Research on the Dissemination Route of Vespa Mandarinia Based on Grey Prediction Model

Yongzhe Liang, Qianmiao Guo, Zheng Zhang

School of Economics and Management, Beijing Jiaotong University, Weihai, Shandong 264401

Keywords: Vespa mandarinia, gray model, machine learning, image recognition.

Abstract: Vespa mandarinia is the world's largest hornets and its existence is dangerous. In September 2019, the nest of Vespa mandarinia was discovered in British Columbia, Canada. After that, there were thousands of sightings reported.In this paper, we will introduce the method of predicting the spread of this pest over time using gray prediction model, and using machine learning tech and classification learner to build models to predict the likelihood of mistaken classification.

1. Introduction

In 2019, a pest called the Asian Hornet was discovered on Vancouver Island, Canada. Even if it was quickly destroyed, the entire incident has spread throughout the region. Based on several confirmed pest sightings and several false sightings, it constitutes a map of bumblebee surveillance and public sightings. Bumblebees are predators of European honeybees and can destroy the entire European honeybee colony in a very short time. Therefore, according to the life characteristics of this pest and some sightings reported by people, we need to establish various models to predict and analyze the spread and existence of this pest.

2. The Prediction of the Spread of the Pest

For the first question, to discuss whether the spread of this pest in a period of time can be predicted, we use the gray prediction model in Matlab for analysis.

White system means that the internal characteristics of a system are completely known, that is, the system information is completely sufficient. A black system is a system whose internal information is unknown to the outside world. It can only be observed and studied through its connection with the outside world.

The gray system is between white and black, and part of the information in the gray system is known, while the other part of the information is unknown, and there is an uncertain relationship between various factors in the system. The gray model (Grey Model, referred to as GM model) is generally expressed as GM (n, x) model, its meaning is: use n-order differential equations to establish a model for x variables. The gray forecasting model is to consider the discrete data scattered on the time axis as a set of continuously changing sequences, using accumulation and subtraction methods to weaken the unknown factors in the gray system, strengthen the influence of known factors, and finally construct a continuous differential equation with time as the variable is used to determine the parameters in the equation through mathematical methods to achieve the purpose of prediction.

The prediction principle of the GM(1,1) model is to generate a set of new data sequences with obvious trends in a cumulative manner for a certain data sequence, build a model for prediction according to the growth trend of the new data sequence, and then use cumulative subtraction method performs reverse calculation, restores the original data sequence, and then obtains the prediction result. Sets the original number column:

$$x^{(0)} = (x^0(1), x^0(2), \dots . x^0(n))$$

$$\text{Make } (k) = \sum_{i=1}^{k} x^{(0)}(i) \quad k = 1,2 \dots , n$$

$$x^{(1)} = (x^1(1), x^2(2), \dots x^1(n))$$

Call the resulting new series of cumulatively generated series

$$x^{(r)}(k) = \sum_{i=1}^{k} x^{(r-1)}(i), k = 1,2 \dots, n, r > 1$$

Generally, the differential equation of GM(1,1) is

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$$

Where a is development index, u is endogenous control index. We need to compute a and u. According to the theory, we generate B and Y_n to compute the two indexes.

$$B = \begin{bmatrix} -\frac{1}{2}\left[x^{(1)}(1) + x^{(1)}(2)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(2) + x^{(1)}(3)\right] & 1 \\ \vdots & \\ -\frac{1}{2}\left[x^{(1)}(n) + x^{(1)}(n+1)\right] & 1 \end{bmatrix}$$

$$Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

Then, the answer is:

$$ans = (B^T B)^{-1} B^T Y_n, where\ ans = [a\ u]$$

And the model function is:

$$\widehat{x^{(1)}}(k+1) = \left[\widehat{x^{(0)}}(1) - \frac{u}{a}\right] e^{-ak} + \frac{u}{a}$$

After that, we need to examine the accuracy with two parameters: p and c.
The residual and relative residual are:

$$e(i) = |x^{(0)}(i) - \widehat{x^{(0)}}(i)|, where\ i = 1,2, \dots n \qquad q = \frac{e(i)}{x^{(0)}(i)}$$

The mean of the residual and variance is:

$$\bar{e} = \frac{\sum_1^n e(i)}{n}$$

$$S = \sqrt{\frac{\sum[e(i) - \bar{e}]^2}{n-1}}$$

Lastly, we get p (small residual error probability) and c (variance ratio):

$$c = \frac{S_2}{S_1}$$

$$p = P\{|e(i) - \bar{e}| < 0.6745 S_1\}$$

Table.1. The result of prediction grade

Prediction Grade	p	c
Wonderful	>0.95	<0.35
Qualified	>0.80	<0.45
Reluctant	>0.70	<0.50
Unqualified	<=0.70	>=0.65

According to the report given in Table 1, this is the original coordinate display, see Figure 1.

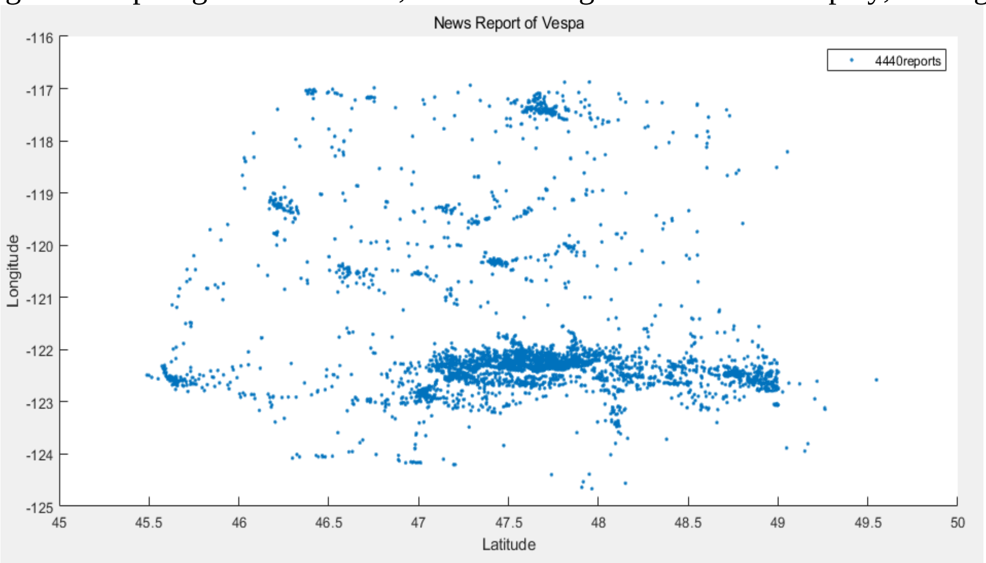

Figure 1. The original coordinate display

According to the meaning of the question and the given data, the blue dots represent all the data in all reports, and the red dots represent the actual witness reports, see Figure 2.

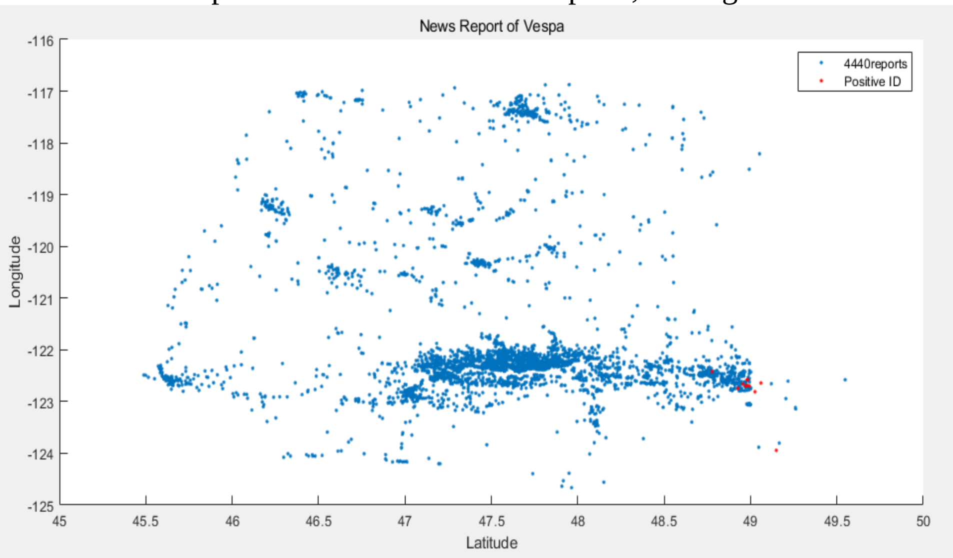

Figure 2. The blue dots and the red dots

Then carry out the prediction and accuracy analysis of its propagation impact. Let's make the first prediction, this prediction does not remove any points. This is the first time the test value of the predicted longitude is compared with the real value. The measurement points are p and c, where c is the quotient of the variance between the true value and the predicted value. The bigger p the better, the smaller c the better. This is the second time to predict the longitude, removing the abnormal point 8, see Figure 3 and Figure 4.

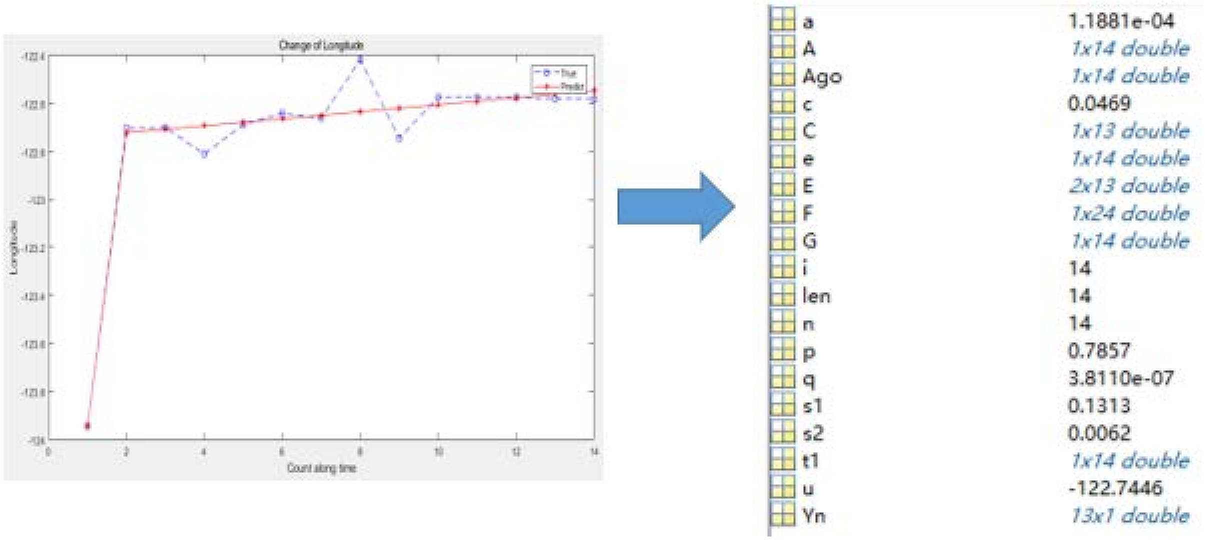

Figure 3. Carry out the prediction and accuracy analysis of its propagation impact

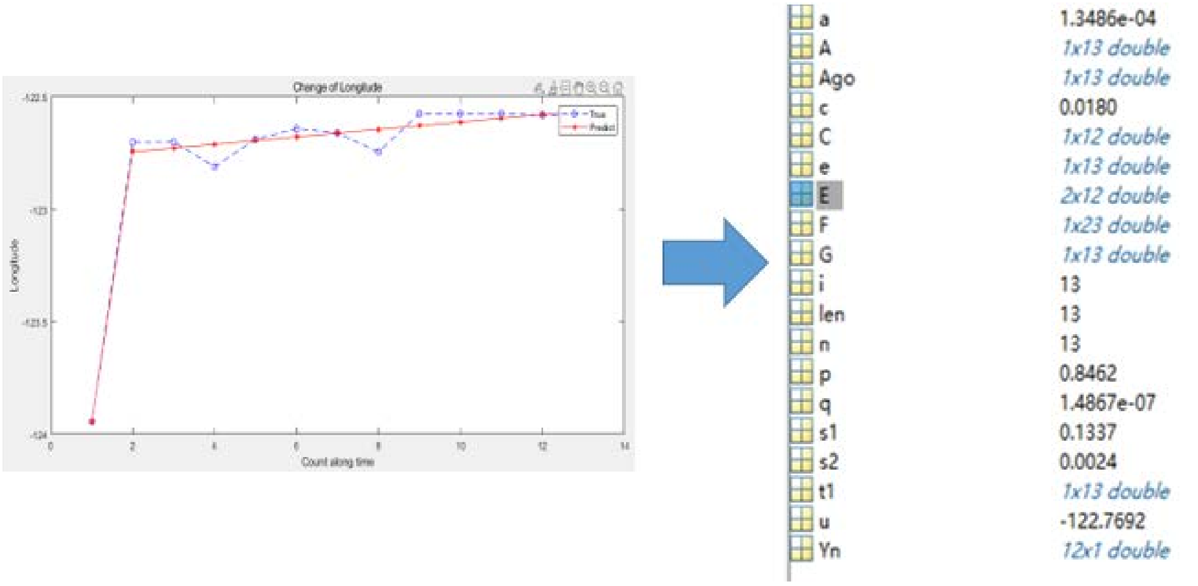

Figure 4. Carry out the prediction and accuracy analysis of its propagation impact

The model will be:

$$\widehat{x^{(1)}}(k+1) = \left[\widehat{x^{(0)}}(1) - \frac{-122.7692}{1.3486e^{-04}}\right] e^{-1.3486e^{-04}k} + \frac{-122.7692}{1.3486e^{-04}}$$

With p=0.84, c=0.018, the prediction is very similar to the reality, with high accuracy.

The latitude prediction method is exactly the same as the longitude. The first time is the full prediction, and the second time is to remove the predicted value of the abnormal point, see Figure 5 and Figure 6.

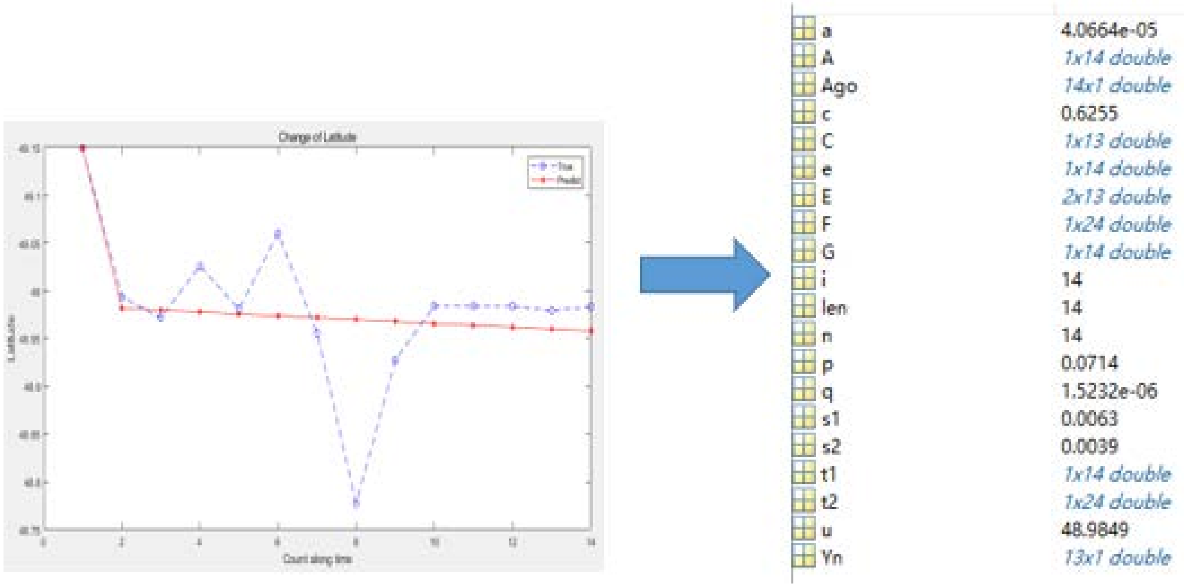

Figure 5. The latitude prediction method is exactly the same as the longitude

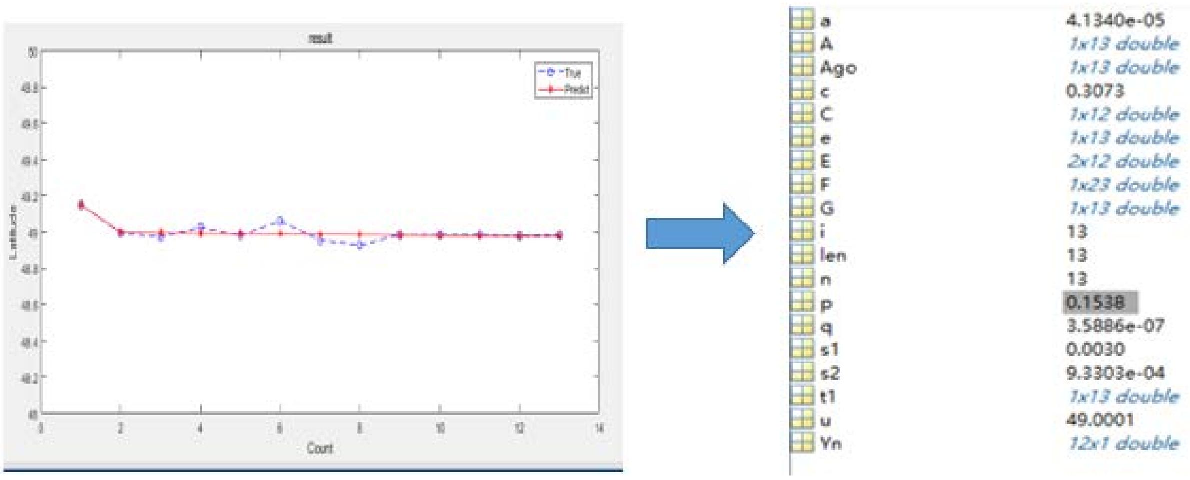

Figure 6. The latitude prediction method is exactly the same as the longitude

$$\widehat{x^{(1)}}(k+1) = \left[\widehat{x^{(0)}}(1) - \frac{49.0001}{4.134e^{-05}}\right] e^{4.134e^{-05}k} + \frac{49.0001}{4.134e^{-05}}$$

With p=0.15, c=0.3, the prediction is not satisfied, but stable.

After the two latitude and longitude predictions are over, the distance is calculated to determine the degree of similarity between the predicted value and the true value, which is presented in an excel sheet. Except for one abnormal point, all predictions meet the requirements (30km), indicating that the spread of this pest can be predicted.

So, start to predict the forecast points for each month from October 2020 to the next ten months. This is the future forecast of longitude and latitude, see Figure 7, Figure 8, Figure 9 and Figure 10.

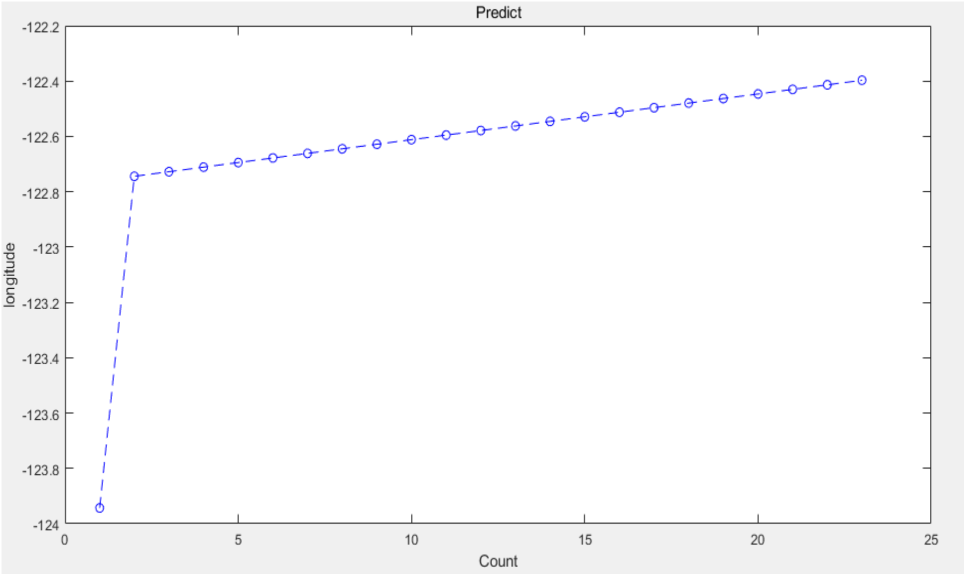

Figure 7. The future forecast of longitude and latitude

	A	B	C	D	E	F
1	Latitude 1	Longitude 1	Latitude 2	Longitude 2	Distance(M)	Distance(KM)
2	49.149394	-123.943134	48.9970986	-122.744172	133643.0848	133.64
3	48.997099	-122.744172	48.9950732	-122.727619	1844.566031	1.84
4	48.995073	-122.727619	48.9930477	-122.711069	1844.314432	1.84
5	48.993048	-122.711069	48.9910224	-122.694521	1844.062837	1.84
6	48.991022	-122.694521	48.9889972	-122.677975	1843.811296	1.84
7	48.988997	-122.677975	48.986972	-122.661432	1843.559835	1.84
8	48.986972	-122.661432	48.984947	-122.644891	1843.308335	1.84
9	48.984947	-122.644891	48.982922	-122.628352	1843.05694	1.84
10	48.982922	-122.628352	48.9808971	-122.611815	1842.805482	1.84
11	48.980897	-122.611815	48.9788723	-122.595281	1842.554165	1.84
12	48.978872	-122.595281	48.9768475	-122.578748	1842.302812	1.84
13	48.976848	-122.578748	48.9748229	-122.562218	1842.051512	1.84
14	48.974823	-122.562218	48.9727983	-122.54569	1841.800276	1.84
15	48.972798	-122.54569	48.9707738	-122.529165	1841.549045	1.84
16	48.970774	-122.529165	48.9687494	-122.512641	1841.29785	1.84
17	48.968749	-122.512641	48.9667251	-122.49612	1841.04667	1.84
18	48.966725	-122.49612	48.9647009	-122.479601	1840.795581	1.84
19	48.964701	-122.479601	48.9626767	-122.463085	1840.544458	1.84
20	48.962677	-122.463085	48.9606527	-122.44657	1840.293449	1.84
21	48.960653	-122.44657	48.9586287	-122.430058	1840.042404	1.84
22	48.958629	-122.430058	48.9566048	-122.413548	1839.791398	1.84
23	48.956605	-122.413548	48.954581	-122.39704	1839.540483	1.84
24	48.954581	-122.39704				

Figure 8. The result of the future forecast of longitude and latitude

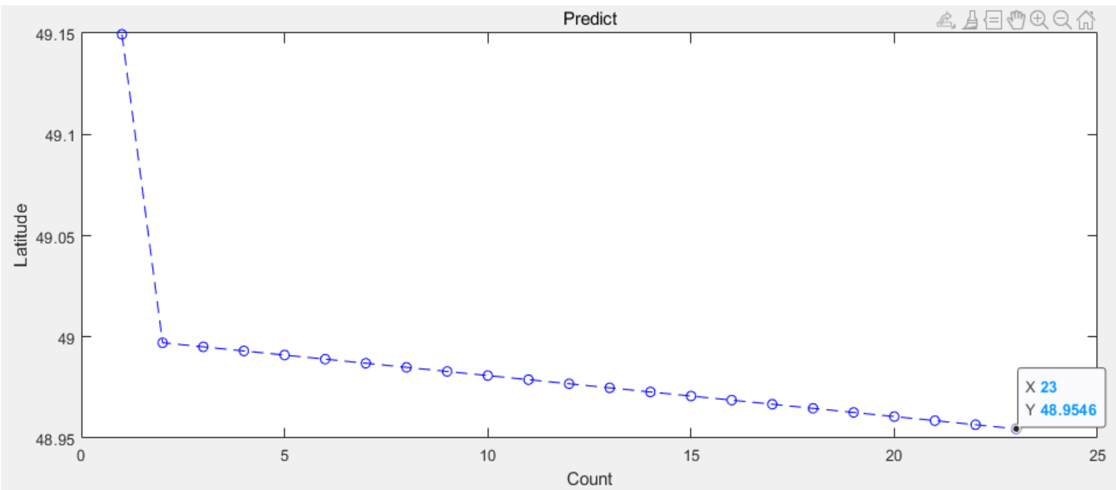

Figure 9. The future forecast of longitude and latitude

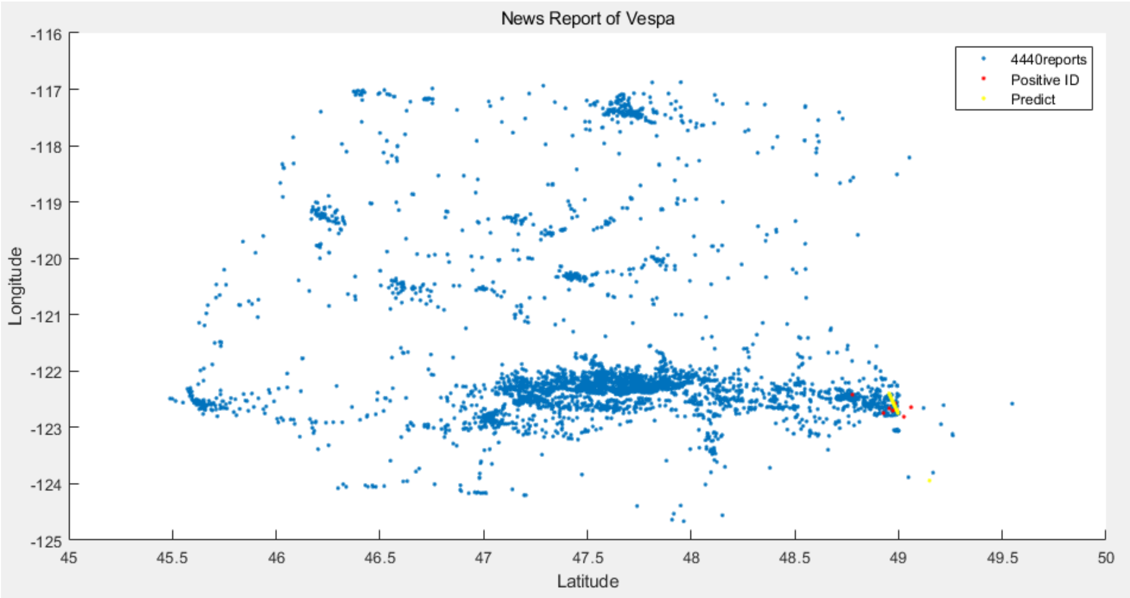

Figure 10. The blue dots and red dots

3. Conclusion and the Update of the Model

For the fourth question, we decided to make this analysis. After providing additional new reports, the new data is recognized by the note model, and then the image is recognized. After the lab comment is obtained, the comment model is trained as new training data. The update time is about three months, because there is no abnormality in the training. Because the bee will slowly disappear by itself, the time interval for collecting data can be extended.

We used Grey Model to solve the first question. Grey Model does not need a large number of data, the effect of its short-term prediction is good, and the operation process is simple.

Grey Model used in question 1 has a poor prediction effect on nonlinear data samples.

In the first question, we used the longitude and latitude of the sighting event to predict its spread using the gray model and calculated its accuracy. The prediction was divided into two times. The second time was based on the first removal of abnormal points. Yes, the p-value and c-value meet the requirements, and it is concluded that the spread of pests can be predicted. At the same time, the distance is also calculated to actually consider the predicted value and the true value.

References

[1] Baike, B. K-NearestNeighbor. Retrieved February 6,2021, from https://baike.baidu.com/item/%E9% 82%B B %E8%BF%9 1%E 7%AE%97%E6%B3%95/1151153?fr=aladdin

[2] Johanna Pingel, MathWorks Avinash Nehemiah, MathWorks (Mar 7, 2017). Object Recognition: Deep Learning and Machine Learning for Computer Vision. Retrieved February 6, 2021, from https://ww2.mathworks.cn/videos/object-recognition-deep-learning-and-machine-learning-for-computer-vision-121144.html?s_iid=doc_rw_VP_footer

[3] Sudo-Wang (June 5, 2018). Gray Model. Retrieved February 6, 2021, from https://blog.csdn.net/wuli_de ar_wang/arti cle/details /80587650