Making the most of invasion records, the case of the spotted lanternfly, part I: isolating jump dispersal and diffusive spread

Nadege Belouard[*] Sebastiona De Bona[†] Jocelyn E. Behm[‡] Matthew R. Helmus[§]

1/6/2021

Contents

[*]Temple University, nadege.belouard@temple.edu
[†]Temple University, seba.debona@temple.edu
[‡]Temple University, jebehm@temple.edu
[§]Temple University, mrhelmus@temple.edu

Aim and setup

The dispersal of a species can be autonomous or vectored, and in the case of the spotted lanternfly, it is strongly suspected that human transportation dramatically increases the spread of the species. While most dispersal events occur over short distances and likely result in a continuous invasive range, anthropogenic dispersal promotes the occurrence of dispersal "jumps", and the establishment of satellite populations away from the core of the invasion. Distinguishing diffusive spread and jump dispersal is important to understand the process of invasion, its evolution, but also to take efficient management measures.

The spotted lanternfly, *Lycorma delicatula* (hereafter SLF) is an insect from China that is an invasive pest in the US. Since the initial detection of SLF in Berks County, PA, in 2014, large-scale surveys were conducted to trace the progression of the invasion, resulting in a large amount of detection and non-detection data. A unique dataset summarizing SLF presence and absence in the US was constructed using the package `lycordata`. This dataset is an opportunity to study the spread of the SLF.

The aim of this first vignette is to differentiate diffusive spread from jump dispersal using a simple and conservative method. We defined a distance that SLF are unlikely to disperse autonomously - here, 10 miles. Then, we calculated the distance of each detection point (established population) to the introduction site, and we looked for gaps in the distribution of this distance that were larger than 10 miles. Every detection of a spotted lanternfly found after such a gap was considered to be a jump event, i.e. an event of anthropogenic dispersal. Considering that the expansion of the invasion is heterogeneous in space, we divided the invasion into disk portions with the introduction site as the origin, to increase the accuracy of the calculations while keeping the analyses reasonably simple.

For the sake of homogeneity with other analyses presented, only established populations (detection involving more than one individual, as defined in `lycordata`) will be used in analyses.

1. Data initialization

Data reshaping

In `lycordata`, each survey appears in a row, and multiple surveys are conducted at the same location the same year, resulting in a long and redundant dataset. It is considered whether SLF are present (one individual found) and/or established (more than one individual or an egg mass found). The table has 276309 rows.

bio_year	latitude	longitude	state	slf_present	slf_established
2015	40.41152	-75.65998	PA	TRUE	FALSE
2016	40.37255	-75.62202	PA	TRUE	FALSE
2016	40.38061	-75.71915	PA	TRUE	FALSE
2016	40.47693	-75.62499	PA	TRUE	FALSE
2016	40.59245	-75.52150	PA	TRUE	FALSE
2017	40.13529	-75.51762	PA	TRUE	FALSE

We reshape the table to summarize the information, so that one line represents the detection status at a given location for each year. We round the geographical coordinates to cells of 1 km^2 (100 m * 100 m). The code is borrowed from Seba De Bona's `lycordata` vignette to homogenize our data.

Note: when several surveys indicate that SLF are "present" the same year at the same location, we could be tempted to categorize them in the "established" category. However, the category "present" often refers to dead individuals, although this information is not explicitly available. We use a conservative approach and kept the same categories while summarizing the data.

bio_year	latitude_rounded	longitude_rounded	slf_present	slf_established
2014	39.79279	-76.83544	FALSE	FALSE
2014	39.83784	-76.30380	FALSE	FALSE
2014	39.84685	-76.43038	FALSE	FALSE
2014	39.85586	-76.60759	FALSE	FALSE
2014	39.86486	-76.88608	FALSE	FALSE
2014	39.88288	-76.50633	FALSE	FALSE

The table now has 47975 rows.

Distances and status calculation

We first calculate the distance between each survey point and the introduction point (-75.675340, 40.415240, from Barringer et al., 2015). This distance will be the basis of all subsequent analyses. The summary of this distance is:
Min. : 0.389 , 1st Qu.: 66.969 , Median : 108.736 , Mean : 190.511 , 3rd Qu.: 233.561 , Max. :3820.250
We also create a variable containing the status of SLF for each point: Undetected, Present, Established.

Space division

We divide the invasion records into 8 disk portions to increase the accuracy of subsequent calculations.

Disk portions are named according to the points of the compass: WNW, NNW, NNE, ENE, WSW, SSW, SSE, ESE. Established SLF populations are represented on a map (Figure 1), colored by disk portions.

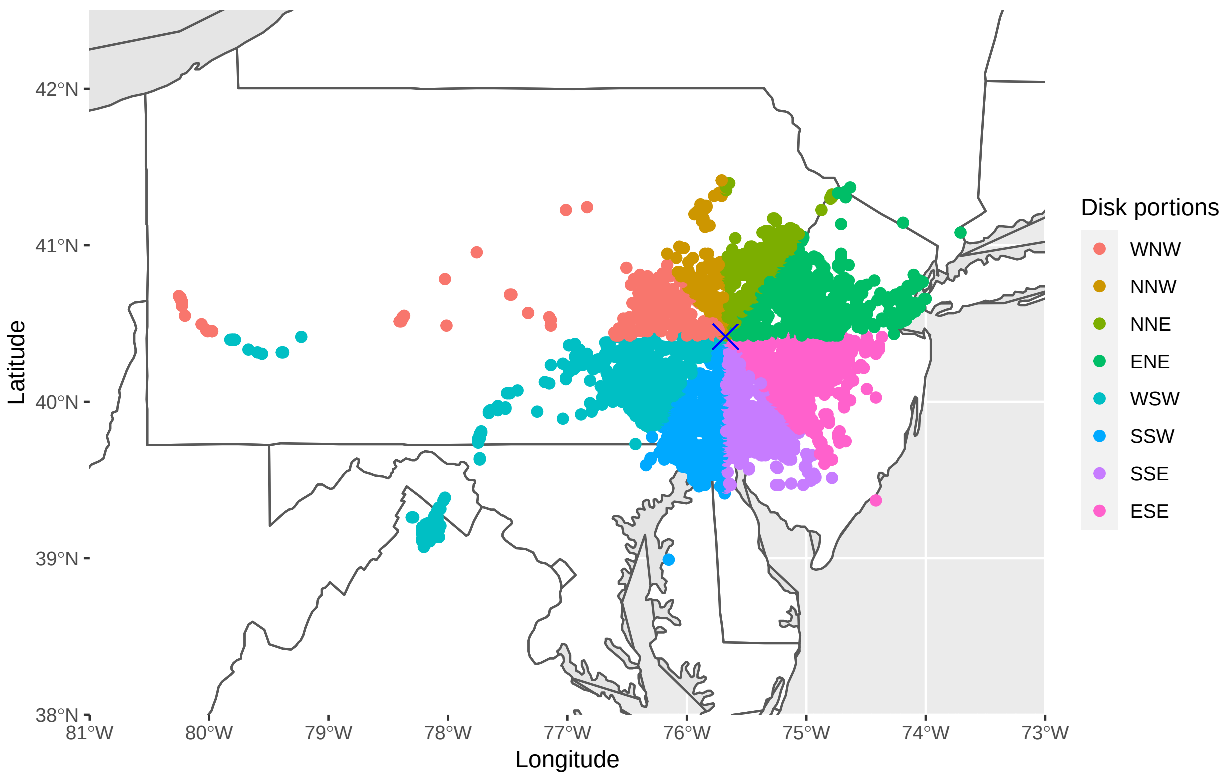

Figure 1: Map of SLF establishment

2. Yearly radius of the invasion

To estimate the spread of the SLF, we extract for each year the radius of the invasion, defined as the maximum distance of an established population to the introduction point. Median or mean distances are not informative here because surveys are preferentially conducted towards the invasion front, and thus bias the distribution of distances.

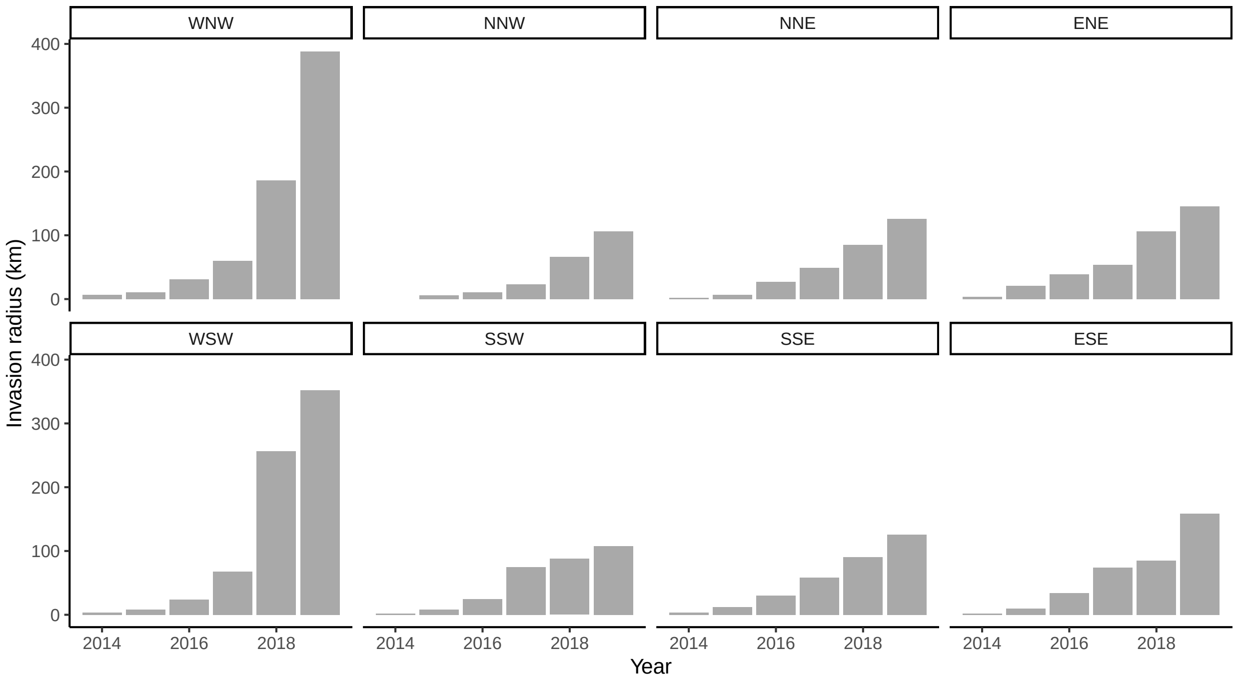

Figure 2: Yearly radius of the invasion

The invasion radius increased regularly until 2017 (Figure 2). From 2018, there is a steep increase in the invasion radius, especially in the westernmost disk portions, denoting the apparition of dispersal jumps.

3. Differentiating diffusive spread and jump dispersal

Exploration of histograms of distances to the introduction site

Let's have a look at the distribution of the distances of established populations to the introduction point (Figure 3).

The fact that non-detection events are always recorded further than detection events indicates that we can be be fairly confident that the spread of the SLF is accurately monitored. We can also see that the distribution of established populations is sometimes discontinuous, with gaps where populations were not detected. Detections that appear after the first gaps are likely the result of jump dispersal, i.e. human-vectored transportation of SLF in new locations (secondary introductions).

We can further understand the yearly spread of the SLF by distinguishing diffusive spread (the continuous progress of the invasion) and jump dispersal (long-distance, human-vectored dispersal). We calculate, for each year, the limit of diffusive spread by finding the first 10-mile gap in the distribution of the distances to the introduction site.

Function differentiating diffusive spread and jump dispersal

A custom program searches the distance at which the gap occurs, and returns both the survey before this threshold (the limit of diffusive spread) and a list of surveys found after this threshold (jump events). The function runs over each year and disk portions in a row.

Here, the program is run with a gap of 10 miles (16 km), from 2014 to 2019, and for 8 disk portions.

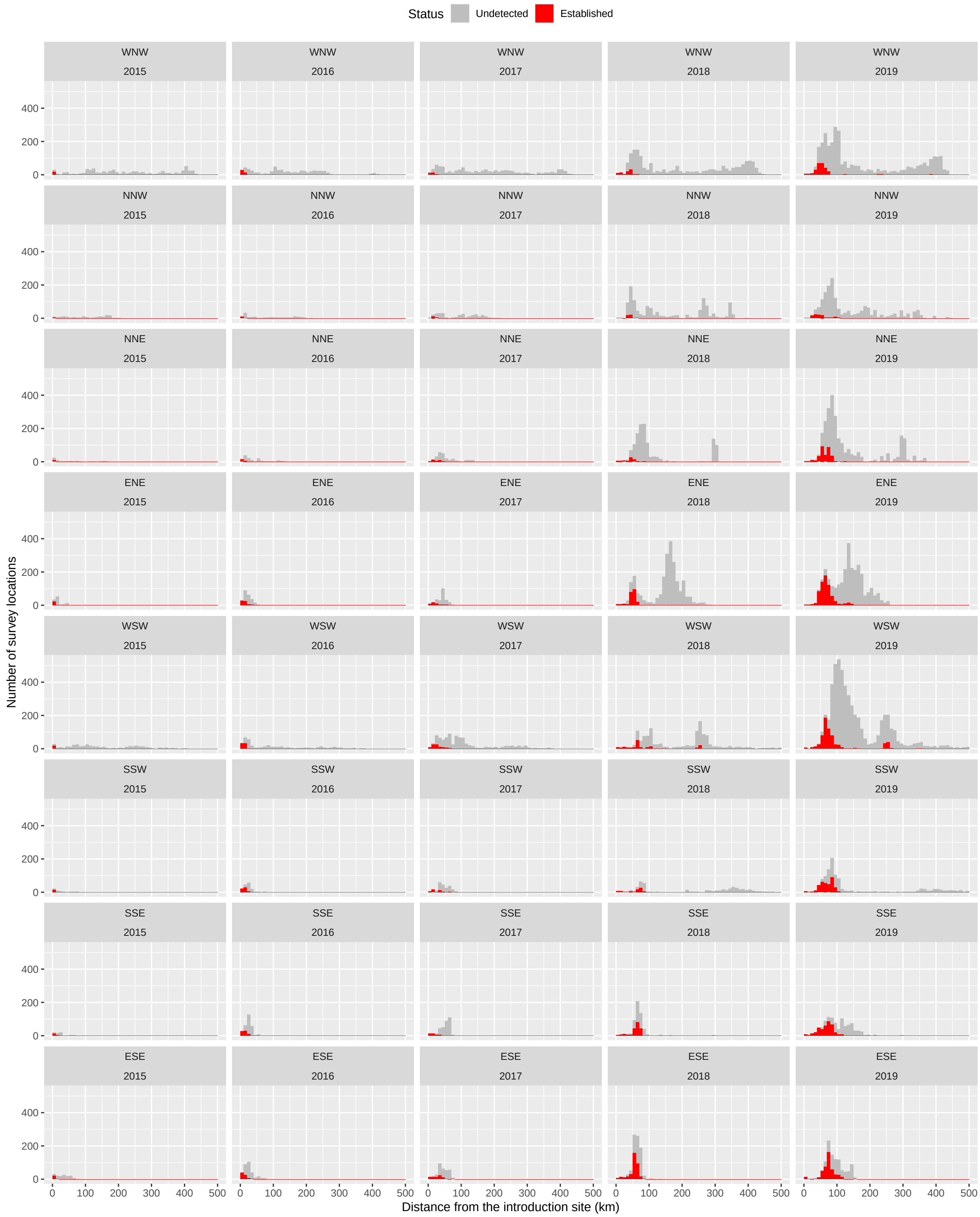

Figure 3: Distribution of the distance between SLF populations and the introduction site, per disk portion and per year

4. Results

Jump locations

The 75 jump events found by the function can be visualized on a map (Figure 4). Jump locations appear in red, among all the established populations in grey. The introduction site is signaled by a blue cross. We note that most jump events occur in northern Virginia or western Pennsylvania. In Winchester (VA), a diffusive spread appears around jump events, indicating that a secondary invasion began in this area. A similar pattern is found around Harrisburg (PA), although the diffusive spread has now reached Harrisburg too.

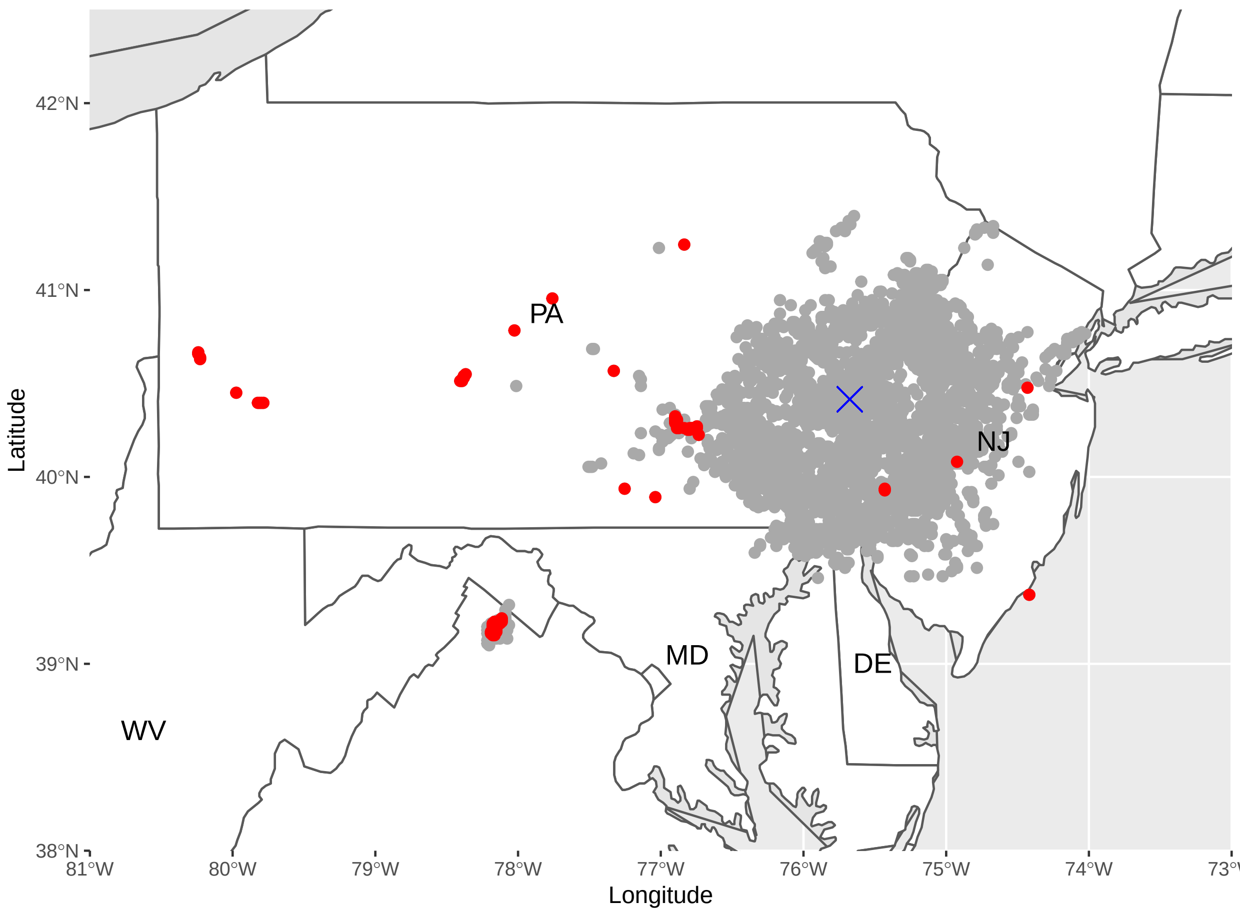

Figure 4: Map of SLF jumps

Evolution of the radius of diffusive spread and jumps over time

We can now look at how the radius of the invasion increases over time, when differentiating diffusive spread and jump dispersal (Figure 5). In the westernmost disk portions, jump dispersal is responsible for the very high increase in the invasion radius. In the other disk portions, the spread seems to be mostly linked to diffusive dispersal.

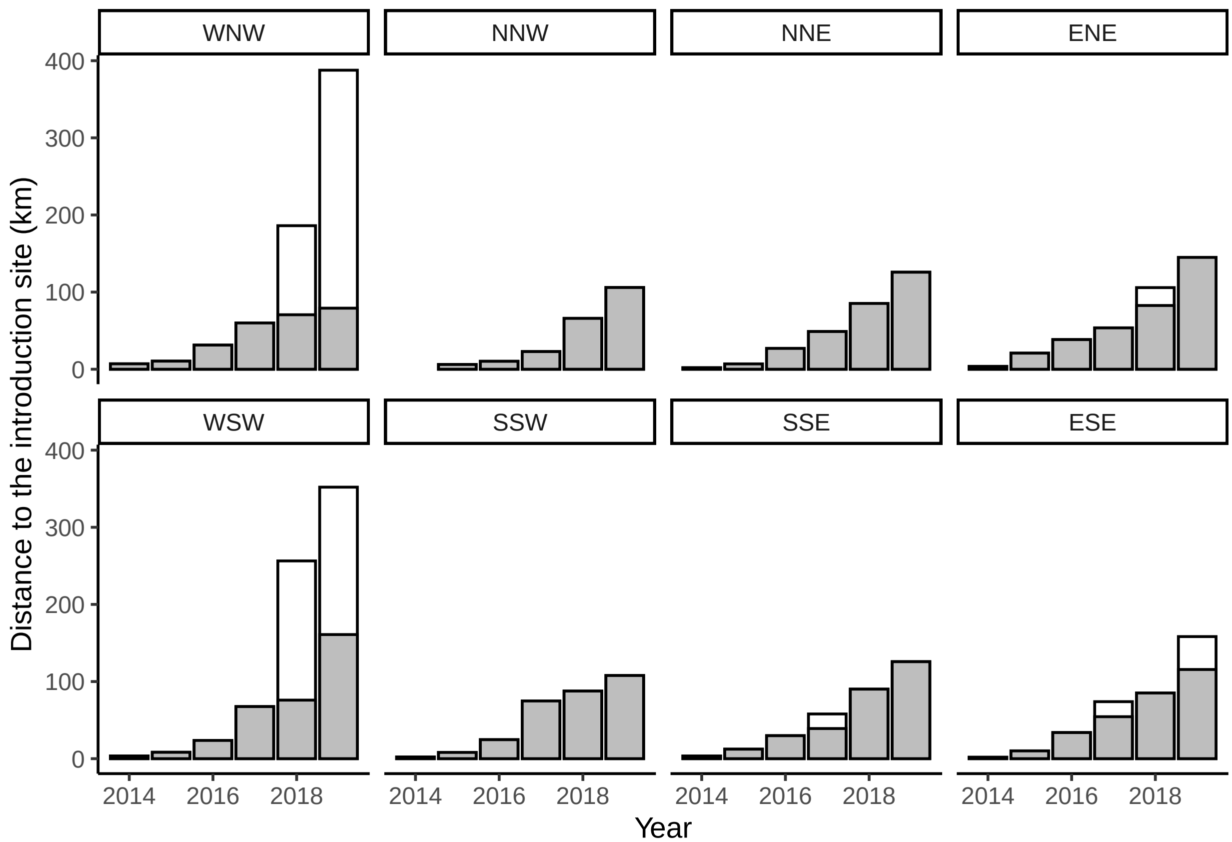

Figure 5: Evolution of the radius of the invasion over time, when diffusive spread and jump dispersal are separated

Yearly net spread of the invasion

Lastly, we can monitor the yearly progress of the diffusive spread (Figure 6). Different biological and evolutionary mechanisms, including spatial sorting, can lead to an increase of the spread of invasive species over time. This can be monitored by looking at the yearly increase in the invasion radius. Here, the tendency is variable between disk portions, but in some cases, the spread seems to increase every year, somewhat conform with the prediction.

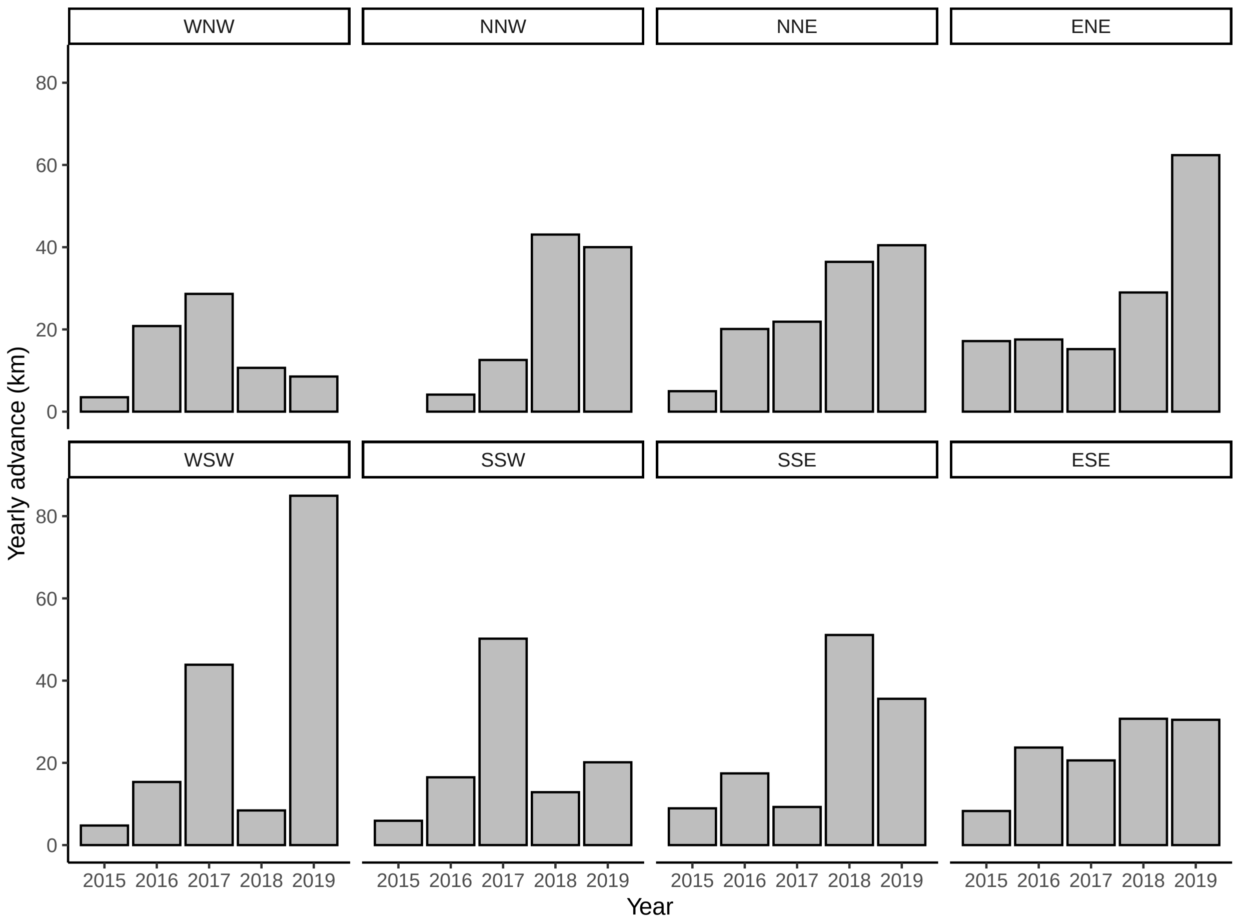

Figure 6: Yearly progress of the diffusive spread

5. Conclusion

The spread of the spotted lanternfly in the US is likely due both to diffusive spread and human-assisted jump dispersal. 75 jump occurrences have been identified, and most of them are situated in Winchester (north VA) and western Pennsylvania (Harrisburg through Pittsburgh). The spread of SLF in the US continues every year, at an increasing rate in most directions.

Jump events are likely be caused by SLF hitchhiking on human transports, and establishing near transport infrastructures: railroads, roads, and airports. In the next vignette, we will test the significance of the proximity between jump events and transport infrastructures by a comparison with a random distribution. We will also compare these distances to those of diffusers (SLF spread through diffusive spread) and of points where SLF were not detected, to check for a potential bias in survey locations.