2 Prediction

2.1 Examination

As the start point for the further research, the spread of hornets is need to related to time. The first problem here is how to use data to represent the spread of hornets. we use time series analysis to examine this time series is a weak stationary process and this value is relavant to time. The amount of hornet sighting is adopted to illustrate the increasement of hornets amount, so the amount variation of hornets by time is discuessed. However, since the amount of positive samples are 0 is common in each month, it's not suitable to discuss this varietion in month. On the other hand, it's still unreachable to see the variation in years since the data only mainly include two years. Under the comprehensive consideration of the Seasonal habits of hornets as the profile said, using quarter (three months) as the time window to have an effective statistics (Figure 4). So devide 2019.01~2020.12 to 8 quarters.

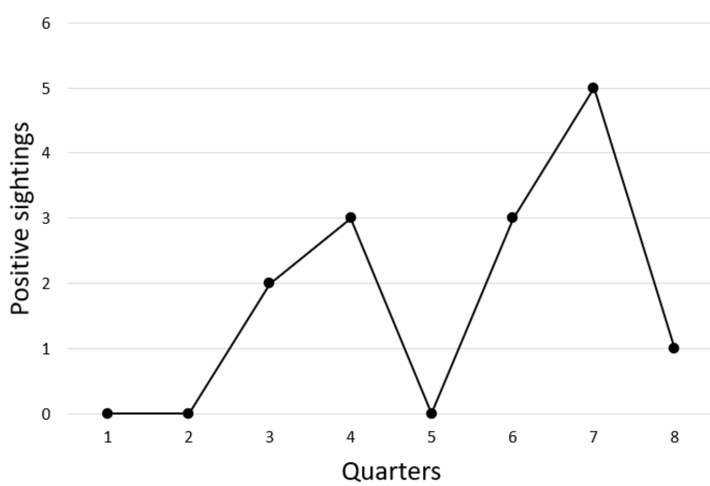


Figure 1: Number of Positive Sighting by Quarters

2.1.1 Weak Stationary Process

First, we examine weak stationary process of this time series, using augmented Dickey-Fuller test (ADF).

Hypothesis: $H_0 : \rho \geqslant 1, H_1 : \rho < 1$; H_0 represents the series is stationary and H_1 represents the series is not stationary.

Test statistic:

$$t = \frac{\hat{\rho} - 1}{\hat{\sigma_p}}$$

In this series, the sample data is the amounts of quarters, $Y = ()$

Through the calculation of test statistic t, and designating the confidence level, the probability of t in the confidence interval (CI) could get by seeking existed table. Another method is calculate the probability of test statistic t not in the CI, $p - value$. If this value is far less than confidence level C, it is considered H_0 is true with sufficient confidentce, or it is considered H_1 is true.

Let confidence level C to be 0.05, after calculated the $p - value = 0.846$. Since $p - value$ is far greater than confidence level, the series is related to its own history. In other words, the data of each moment is related to the previous historical data.

2.1.2 Correlation

The time was regarded as the independent variable and the number of sightings as the dependent variable, and the correlation was directly calculated.

After calculation, $p - value = 0.167$, far less than 0.05 of confidence level. And the correlation level $C = 0.540$, shows a moderate positive correlation between the two.

2.2 Prediction Method

2.2.1 Area Division

Cellular automata (CA) is mainly used to predict the spread of hornet. To use this method, the map is needed to be divided up to Cellulars before further process. We divide the area in this ways:

- The rectangular area of 45° N to 50° N and 120° W to 130° W was taken as the research object.
- Latitude and longitude are drawn every 0.2 degrees, dividing the graph into small rectangles.

According to HaverSineg equation of spherical distance

$$haversin(\frac{d}{R}) = haversin(\varphi_2 - \varphi_1) + cos(\varphi_1)cos(\varphi_2)haversin(\Delta\lambda) \quad (1)$$

$$haversin(\theta) = sin^2(\frac{\theta}{2}) = \frac{1 - cos\theta}{2}$$

Here, R is the redius of earth, φ_1 and φ_2 is the latitude of two points and $\Delta\lambda$ is the absolute value of the difference between the longitude of two points.

In this way, the division ensures that the center of one rectangle is about 20~30 km away from the eight adjacent rectangles, which is the maximum distance that a fertilized female can fly.

2.2.2 Cellular Automata

The cellular automata consists of the following elements:

- Space: Two-dimensional grids on the map.
- State: The state of each grid is defined as $s_{i,j}$, which represents the number of active nests of hornet in each grid.
- Neighbor: The neighbors of a certain cell are defined as its 8 adjacent cells.
- Evolution rules: The rules of evolution can be summarized as

$$s_{i,j}^{t+1} = f(s_{i,j}^t, s_{neighbor_{ij}}^t)$$

Assuming that the flight distance of the female hornet obeys the Poisson distribution, the eight orientations of the flight process are equal probabilities. Because the probability of the Poisson distribution is maximized around parameter $\lambda = \frac{30+0}{2}$, meaning in the best condiction it fly 30 km away, and in the worst consition it only fly hundreds of meters away, which is rounded to 0 km. At the same time, the population size limit of hornet within a grid is mainly dependent on latitude and can be simply assumed to be linearly negative.[1] At the same time, suppose that an hornet produces a total of $eggs$ during the breeding period, and the probability of each egg successfully hatching is $alive\_prob$. Here, according to the history record, let $egg = 20000$, $alive\_prob = 0.00001$ and the probability that a female bee is fertilized is $fert\_prob = 0.35$.[2]

Evolution rules:

- $0 \leqslant s_{i,j} \leqslant limit$
- Only the influence of 8 cells around one cell and its own state on the next state is considered. Since the distance between neighbor cells in the southeast and northwest is slightly different from that of neighbor cells in the southeast, northwest, northeast and southwest, their contributions are considered separately. Note that the probability of female hornets flying along the southeast, southwest, northeast and northwest directions without flying out of the current grid is p_1, the set of neighbors in these four directions is S_1, and the probability of female hornets flying along the southeast, southwest, northeast and northwest directions without flying out of the current grid is p_2, the set of neighbors in these four directions is S_2.

$$
\begin{aligned}
s_{i,j}^{t+1} = & fert\_prob \cdot eggs \cdot alive\_prob \\
& \cdot \left(s_{i,j}^{t} \cdot (p_1 + p_2) + 0.25 \sum_{k=1}^{2} (1 - p_k) \cdot \sum_{(i,j) \in S_k} s_{i,j}^{t} \right)
\end{aligned} \tag{2}
$$

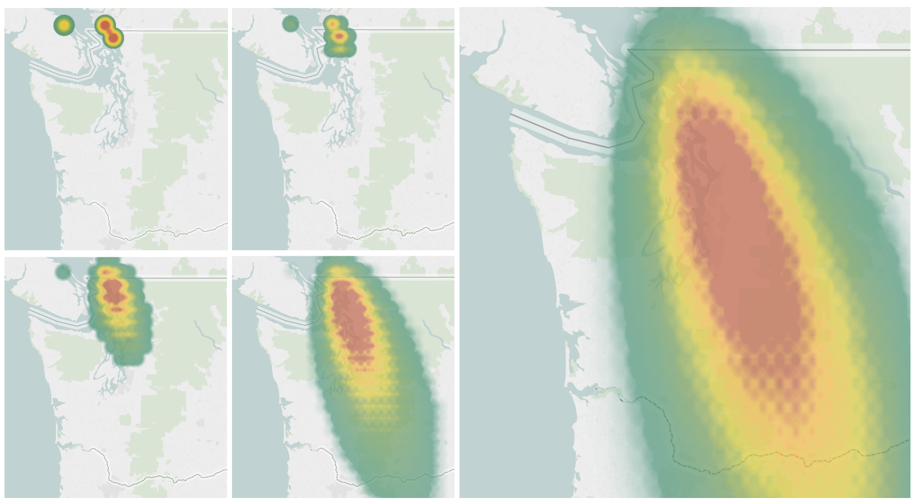
2.3 Result

The initial data starts with several sightings reported in 2019 at the bumblebee invasion site, where there is initially only one active hive.

It can be seen from the predicted results that the initial population size and the area involved were small, but with the passage of time, the population size increased rapidly and the range expanded rapidly, which

Global ID
{5AC8034E-5B46-4294-85F0-5B13117EBEFE}
{5EAD3364-2CA7-4A39-9A53-7F9DCF5D2041}
{124B9BFA-7F7B-4B8E-8A56-42E067F0F72E}
{F1864CC3-508C-4E60-9098-B158AB413B03}
{7F3B6DB6-2ED4-4415-8DC2-3F03EC88F353}

Table 1: Positive Reports in 2019



(a) 2020-2023 (b) 2024

Figure 2: Prediction

is in line with the characteristics of the rapid spread of invasive species.By 2024, the western part of Washington state is almost completely lost, with a trend of rapid growth in the central part.In some areas, populations of Asian bumblebees could reach more than 1,000.From the direction of transmission, more inclined to the south of the spread, and the north of the spread is slower.This is because high latitudes largely limit the maximum population size in each grid, which makes the population size in high latitudes spread more slowly than that in low latitudes, resulting in a situation that the south is fast and the north is slow.