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.

FIXME- I have just realized that this is an inappropriate use of AIC. These univariate models are not nested and are based on different datasets- different lagged meteorological data. I should actually use GCV- generalized cross validation.

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. The generalized cross-validation (GCV) score and deviance explained of these multivariate models were compared.

Results

Figures/Charts:

1. time series of abundance, temp, and precipitation
2. partial dependency plot of multivariate GAM of a sample species
3. Table comparing deviance explained of different temporal resolutions

Median Deviance Explained at Various Temporal Resolutions			
County	Weekly	Biweekly	Monthly
Lee	31.9%	42.2%	*52.2%
Manatee	24.8%	21.6%	*34.1%
Orange	17.0%	15.8%	*21.3%
Saint Johns	*8.5%	7.4%	8.4%
Walton	23.3%	40.0%	*49.9%
All Counties	20.0%	22.2%	*31.9%

Table 2. This table makes me hate Saint Johns

5. Plot showing effect of autoregressive term. Also include in words what the AIC difference was between AR and non-AR model

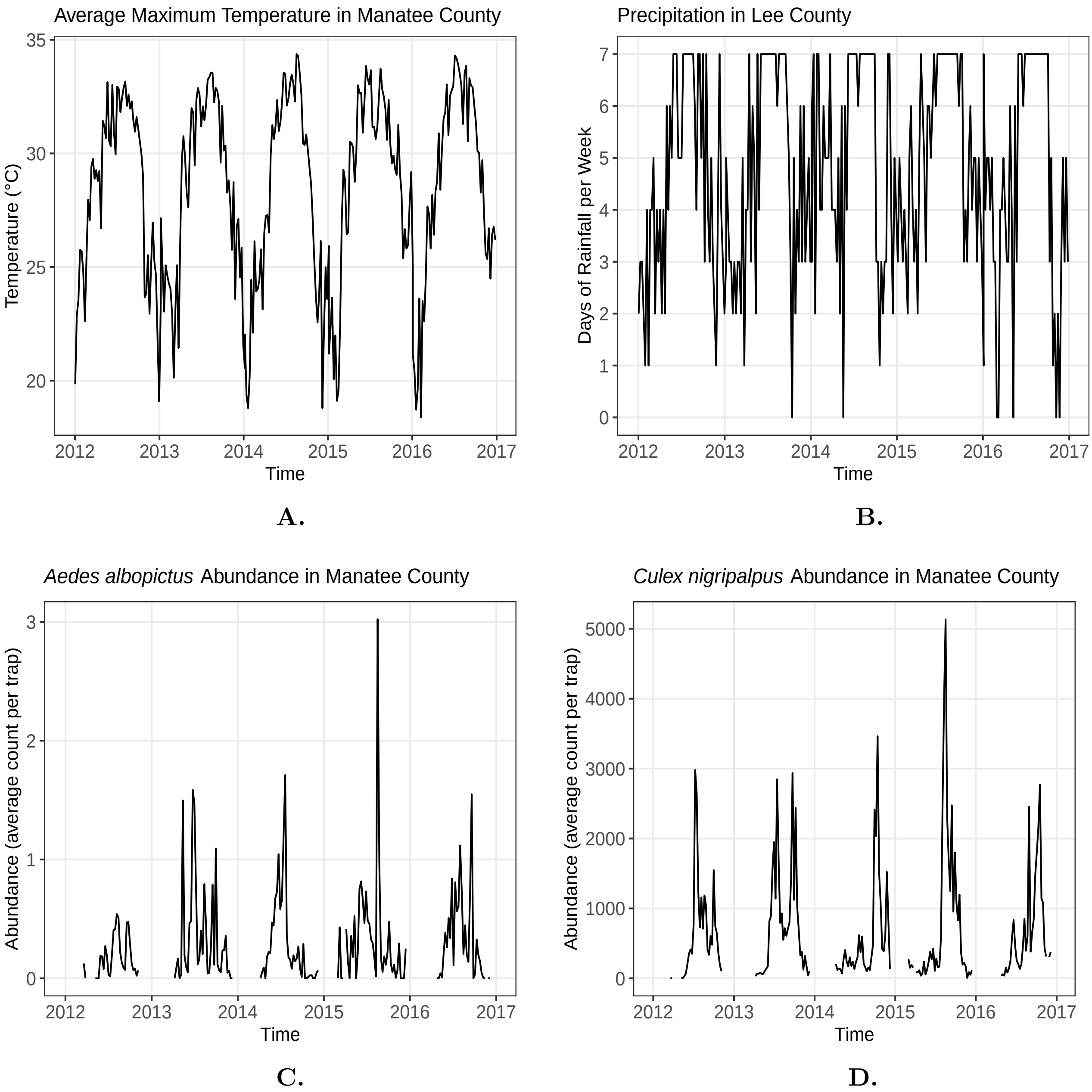
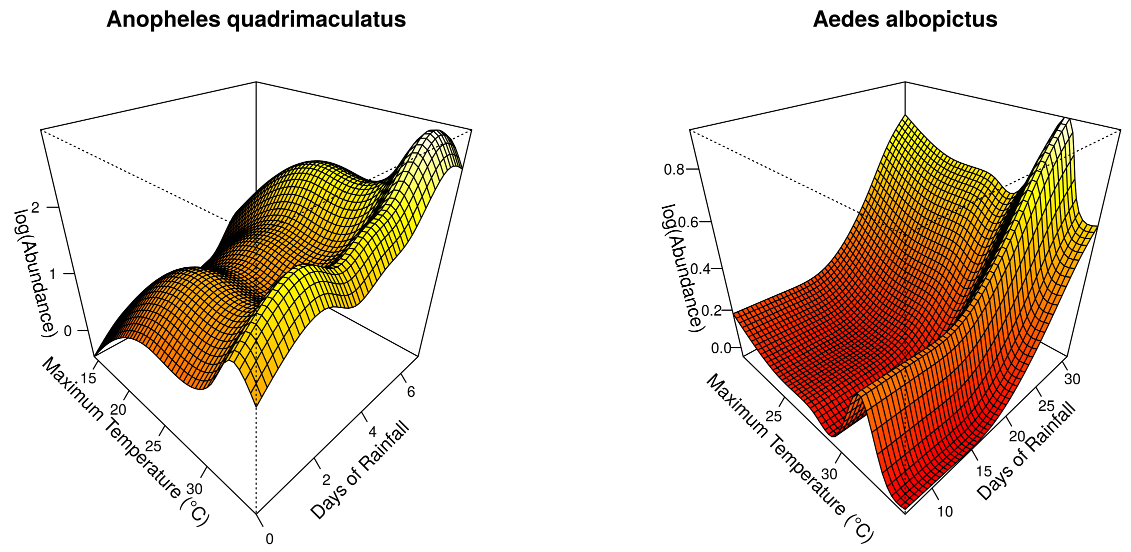


Fig. 2. Sample time series data of temperature, precipitation, and two vector species from Manatee County, Florida, 2012-2012. Data shown here are aggregated at the weekly scale. In **A.**, average daily maximum temperature per week shows regular seasonality with a range of about 15 °C. In **B.**, precipitation 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.



A. **B.**

Fig. 3. This a sample of datasets fit with multivariate models of temperature and precipitation at the best fit lags (Equation 3). In **A.**, average *Anopheles quadrimaculatus* count per trap at a weekly resolution was best fit according to univariate model comparison (Equations 1 and 2) with temperature at a lag of six weeks and precipitation with a lag of four weeks. Both temperature and precipitation were significant. In **B.**, a Aedes albopictus abundance at a monthly resolution was best fit with temperature at a lag of one month and precipitation in the contemporary month. In this dataset, only precipitation was significant.

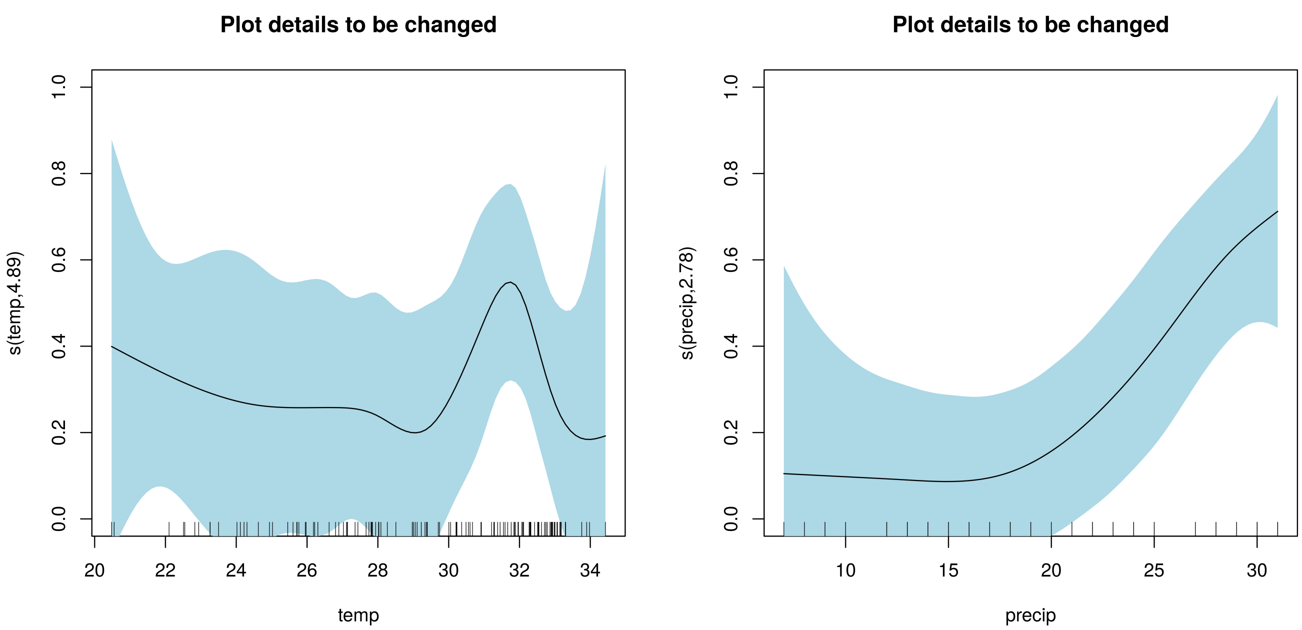
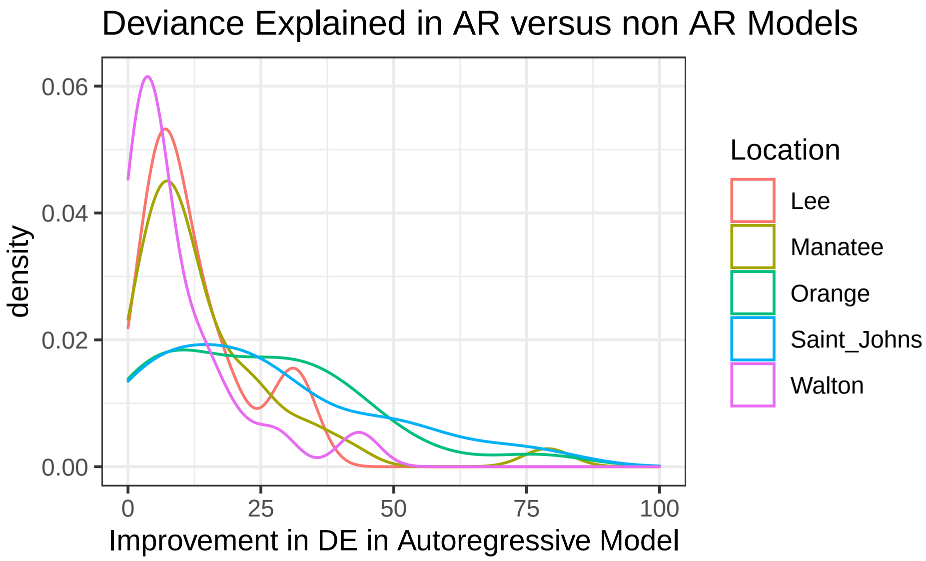


Fig. 4. This is my second option for demonstrating multivariate model fits (versus Fig. **??**)



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)

Bibliography

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

(2020). Mosquito-Borne Disease Surveillance.

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).

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.

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.

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.

Shaman, J. and Day, J. F. (2007). Reproductive Phase Locking of Mosquito Populations in Response to Rainfall Frequency. *PLoS ONE*, 2(3):e331.

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.

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.

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).