Spread Paper Revisions

Statistical Analysis

JMG

Statistical analysis to determine relevance of factors

For a statistical test of our hypothesis, we have use R version 4.3.1 and package `mgcv` version 1.9 to fit a generalized additive model (R Core Team 2023; S. N. Wood 2003; Simon N. Wood 2017). Specifically, we model the log odds of county infestation in 2021 with tensor product smoothing for longitude and latitude to control for spatial autocorrelation and parametric terms for presence/absence of two-digit interstate highway, number of garden centers, 2019 county population, and an interaction term for presence/absence of two-digit interstate highway and number of garden centers. To deal with issues of convergence and variables of different scale, we log transformed the population and normalized the number of garden centers. We used the packages `DHARMa` version 0.4.6 and `gratia` gratia version 0.8.1.34 for diagnostics to assess model assumptions (Hartig 2022 ; Simpson 2023).

Figure 1 shows the estimates on the odds-ratio scale for the parametric terms in our generalized additive logistic regression model while Table 1 displays the values for the same estimated coefficients on the log-odds scale. The model results indicate that the data provides evidence for the increase in likelihood of infestation in a county in 2021 with the presence of interstate highways and larger populations. Additionally, our statistical analysis also suggests that the data provides evidence for the increase in likelihood of infestation in a county in 2021 with an increase in the number of garden centers only when there is also the presence of a two-digit interstate highway for that county.

Table 1: Summary table for estimates for linear terms in generalized additive model. Results are on the log-odds scale.

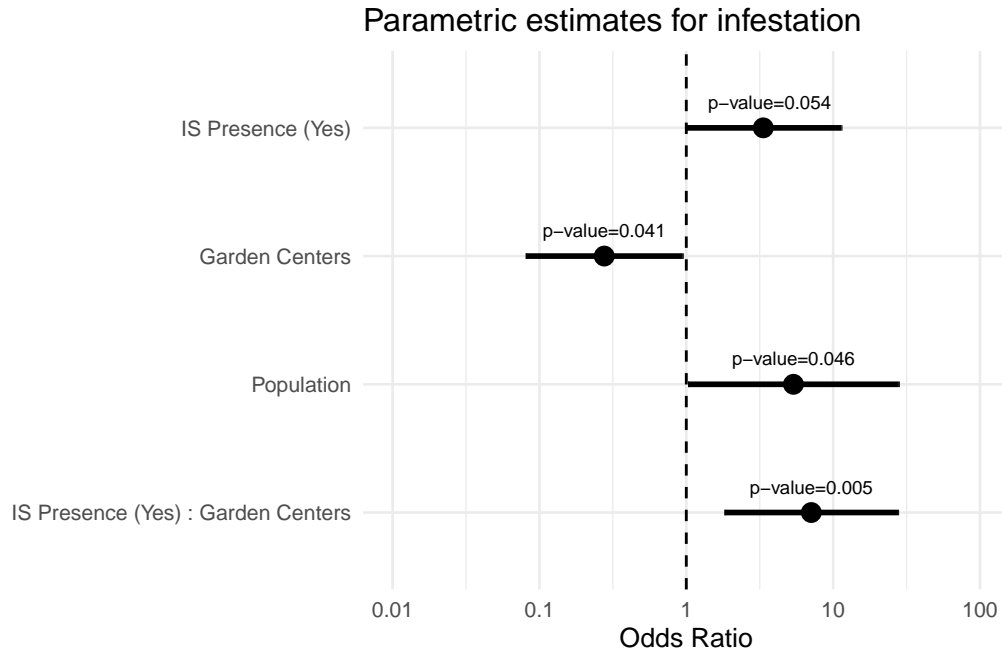| Coefficient	Estimate	SE	p-value
Intercept	-8.794628	4.2781470	0.0398105
IS Presence - Yes	1.207274	0.6253652	0.0535437
Garden Centers	-1.286439	0.6280036	0.0405151
Population (log)	1.679994	0.8431980	0.0463264
IS Presence - Yes : Garden Centers	1.959781	0.6944467	0.0047714

Figure 1: Estimated coefficients for the linear predictors in the generalized additive model. Results are on the scale of odds-ratio. The likelihood of infestation for a county in 2021 is predicted to increase with the presence of interstate highways and larger populations. Interestingly, the likelihood of infestation for a county in 2021 is predicted to increase with an increase in the number of garden centers only when there is also the presence of a two-digit interstate highway for that county.

Supplementary analysis

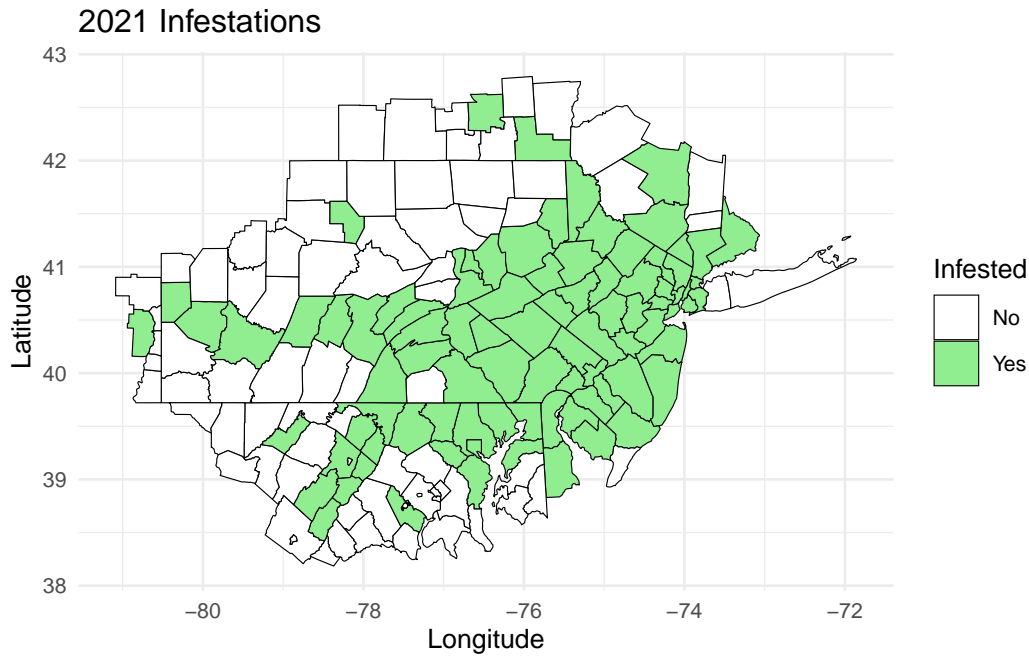Figure 2 displays the 2021 infestation status for the 166 Mid-Atlantic region counties in 2021:

Figure 2: Counties in the defined Mid-Atlantic regions that were designated as infested by the spotted laternfly in 2021.

Table 2 displays the number of counties per year in the defined Mid-Atlantic region of 166 that were listed as infected.

Figure 3 shows the 166 counties in the defined Mid-Atlantic region together with their 2021 infestation status and the number of two-digit interstate highways that transect the county.

Figure 4 shows the 166 counties in the defined Mid-Atlantic region together with their 2021 infestation status and the number of garden centers in the county.

Figure 5 shows the 166 counties in the defined Mid-Atlantic region together with their 2021 infestation status and the county population on the log scale as estimated in 2019.

Figure 6 shows the 166 counties in the defined Mid-Atlantic region together with the presence/absence of two-digit interstate highways and the number of garden center per county.

Infestation Trend Over Time

(Simpson 2023)

Table 2: The number of counties per year in the defined Mid-Atlantic region of 166 that were listed as infected.

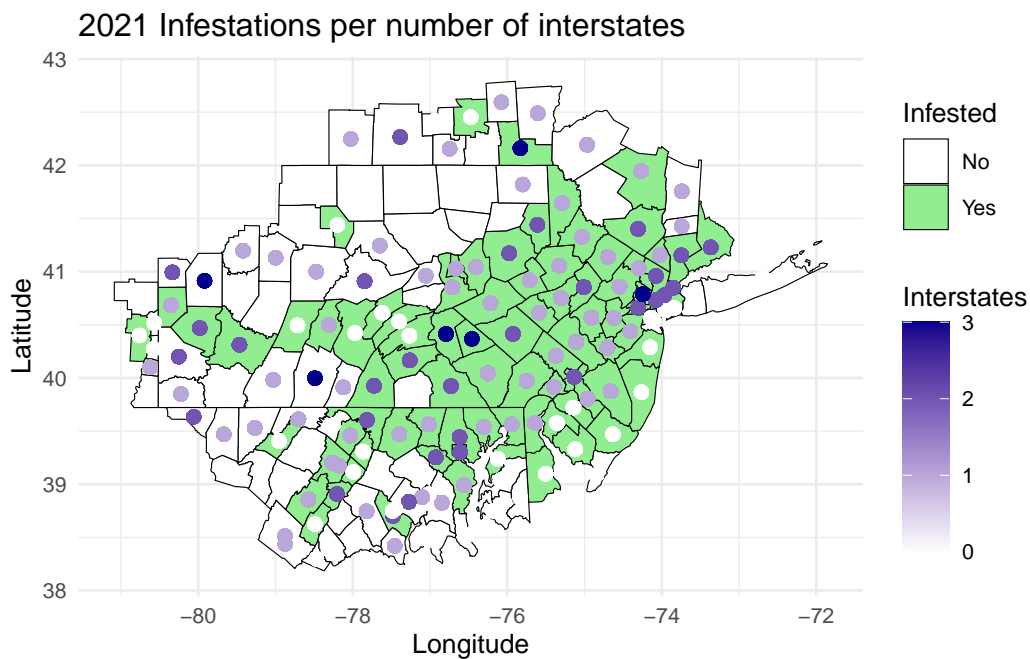| Year	Number Infested
2014	1
2015	4
2016	6
2017	6
2018	18
2019	26
2020	50
2021	88

Figure 3: Spotted laternfly infested counties for 2021 together with the number of two-digit interstate highways transecting the counties.
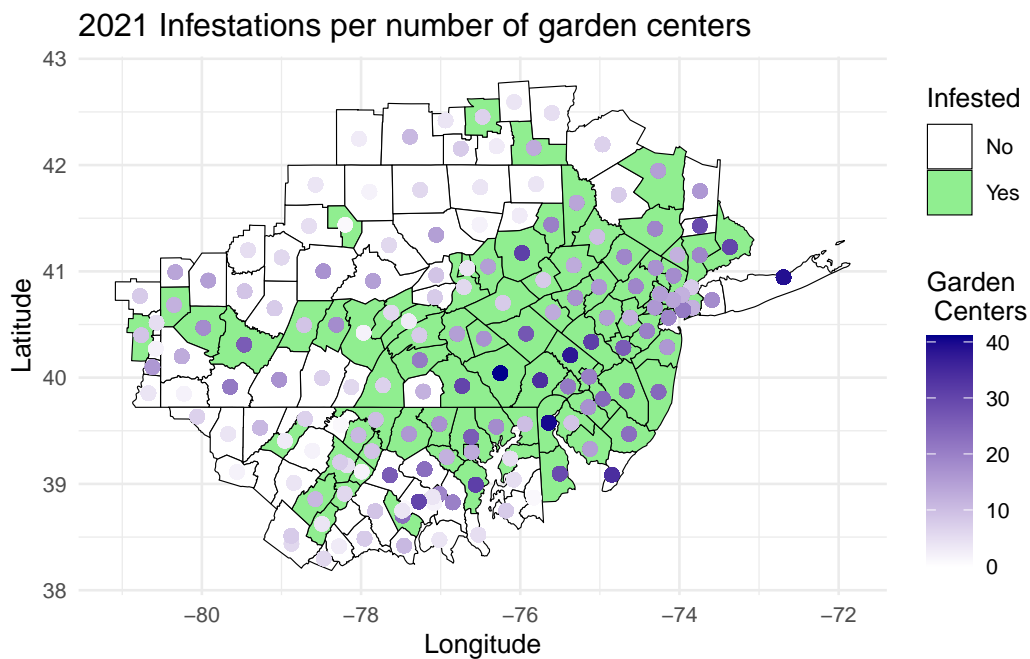
Figure 4: Spotted laternfly infested counties for 2021 together with the number of garden centers in the counties.
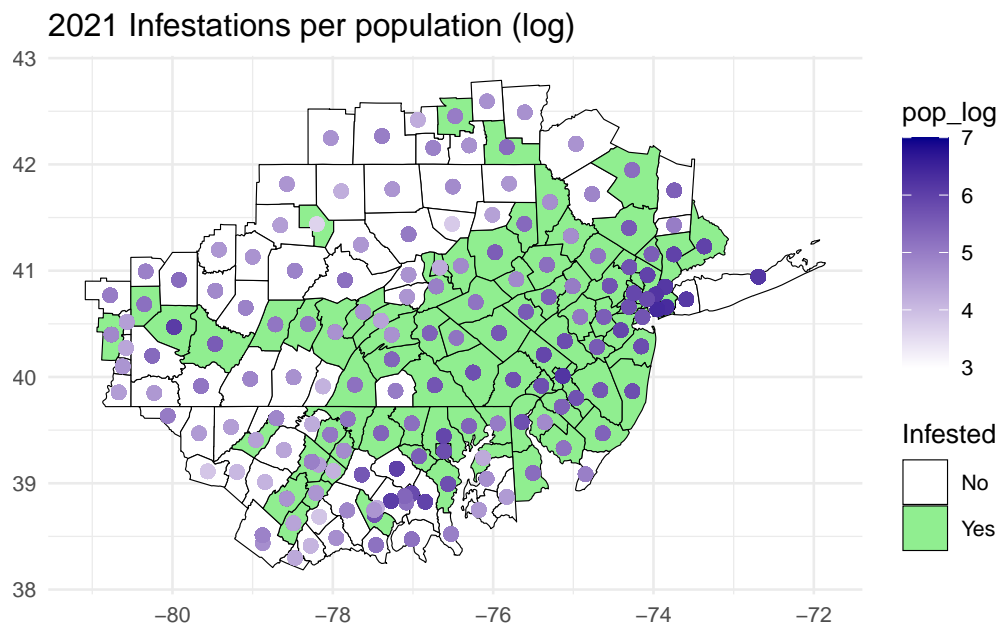

Figure 5: Spotted laternfly infested counties for 2021 together with the 2019 estimated population (log scale) for the counties.
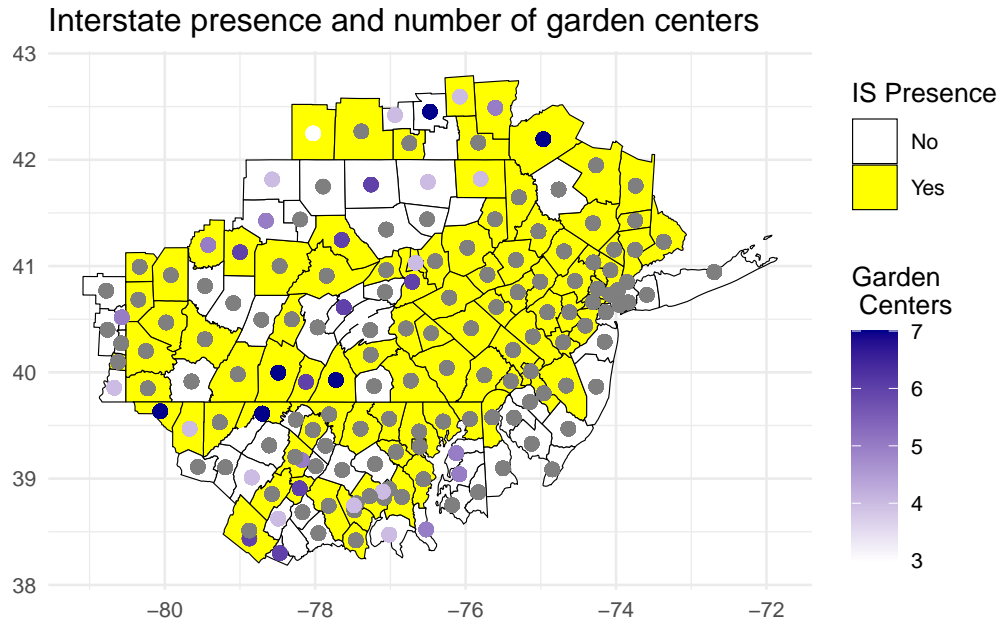
Figure 6: Spotted laternfly infested counties for 2021 together with the presence/absence of two-digit interstate highways and the number of garden center per county.
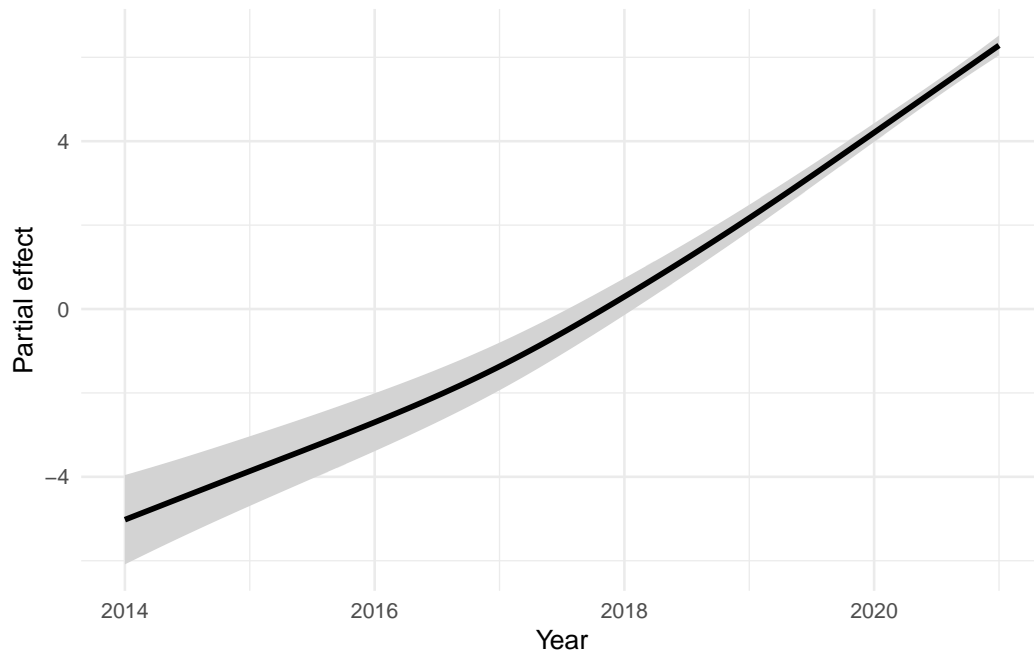

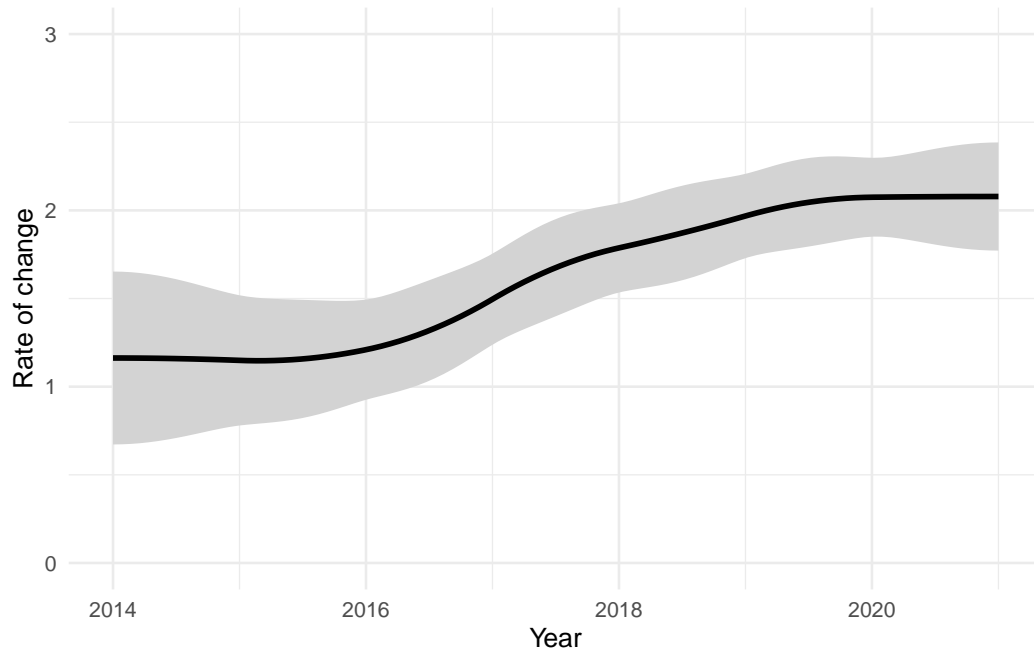Figure 7: Partial effect plot for the log-odds of infestation over time from 2014 - 2021.

Figure 8: Time rate of change (i.e., first derivative) of partial effect of the log-odds of infestation over time from 2014 - 2021.

References

Hartig, Florian. 2022. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models.* https://CRAN.R-project.org/package=DHARMa.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Simpson, Gavin L. 2023. *gratia: Graceful ggplot-Based Graphics and Other Functions for GAMs Fitted Using mgcv.* https://gavinsimpson.github.io/gratia/.

Wood, S. N. 2003. "Thin-Plate Regression Splines." *Journal of the Royal Statistical Society (B)* 65 (1): 95–114.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with r.* CRC press.