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`, and constitutes 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 calculated the distance between each detection point and the introduction site. We defined a distance that SLF are unlikely to disperse autonomously - here, 10 miles (16 kilometers). Then, we looked for gaps larger than 10 miles in the distribution of the distance to the introduction site. Every detection of SLF found after such a gap was considered to be a jump event, i.e. an event of anthropogenic dispersal potentially leading to the establishment of a new population. The threshold of the diffusive spread was considered to be the last positive survey before this gap. 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 states whether SLF were present (one individual found) and/or established (more than one individual or an egg mass found). Multiple surveys were conducted at the same location during the same year, resulting in a complex and redundant dataset of 327346 rows (surveys).

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 by rounding the geographical coordinates to cells of 100 m^2 (100 m * 100 m), so that one line represents the detection status at a given location for a given year. 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 57276 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 (in kilometers):
Min. : 0.389 , 1st Qu.: 71.885 , Median : 117.549 , Mean : 196.430 , 3rd Qu.: 238.876 , Max. :3884.595
We also create a variable summarizing the status of the survey for each point: SLF 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. All surveys with established SLF are represented on a map (Figure 1), colored by disk portions.

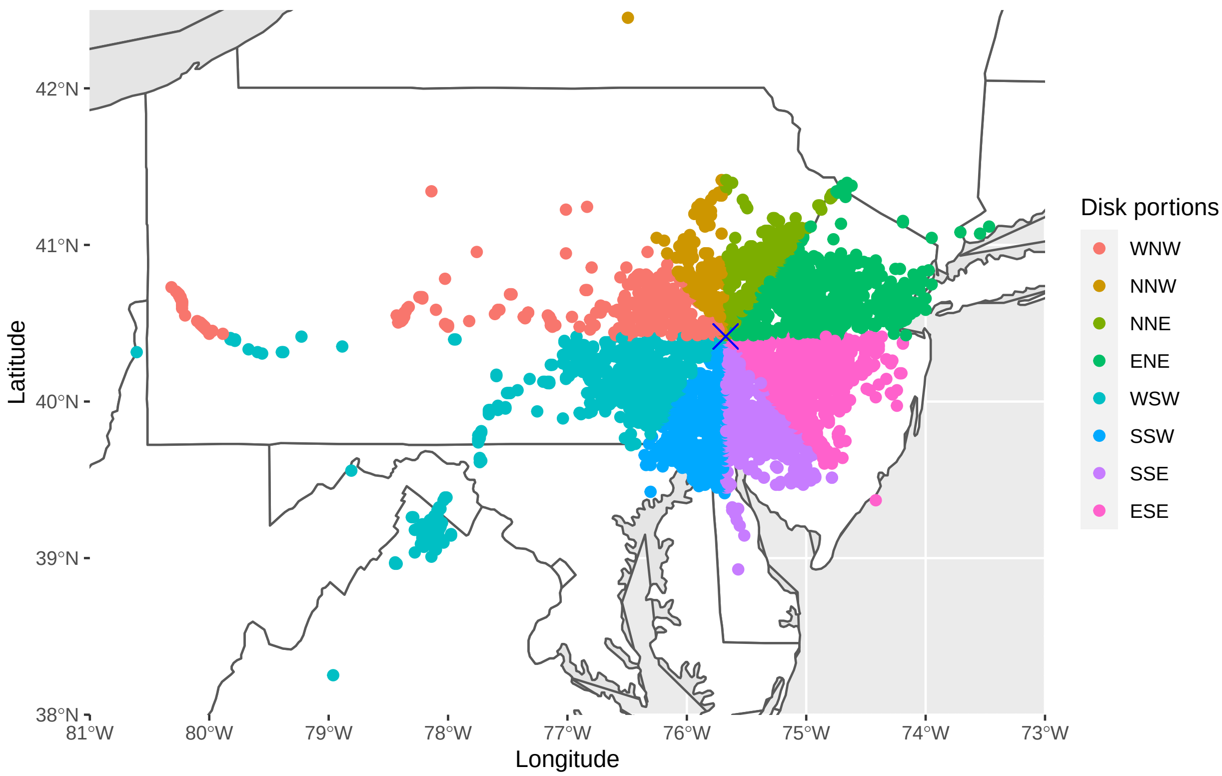

Figure 1: Map of surveys with established SLF colored by disk portion (2014-2020)

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 a survey with established SLF 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.

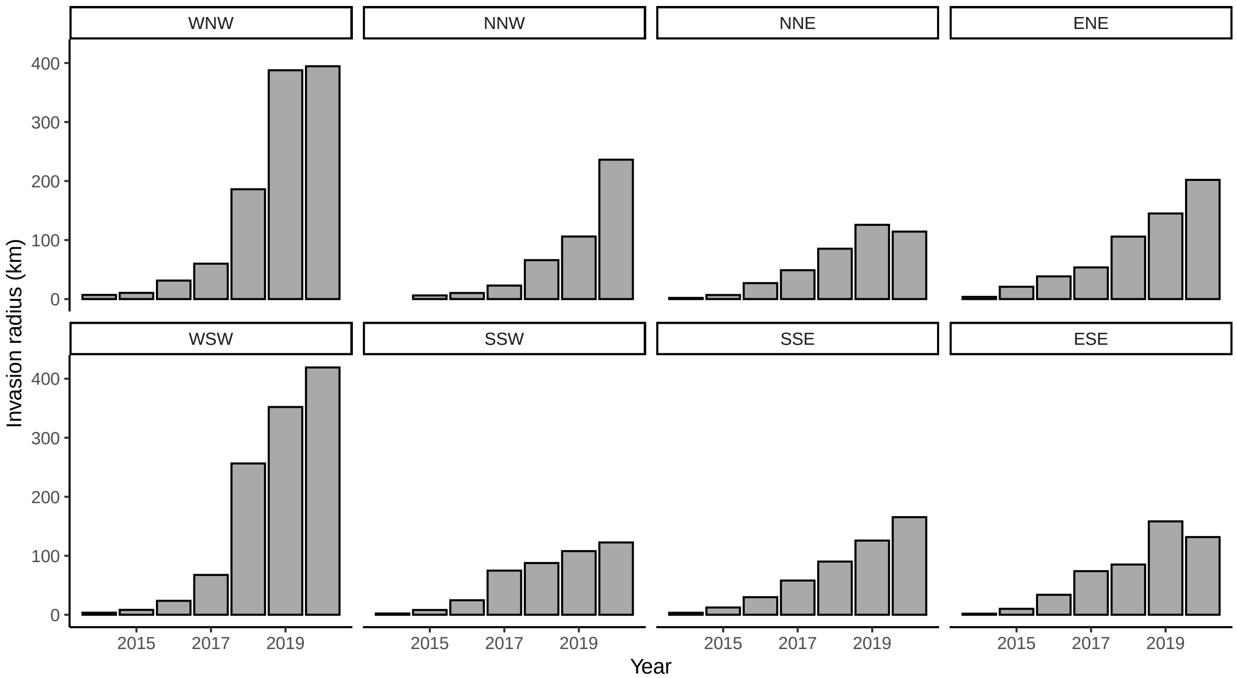

Figure 2: Yearly radius of the invasion

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. Note that in 2020, the invasion radius did not increase dramatically, and even decreased in the NNE and ESE portions. This might be due to fewer surveys being conducted that year because of the covid19 pandemic, or be an actual biological pattern. The accuracy of these values can be investigated by checking whether negative surveys are found further than positive surveys, to make sure that SLF were not missed.

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 positive vs negative surveys to the introduction point (Figure 3).

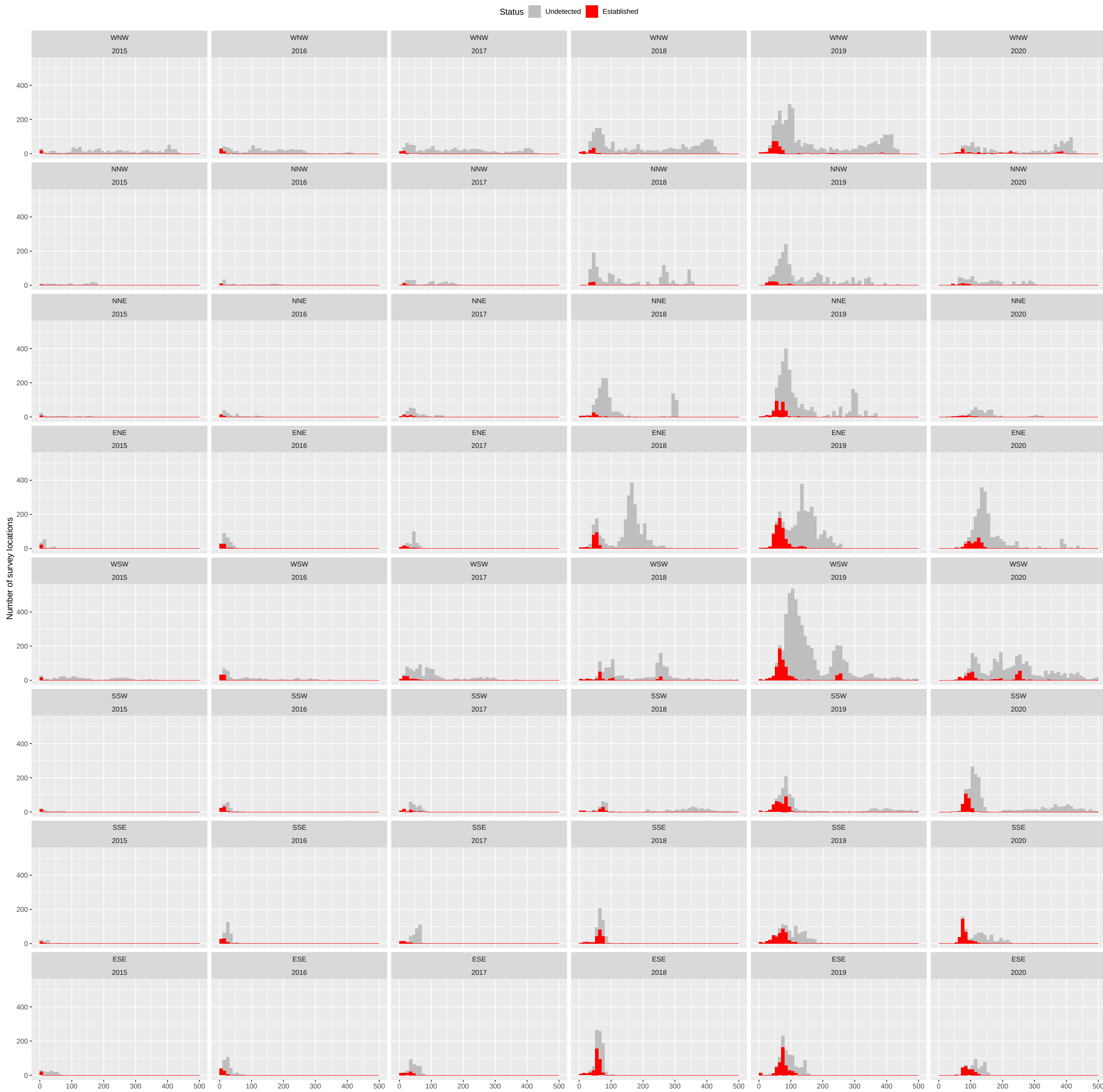

Figure 3: Distribution of the distance between SLF populations and the introduction site, per disk portion and per year

The fact that surveys with SLF undetected are always recorded further than detection events indicates that we can be be fairly confident that the spread of the SLF is accurately monitored. It is also the case in 2020 in the NNE and ESE portions. 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 gap 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 calculated, 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 for each year 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).
* Note that here, we consider that populations do not go extinct, so that the limit of the diffusive spread cannot be lower in year y than in year y+1. This is because fewer and fewer surveys are conducted near the introduction site over time, leading to the appearance of a false first gap near the introduction site (see Figure 3, surveys are shifted on the right in 2019 and 2020).
* If a jump event is identified closer than 10 miles to a jump from the previous year, it is removed from the list, as SLF likely spread from the jump of the previous year. * The function runs independently for each disk portion, generating false-positive and false-negative (see troubleshooting with disk rotation).

Here, the program is run with a gap of 10 miles (16 km), from 2014 to 2020, and for 8 disk portions.

4. Results

Jump locations

The 82 jump events found by the function can be visualized on a map (Figure 4). Jump locations are colored according to their year of appearance, 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.

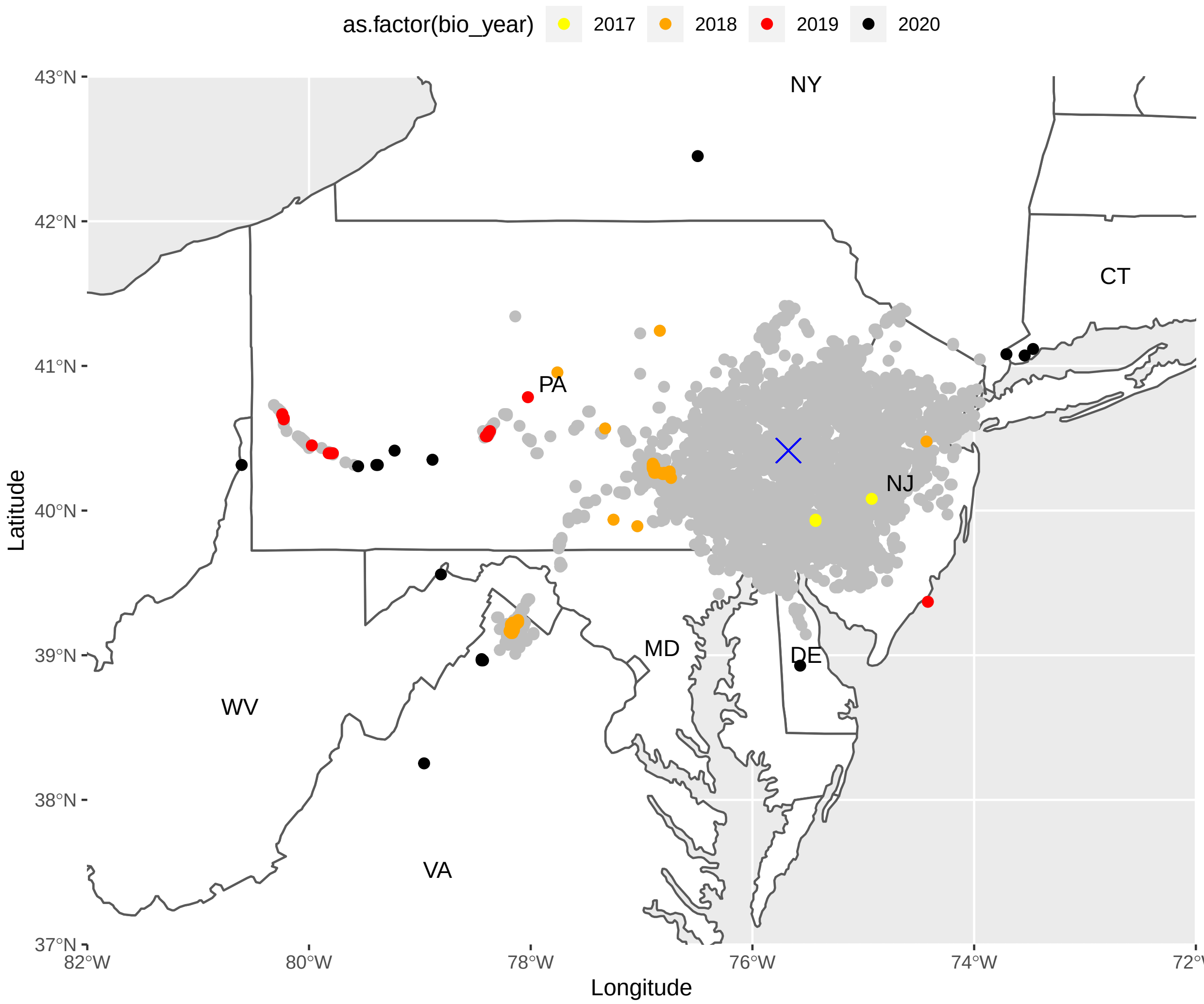

Figure 4: Map of SLF jumps

	Harrisburg	Winchester	Other
2017	0	0	3
2018	21	28	6
2019	0	0	17
2020	0	0	16

Most jump events occurred in Harrisburg, PA (21 jumps), and Winchester, VA (28 jumps) in 2018. They might be true independent jumps, i.e. SLF hitchhiked multiple times to these locations the same year. Alternatively, they might be the result of SLF quickly spreading from a single jump event. Finally, they can be a mix between these two hypotheses. For the rest of the analyses, we will test the two most contrasted hypotheses in parallel in order to test whether results vary. For the first hypothesis (all points are independent introductions), the dataset consists of all 82 jump points. For the second hypothesis (only one introduction in Harrisburg and Winchester), the dataset consists of 44 jumps, with the jumps in Harrisburg and Winchester summarized each by their most central point.

Figure 5 is the map of the 44 jump events under the hypothesis of non-independent jumps to Harrisburg and Winchester (replaced by their most central point).

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 6). 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.

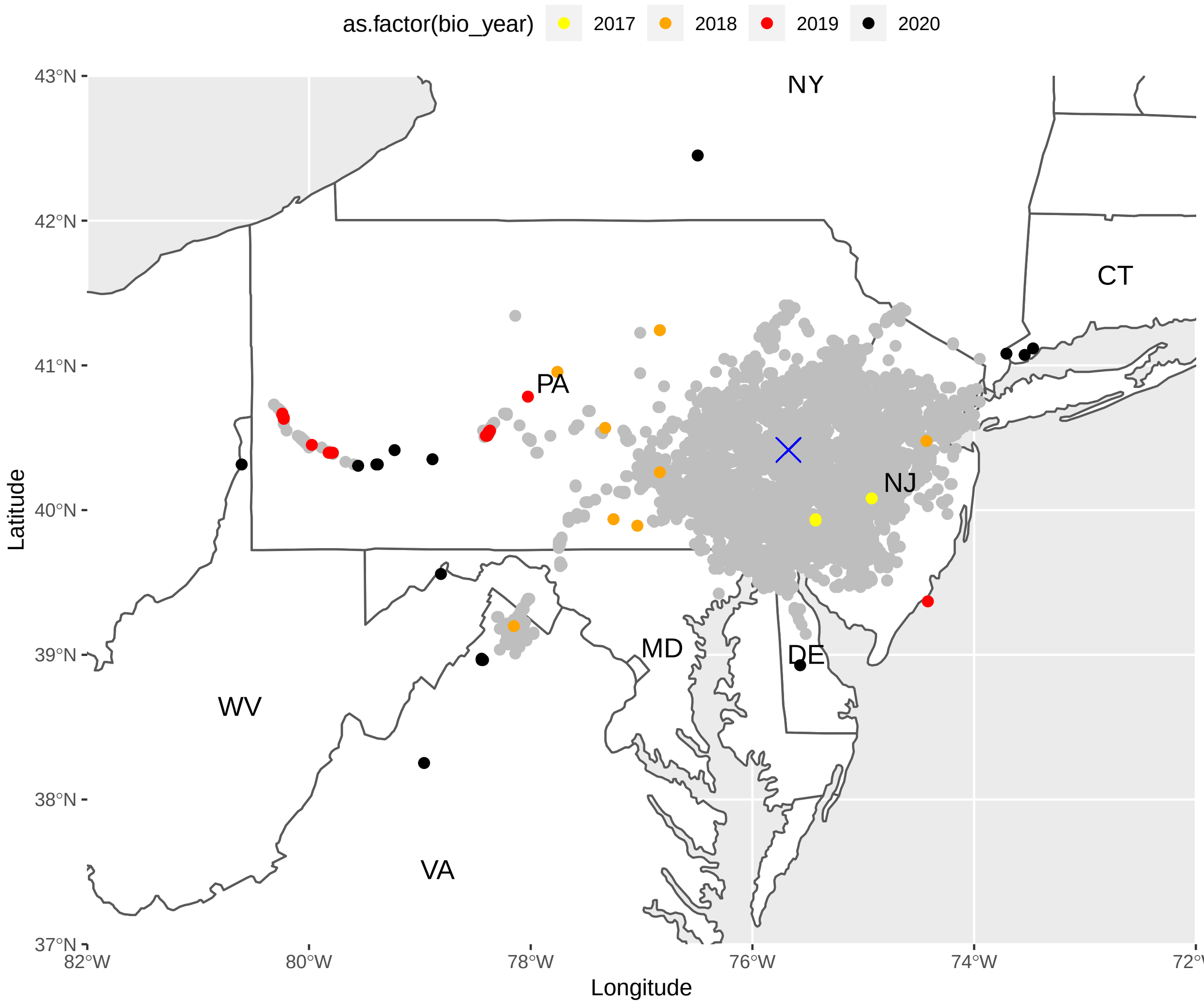

Figure 5: Map of SLF jumps without simultaneous jumps

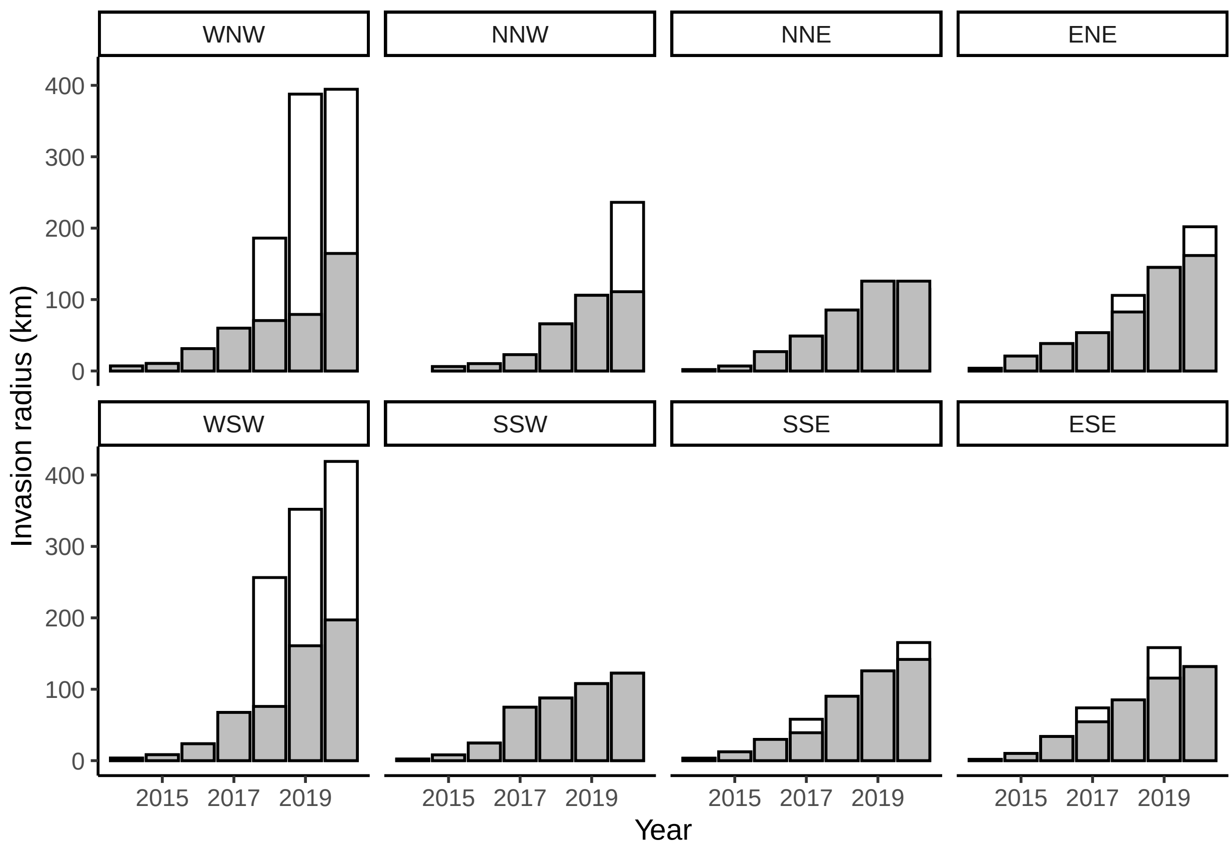

Figure 6: Evolution of the radius of the invasion over time, when diffusive spread and jump dispersal are separated

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 (especially Harrisburg).

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.