PubH 7440 Project Proposal

Jake Wittman

2020-04-13

1. A high-level statement of the problem you intend to address

The emerald ash borer (*Agrilus planipennis*) is a beetle native to eastern Asia that attacks trees ash trees (genus *Fraxinus*). This beetle was introduced to North America sometime in the 1990s near Detroit, MI and has since spread to 34 differet states in the US and five Canadian provices. *Fraxinus* species native to North America do not have a coevolved history with this invasive pest, and as a result the beetle is devestating ash trees across the continent, with nearly 100% mortality of ash in invaded areas (Herms and McCullough, 2014). The cost to communities to manage the emerald ash borer (EAB) is enormous; the estimated 10 year cost to communities in 25 states is approximately $10.7 billion USD (Kovacs et al., 2010). Slowing the spread of invasive species that cannot be eradicated has been shown to help reduce and spread out costs associated with managing the pests, easing the financial burden on affected communities (Fahrner et al., 2017; McCullough and Mercader, 2012). The goal of this study is elucidate the factors that have drive the spread of EAB, such as environmental (e.g. temperature, precipitation, amount of available host tree), geographic (e.g. density of infestation in neighboring areas), and social (e.g. population density in a given county). Of particular interest is how the spread rate of EAB may vary with the direction of spread. There is some evidence that the pest may spread more slowly to the north and to the south of its currently invaded range. This study will focus on quantifying the spread rate of the beetle, which factors appear to be driving it, and if the beetle is spreading more slowly in some directions that others.

2. The data source(s) you intended to use

The data for this project come from county level EAB detection data provided by USDA-APHIS in the US and the Canadian Forest Service in Canada. County level environmental data will come from the PRISM Climate Group at Oregon State University. Geographic and social data will come from the 2010 US Census and similar Canadian sources.

3. How you plan to obtain that data

The county level detection data has been provided by collaborators. Other data is freely available online. Spread distances and directions will be calculated using R.

4. The goals of your analysis

Quantify the spread rate of EAB, estimated in kilometers moved per year in two different ways. Identify which factors appear to be driving the spread of EAB. Determine if spread rate depends on the direction of spread, and if so, what factors drive this pattern. Compare how estimates change depending on which calculation of spread rate is used.

5. A description of the Bayesian analysis tools you plan to use

I plan to build a hierarchical regression model using Stan.

$Y_{it} \sim N(X\beta + \alpha_i + \gamma_t, \sigma_m^2 + \sigma_d^2)$

$\alpha_i \sim N(\eta_s, \tau^2)$

$\eta_s \sim N(0, \phi^2)$

$\gamma_t \sim N(0, \epsilon^2)$

Where Y_{it} is the spread rate calculated as the distance between a detection in county i in year t and the closest infested county $j \neq i$ in year $t-1$ or the linear distance to the point of introduction in Detroit, MI. $X\beta$ is the design matrix containing county level covariates, α_i are random effects for county which are nested within random effects for state η_s. Lastly, there is a random effect for year, γ_t to allow for yearly variation in spread rate detection. The error component, σ^2 will be divided into two components: the residual error σ^2_m and the error due to imperfect detection data σ^2_d (Yarkoni, 2019). Emerald ash borer spends the juvenile portion of its life cycle under the bark of trees and the adult portion in the canopy of trees. This beetle is also in the family Buprestidae, which to our knowledge, only use very short range signaling pheromones to attract mates, making it very difficult to detect emerald ash borer when populations are small. The cryptic life cycle of the insect coupled with the difficulties of luring it into traps mean detection data are imperfect and may not reflect when the insect actually arrived in a given county. The value σ^2_d is thus unknown and not identifiable in this dataset. This will be accounted for in the model by using various strong priors for $\pi(\sigma^2_d)$ as a test of various assumptions of the strength of this error term.

Spread rate will be calculated in two ways. First, by identifying counties that were invaded in year t ($county_{i,t}$), locating the nearest county that was invaded in year $t-1$ ($county_{j \neq i,t-1}$), and calculating the distance between the newly invaded county's centroid and its closest previously invaded neighbor. The second method will calculate rate of spread per year as the distance between new detections and the point of introduction in Detroit, MI divided by the years since introduction. These calculations will provide multiple estimates of spread rate $Y_{i,t}$, which will allow comparisons between the two. The direction of this spread will also be recorded in radians. Spread from one county to another positioned directly north of it would be recorded as 0 radians, while spread to a county directly south would be recorded as π radians.

County level covariates included in fixed effects portion of the model will be the direction of spread, climatic variables in county i that are known to affect species distributions (average winter temperature, average minimum winter temperature, average yearly precipitation, average summer temperature, average high summer temperature), the amount of host material (i.e. basal area of *Fraxinus* trees) available in county i, the number of counties in year $t-1$ that are infested and adjacent to a county i newly infested in year t, and the population density for county i. To appropriately model direction of spread, this variable will be fit as $\beta_1 * cos(direction_i) + \beta_2 * sin(direction_i)$, which should allow flexibility in the relationship between direction of spread and distance. The model will be fit with and without the direction covariate, as well as state-level random effects, and WAIC and LOO to determine how much the addition of direction improves the fit of the model.

6. The products you plan to build, ideally including visualizations, and a report of outcomes.

Posterior density plots with shaded 95% credible intervals will be included for all fixed effects coefficients. A graph of spread rate against direction with fitted values from the model will be used to visualize model predictions and uncertainty. A table showing mean and median point estimates of coefficients along with standard error of the coefficients will be presented for each different strong prior assigned to detection error σ^2_d.

References

Fahrner, S.J., Abrahamson, M., Venette, R.C., Aukema, B.H., 2017. Strategic removal of host trees in isolated, satellite infestations of emerald ash borer can reduce population growth. Urban Forestry & Urban Greening 24, 184–194. doi:10.1016/J.UFUG.2017.03.017

Herms, D.A., McCullough, D.G., 2014. Emerald Ash Borer Invasion of North America: History, Biology, Ecology, Impacts, and Management. Annual Review of Entomology 59, 13–30. doi:10.1146/annurev-ento-011613-162051

Kovacs, K.F., Haight, R.G., McCullough, D.G., Mercader, R.J., Siegert, N.W., Liebhold, A.M., 2010. Cost of potential emerald ash borer damage in U.S. communities, 2009–2019. Ecological Economics 69, 569–578. doi:10.1016/J.ECOLECON.2009.09.004

McCullough, D.G., Mercader, R.J., 2012. Evaluation of potential strategies to SLow Ash Mortality (SLAM) caused by emerald ash borer (Agrilus planipennis): SLAM in an urban forest. International Journal of Pest

Management 58, 9–23. doi:10.1080/09670874.2011.637138

Yarkoni, T., 2019. The generalizability crisis. PsyArXiv 1–26. doi:10.31234/osf.io/jqw35