Scientific Journal of Intelligent Systems Research Volume 4 Issue 6, 2022
ISSN: 2664-9640

Extreme weather time and space research and risk assessment and prediction

Junting Dong*
China University of Petroleum - Beijing At Karamay, School of Business Administration
Karamay City, Xinjiang Uygur Autonomous Region, 834000, China

Abstract

This paper studies the probability calculation model of extreme events, completes the improvement of the old model, and analyzes the law of space-time change and predicts it by using extreme weather data. Finally, the losses caused by extreme weather are analyzed, and the losses are effectively evaluated by establishing a multi-regression model. We modify the original probability model, transform a static probability model into a dynamic model that changes over time, and establish dataset for descriptive statistics. Calculating the mean, variance, and degree of dispersion of the dataset, we compare who is least discrete under different statistical rules, and we can consider the event to be a once-in-a-decade event. The random selection cz_ fips for this model is 1, state_ fips 0, event_ type Hail. The model is used to calculate that the probability of the event occurring is 4.65%. It's a once-in-a-30-year event. Then, we study the laws of space-time in which extreme events occur. For the time law to find the effect of probability change on the time occurrence period, the scatterplot is drawn to find the fitted equation. It is found that the probability of time increases over time and the cycle decreases gradually. For the law of space, count the number of times different places occur in a city, calculate the relative probability of occurrence, and use space-time interpolation algorithm to predict the occurrence of events in space .different kinds of events are treated as $D_1, D_2, D_3 \dots D_n$ and the corresponding event period is $T_1, T_2, T_3 \dots T_n$.When the probability of an extreme event once in a century or once in 50 years does not change, the variety of extreme events that can occur in a given area increases over time.

Keywords

Multiple linear regression; The Laida Code; Spatial interpolation model

1. Introduction

1.1. Problem Background

In July, a rare storm and flooding in Western European countries such as Germany, the Netherlands and Belgium plunged Europe, which had just recovered from the outbreak, into disaster. The death toll in Europe in a major flood is 200, with hundreds more still unaccounted for and little hope. In 2021, the biggest worry is no longer the impact of the new crown on the world, but the global succession of extreme weather events, storms, extreme heat and cold, floods and landslides, these extreme weather events in many experts believe are difficult for years, but now the frequency of occurrence has to let the world start to pay more attention to the weather.

Extreme weather events are weather and climate events that occur in a region or place with a low frequency or considerable intensity that have a significant impact on human society over a period of time. But in recent years it has been found that the occurrence of these list of extreme events is not isolated and accidental, as people over-consumption of resources to destroy the

ISSN: 2664-9640

environment, the global environment is deteriorating, global warming is becoming more serious, which may make extreme weather occur more frequently.

At the same time, scientists believe that extreme weather events may occur by accident, the probability of occurrence should be very small, look at the current global climate conditions we find that the occurrence of these events are interrelated, in a certain time frame, if an extreme weather event occurred in one area, often affect the surrounding area and other areas are more prone to extreme weather. With the frequent occurrence of extreme weather events, we should master the scientific prediction and evaluation method, can rationally deal with the storm, high temperature and cold, heavy rainfall and flooding under the trend of climate change and other disasters affected us.

2. Problem analysis

We build a model that can assess the probability of weather events based on the data given by the topic. First, we should understand that the assessment data weather event definition is the probability of extreme weather occurring within a unit of time. In fact, there is an initial model to judge the probability of events, but as the title says, the original method of judgment itself has certain defects, can not be a good assessment of the probability of events. In this regard, we can modify the original model to convert the static evaluation model into a dynamic model. The data of a region is selected for stability testing and the probability of the event occurring is obtained.

For question to explore the law of space-time change, we can divide into time and space to study, first of all, in the law of time, the occurrence of extreme events contains nothing but the probability and cycle of the change. The trend of the cycle can be judged according to the study of the probability of question one. With regard to space exploration, this requires us to study the location and probability of extreme events, find their correlations and predict them.

And we analyzed the probability of extreme events and the laws of space-time, and now think about what makes extreme weather more likely. In fact, in order to calculate the probability of the occurrence of the whole extreme event, we must overlay the probability of each event, and we can determine the cause of the specific frequency by establishing the formula of the total event probability calculation.

3. Model establishment and solution

3.1. The establishment of an evaluation model for dynamic events

This article has explained the events, which are called 30 years once, 50 years once, or once in a hundred years, may taste how many years will occur once. But that doesn't mean it happens every 30 years, every 50 years, or every 100 years. Meteorologists understand that once in a century is an event of equal or greater size that takes place on average once every 100 years. In mathematics, we can fully understand as an assessment of the probability of occurrence. For the probability of an event occurring, we usually use the following formula:

$$P = \frac{n}{t} \times 100\%$$

Thereinto, t represents the selected time range, and n represents the number of times events occurred during that time. In the probability expression above, we can see that this is a static model, the P value may always be fixed in a certain time range, such a probability method is to have a relatively large error and inaccuracy. To do this, we add dynamic changes to the original model, making P a series of numbers that change over time periods, which improves accuracy.

According to the analysis of the original model, we should increase the time variability to improve the accuracy of the probability solution results, the dynamic event evaluation model is as follows:

$$P_{(t)} = \frac{n_{(t)}}{t_{1(t)} - t_{2(t)}} \times 100\%$$

Thereinto, $t_{1(t)}$ represents the year in which the statistics were closed, $t_{2(t)}$ represents the year in which the statistics began,and $T = t_{1(t)} - t_{2(t)}$ is a time difference constant, n represents the number of times an event occurred during $t_{2(t)}$ and $t_{1(t)}$ consecutive periods. With the improved model, we can calculate the number of columns whose probability of an event will occur over time. Over time, we can get a number of columns $P_0, P_1, P_2, P_3 \ldots P_n$. So we construct the dataset $P = \{P_0, P_1, P_2, P_3 \ldots P_n\}$, reuse the data in the dataset to describe the statistics.

Descriptive statistics of events occur, which cannot be avoided in dynamic data, but they can have a greater impact on the results in the model of probability calculation. First of all, the ractracity data in the data set is rejected by the Laida criterion, and the new data set P^* is obtained. The data in the new dataset P^* is then calculated as mean and variance, as follows:

$$\overline{P} = \sum_{i=1}^{i=h} P^*_{(i)} / n$$

$$S^2 = \sum_{i=1}^{i=h} (P^*_{(i)} - \overline{P})^2 / n$$

Where $\overline{P}$ is the average of the data, it can be used as an evaluation value for the dataset, which is the probability of the model solving. S^2 can represent the stability of the evaluation value $\overline{P}$. Here, in order to further judge the effect of solving the evaluation value $\overline{P}$, we introduce deviation degree u to judge whether the evaluation value is in the deviation range, the formula is:

$$\min u = \frac{\overline{P} - P}{P}$$

$\overline{P}$ is just the evaluation probability value we calculatedǎ and in the question we have to judge whether the data event belongs to one in 30, one in 50 or one in 100 years. So we introduce event classification variables again. Find out the group with the smallest R value of the classification variable $R = S^2 \times u$,and you can determine whether the event is a ten-year-old event, a thirty-year-old event, a fifty-year event or a hundred-year-old event.

The model has been built above, and here we select a set of extreme weather events as a test of the model. The title gives a time span of up to 70 years for storm events in various parts of the United States from 1950 to 2021, and the data do not allow for the determination of once-in-a-century events. But we can judge whether the event is a ten-year-old, thirty-year or fifty-year event, using the above three groups as a classification, we began to verify the model.

1.The choice of data

There are three categories of data files provided by the title: details of the incident, the location of the incident and the number of deaths. Here we select the event detail file and filter the data. There is too much information in these files, and we only need to determine which years of extreme weather have occurred in a given place, and only four headings in the table are useful to us: (1) cz_fips: a unique number assigned to the county by the National Institute of Standards and Technology (NIST) or NWS Forecast Area Number. (2) state_fips: unique number assigned to the state by the National Institute of Standards and Technology (NIST) (state federal information processing standard). (3) Year: The year in which the event occurred. (4) event_type: event categories such as hail, thunderstorms, snow, ice (spelled out); Not abbreviated).

To determine that an event occurs in the same region, simply find the cz_fips, state_fips, and event_type the same event and record the year in which it occurred. We randomly select cz_fips here as 1, state_fips as 0, event_type as Hail. The years in which these incidents occurred were recorded in 1954, 1972, 1999 and 2013. Based on the calculation of the new model, the dataset P is obtained as shown in Tables 1 and 2:

Table 1. Probability P is counted every 30 years and every 50 years

statistics end	30 years	30 years	statistics end	30 years	30 years
1979	6.67%	/	2000	6.67%	6.00%
1980	6.67%	/	2001	6.67%	6.00%
1981	6.67%	/	2002	3.33%	6.00%
1982	6.67%	/	2003	3.33%	6.00%
1983	6.67%	/	2004	3.33%	4.00%
1984	3.33%	/	2005	3.33%	4.00%
1985	3.33%	/	2006	3.33%	4.00%
1986	3.33%	/	2007	3.33%	4.00%
1987	3.33%	/	2008	3.33%	4.00%
1988	3.33%	/	2009	3.33%	4.00%
1989	3.33%	/	2010	3.33%	4.00%
1990	3.33%	/	2011	3.33%	4.00%
1991	3.33%	/	2012	3.33%	4.00%
1992	3.33%	/	2013	6.67%	6.00%
1993	3.33%	/	2014	6.67%	6.00%
1994	3.33%	/	2015	6.67%	6.00%
1995	3.33%	/	2016	6.67%	6.00%
1996	3.33%	/	2017	6.67%	6.00%
1997	3.33%	/	2018	6.67%	6.00%
1998	3.33%	/	2019	6.67%	6.00%
1999	6.67%	/	2020	6.67%	6.00%
			2021	6.67%	6.00%

Table 2. Probability P is counted every 10 years

statistics end	30 years	30 years	statistics end	30 years	30 years
1959	10.00%	1980	10.00%	2001	10.00%
1960	10.00%	1981	10.00%	2002	10.00%
1961	10.00%	1982	0.00%	2003	10.00%
1962	10.00%	1983	0.00%	2004	10.00%

ISSN: 2664-9640

1963	10.00%	1984	0.00%	2005	10.00%
1964	0.00%	1985	0.00%	2006	10.00%
1965	0.00%	1986	0.00%	2007	10.00%
1966	0.00%	1987	0.00%	2008	10.00%
1967	0.00%	1988	0.00%	2009	0.00%
1968	0.00%	1989	0.00%	2010	0.00%
1969	0.00%	1990	0.00%	2011	0.00%
1970	0.00%	1991	0.00%	2012	0.00%
1971	0.00%	1992	0.00%	2013	10.00%
1972	10.00%	1993	0.00%	2014	10.00%
1973	10.00%	1994	0.00%	2015	10.00%
1974	10.00%	1995	0.00%	2016	10.00%
1975	10.00%	1996	0.00%	2017	10.00%
1976	10.00%	1997	0.00%	2018	10.00%
1977	10.00%	1998	0.00%	2019	10.00%
1978	10.00%	1999	10.00%	2020	10.00%
1979	10.00%	2000	10.00%	2021	10.00%

2.Exclusion of outliers

In order to eliminate the influence of outliers on our judgment, using the Laida criterion to judge and exclude outliers of the data, after python outlier detection, we found that the data did not have outliers, so we will retain the data, that is $P = P^*$.

3. The calculation of the mean and variance

Table 3. The mean and variance of probability P are obtained by different statistical Methods

Statistical method	1 statistic in 10 years	1 statistic in 30 years	1 statistic in 50 years
average value	5.40%	4.65%	5.18%
variance	0.00248425	0.000265609	0.00009669

4. The calculation of minimum deviation u

ISSN: 2664-9640

Using the results of Table 3 to obtain the degree of deviation under each statistical method, the results are shown in the following table:

Table 4. The degree of deviation of mean P under different statistical methods

Statistical method	1 statistic in 10 years	1 statistic in 30 years	1 statistic in 50 years
Degree of deviation	0.851851	0.2753623	0.6124031

As can be seen from the results of the table above, the 30-year statistical deviation is minimal, so in counties with a cz_fips of 1,state_fips to 0, the extreme weather event event_type Hail is a once-in-30-year event.

3.2. A spatiotemporal forecast model for extreme weather

In question one, we get the data set P by using the improved probability calculation model, at which point we can use EXCEL to make scatter plots of probability data sets and time, and make trend lines to complete the linear fitting equation. The resulting linear regression equation is: $Y = 0.0032x + 0.037$

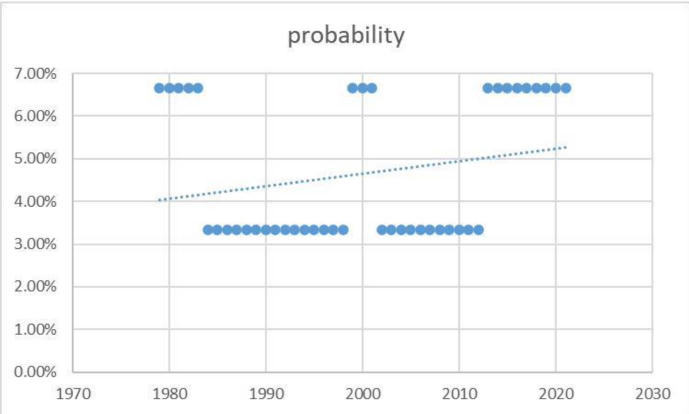

Figure 1: Scatter plot of probability and time of extreme weather occurrence

According to the trend of the figure above, the a of the regression equation is greater than 0, and it can be found that the probability increases gradually with the change of time. For period T, the cycle of extreme events is gradually smaller when $a > 0$.

The prediction of the laws of space is also to find out how the probability of extreme weather occurring in different regions (coordinates) changes. To do this, we first collect the probability of extreme weather occurring in different regions at different times. The data given by the title includes data on regional coordinates. Here we select the cz_fips for the location and probability of a county in 1 city over a period of 70 years, the results are as follows:

Table 5. Number of county towns and extreme weather events

County no	The number of occurrences	County no	The number of occurrences	County no	The number of occurrences
0	4	11	17	21	21
1	18	12	21	22	15
2	6	13	25	23	27
3	4	14	3	24	18
4	21	15	21	25	4
5	7	16	12	26	16

6	21	17	18	27	27
7	9	18	3	28	8
8	5	19	16	29	9
9	29	20	2	30	27
10	15	21	18	31	4

Using the information in the table above, we can know the relationship between location and relative probability of occurrence, and we use spatial interpolation to predict the probability of space. The probability prediction for city 1 is shown in the figure below, in which we can see that the probability of extreme weather in urban center is relatively large.

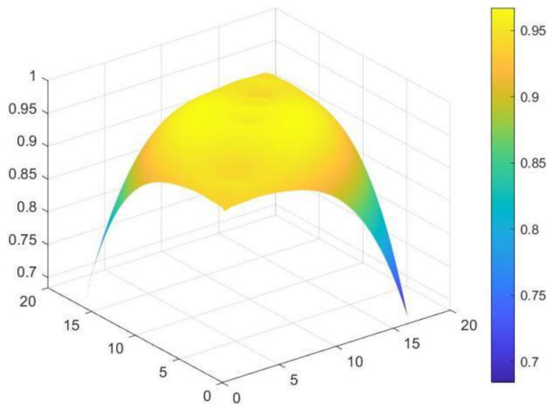

Figure 2: Spatial prediction of extreme weather events in City 1

3.3. Model establishment and solution of problem three

3.3.1 Model establishment

According to the data given by the title, we can know how many kinds and numbers of extreme weather can occur in different regions. Suppose there are n types of extreme weather that can occur in an area, different kinds of events are considered $D_1, D_2, D_3 \ldots D_n$ and the corresponding events occur in a period of $T_1, T_2, T_3 \ldots T_n$. Now we can assume that the type of extreme weather will not change in a short period of time. Here we can use periods to indicate the probability of extreme weather events, which are as follows:

$$P_i = \frac{1}{T_i}$$

Then we set G to the probability of all extreme weather occurring in this area over a period of time(the time span cannot be too large), in the expression:

$$G_n = \frac{T(\frac{1}{T_1} + \frac{1}{T_2} + \frac{1}{T_3} + \cdots + \frac{1}{T_n})}{T} = \sum_{i=1}^{n} \frac{1}{T_i}$$

In the upper formulaā there is always $P_i \le G_n$ or the occurrence of P_i for an extreme event. So when the probability of an extreme event once in a hundred or 50 years does not change, the variety of extreme events that may occur in a region over time increases, which leads to why the once-in-a-century time has occurred frequently in recent years.

3.3.2 The test and analysis of the model

In the above, we find that the frequent occurrence of extreme events is due to an increase in the variety of events in a region based on a simple formula for calculating the total probability of extreme events. Based on the above analysis, we select a location as the subject of the study. By dividing the time periods, observe whether the variety of extreme events is increased over time, resulting in an increase in the total probability.

In this question, we still choose the first question to choose the county with a cz_fips of 1,state_fips of 0, from 1950 to every 10 years for a time period to divide, using EXCEL screening function, statistics in each time period may occur extreme weather time category, the results are shown in the table below:

Table 6. Number of types of extreme weather over a different time period in a region

The time period	The number of extreme weather categories	The time period	The number of extreme weather categories
1950-1959	0	1990-1999	7
1960-1969	2	2000-2009	7
1970-1979	4	2010-2021	10
1980-1989	5		

According to the statistics in the table above, we found that the types of extreme weather events in the region increased between 1950 and 2021, resulting in an increase in the G value. This shows that when the probability of some kind of extreme weather does not change, the frequency of total extreme weather time occurs frequently, mainly due to the increase of some unknown extreme weather.

4. Conclusion

This model can use data to accurately predict extreme weather events. Multiple regression analysis is the most basic and important method in data analysis, and the purpose of predicting Y through X is achieved by studying the correlation between argument X and dependent variable Y. The vast majority of data analysis problems can be solved using the idea of regression. But multiple regression models require strict assumptions. Outliers need to be handled, are sensitive to outliers, and are poor for input data.

References

[1]Guo Weihui, Feng Xuefei, Feng Xuejun, Liu Pengfei, Yu Fangjian. T-YearLy Extreme Weather Event Prediction Methods (J.J.Gilling

[2]Agriculture, 2016(06): 119-120.DOI: 10.14025/j.cnki.jlny.2016.06.046.

[3]Liu Zi. Even more ferocious than the new crown is the climate contemporary workers (version C), 2021 (05): 37-38.

[4]Wang Bin. Extreme rainfall is a wake-up call for climate change. Science News, 2021, 23 (04): 47-49.

[5]Wu Daming. Analysis of global extreme weather events and impacts in recent years. China Disaster Reduction, 2021 (15): 22-25.

[6]The editorial department of this magazine. Focus on Extreme Weather And Weather Events (J).China Disaster Reduction, 2021 (15): 8-9.