Possible Analyses for Revisions

Table of contents

Overview

Using data collected on infestation by the spotted laternfly across 166 counties in the Mid-Atlantic region of the United States, we are interested to know about factors that likely contribute to spread of the organism. Here, we focus on factors related to human activity and that can be targeted for control or mitigation measures. The factors we focus on are:

- The number of businesses designated as garden centers in a county.
- The number of two-digit interstate highways that transect a county.

Further, we also examine the influence of the county population as previous work has established this as an important contributor to lanternfly spread.

A reasonable question: does the data provide evidence that the transportation of materials sold at garden centers over interstate highways contribute to the spread of the spotted laternfly?

The Data

Table 1 displays the first few rows of the data collected for the 166 county Mid-Atlantic region.

For analysis, we add to this data the boundaries for all included counties.

Figure 1 displays the 2021 infestation status for the 166 Mid-Atlantic region counties in 2021:

Table 1: Data collected for the 166 county Mid-Atlantic region.

infested	year	county	state	nis	gc	pop
0	2014	suffolk	NY	0	39	1477000
0	2014	nassau	NY	0	18	1357000
0	2014	fairfield	CT	2	30	943332
0	2014	ocean	NJ	0	24	607186
0	2014	queens	NY	0	12	2253858
0	2014	kings	NY	0	21	2559903

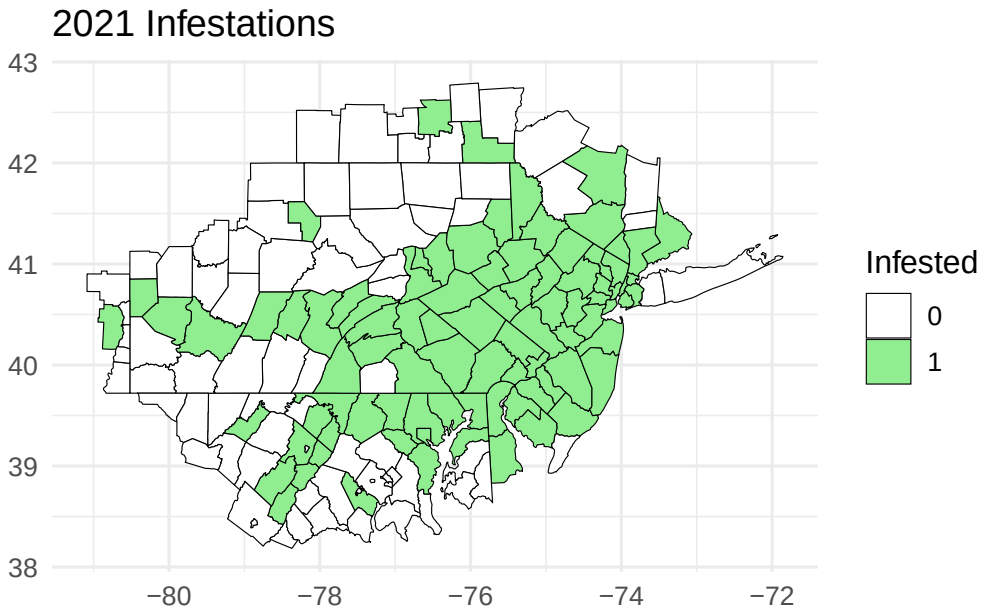

Figure 1: Counties in the defined Mid-Atlantic regions that were designated as infested by the spotted laternfly in 2021.

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

Initial Exploratoration of the Data

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

Figure 2 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.

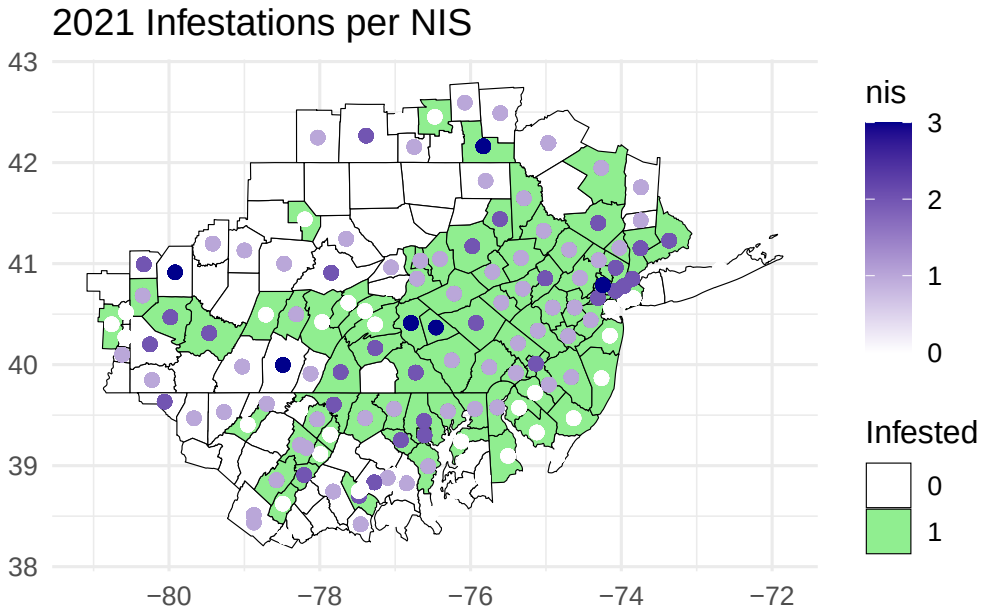

Figure 2: Spotted laternfly infested counties for 2021 together with the number of two-digit interstate highways transecting the counties.

Figure 3 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.

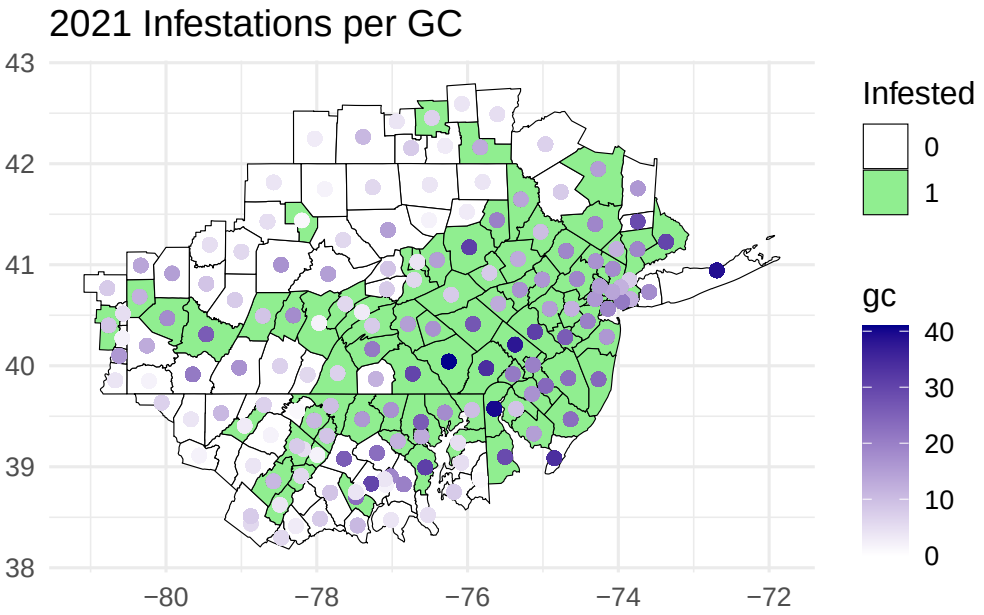

Figure 3: Spotted laternfly infested counties for 2021 together with the number of garden centers in the counties.

Figure 4 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 5 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.

Statistical Analysis

For a statistical test of our hypothesis, we have use R version 4.3.1 and package `mgcv` 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 with linear predictors: 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 centered and scaled the number of garden centers. We used the packages `DHARMa` and `gratia` for diagnostics of model assumptions (Hartig 2022 ; Simpson 2023).

```
mod_form_21 <- infested ~ is_presence + std_gc + is_presence:std_gc + pop_log +
   te(longitude,latitude)
```

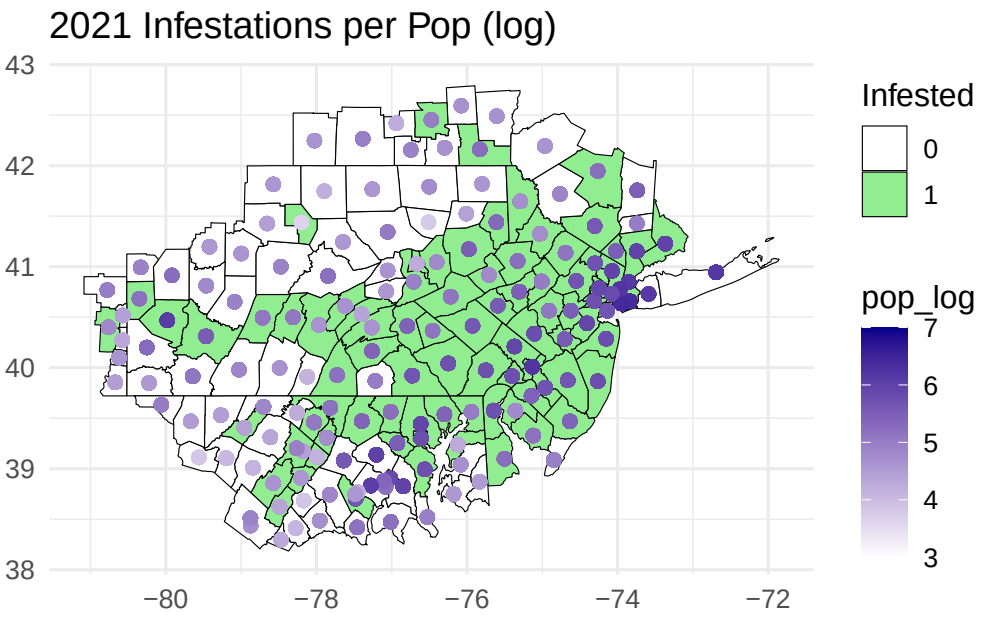

Figure 4: Spotted laternfly infested counties for 2021 together with the 2019 estimated population (log scale) for the counties.

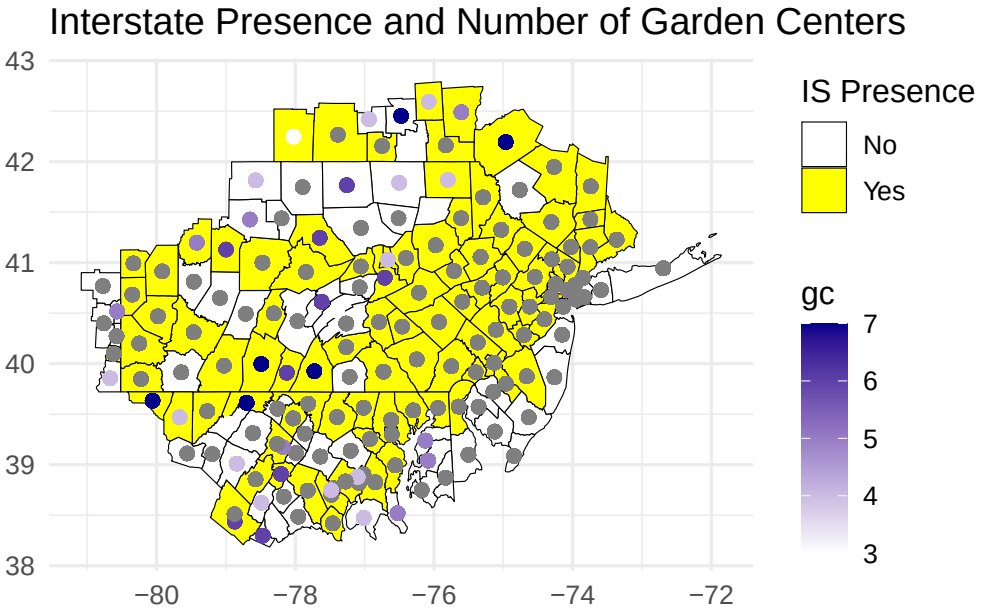

Figure 5: Spotted laternfly infested counties for 2021 together with the presence/absence of two-digit interstate highways and the number of garden center per county.

Table 3: Summary table for estimates for linear terms in generalized additive model.

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
Pop (log)	1.679994	0.8431980	0.0463264
IS Presence - Yes : Garden Centers	1.959781	0.6944467	0.0047714

```
gam_mod_21 <- gam(mod_form_21,data=data_1,
                        family = binomial(),
                        method = "REML")
```

Table 3 summarizes the estimated coefficients for the linear predictors in the generalized additive model.

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.

Expand for Session Info

```
- Session info ---------------------------------------------------------------
 setting  value
 version  R version 4.3.1 (2023-06-16)
 os       macOS Ventura 13.5.2
 system   aarch64, darwin20
 ui       X11
 language (EN)
 collate  en_US.UTF-8
 ctype    en_US.UTF-8
 tz       America/New_York
 date     2023-09-17
```

```
 pandoc   3.1.1 @ /Applications/RStudio.app/Contents/Resources/app/quarto/bin/tools/ (via r
 quarto   1.3.450 @ /usr/local/bin/quarto

- Packages -------------------------------------------------------------------
 package     * version  date (UTC) lib source
 broom.mixed * 0.2.9.4  2022-04-17 [1] CRAN (R 4.3.0)
 DHARMa      * 0.4.6    2022-09-08 [1] CRAN (R 4.3.0)
 dplyr       * 1.1.3    2023-09-03 [1] CRAN (R 4.3.0)
 forcats     * 1.0.0    2023-01-29 [1] CRAN (R 4.3.0)
 ggplot2     * 3.4.3    2023-08-14 [1] CRAN (R 4.3.0)
 gratia      * 0.8.1.34 2023-07-02 [1] Github (gavinsimpson/gratia@f967268)
 kableExtra  * 1.3.4    2021-02-20 [1] CRAN (R 4.3.0)
 lubridate   * 1.9.2    2023-02-10 [1] CRAN (R 4.3.0)
 mgcv        * 1.9-0    2023-07-11 [1] CRAN (R 4.3.0)
 nlme        * 3.1-163  2023-08-09 [1] CRAN (R 4.3.0)
 purrr       * 1.0.2    2023-08-10 [1] CRAN (R 4.3.0)
 readr       * 2.1.4    2023-02-10 [1] CRAN (R 4.3.0)
 sessioninfo * 1.2.2    2021-12-06 [1] CRAN (R 4.3.0)
 stringr     * 1.5.0    2022-12-02 [1] CRAN (R 4.3.0)
 tibble      * 3.2.1    2023-03-20 [1] CRAN (R 4.3.0)
 tidyr       * 1.3.0    2023-01-24 [1] CRAN (R 4.3.0)
 tidyverse   * 2.0.0    2023-02-22 [1] CRAN (R 4.3.0)

 [1] /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/library

------------------------------------------------------------------------------
```