Should Have Seen It Coming: Is Henan Flooding a Karma?

Environmental Analytics Project – Option 2 Natural Hazards.

Candidate id: 013349

2021

Table of Contents

[Introduction 2](#_Toc79486154)

[Data 3](#_Toc79486155)

[Temperature 3](#_Toc79486156)

[*I. Is there any trace of global warming in Zhengzhou?* 3](#_Toc79486157)

[Hydrology 5](#_Toc79486158)

[*II. Has the Risks of Extreme Climate Event Changed?* 5](#_Toc79486159)

[Conclusion: 11](#_Toc79486160)

[Reference 13](#_Toc79486161)

Should Have Seen It Coming: Is Henan Flooding a Karma?

Environmental Analytics Project – Option 2 Natural Hazards.

# Introduction

“When the Yellow River is at peace, the nation is at peace”. The slogan emblazoned on the Sanmenxia, the first dam on the river, built in 1950s. The battle (combat) against flood can be traced as early as five thousand years ago. The hero of story, Da Yu, tamed the long and wild rivers. His status stands tell and grand and his tale /lore told by folks.

Today, the Yellow River is furious. Upstream the longest river is the Capital and central of China’s Henan province, Zhengzhou. The city, once proud for most modernised subway station, now turned into underground rivers (BBC, 2021). CNN (Gan & Wang, 2021) reported a passenger, with flood water reached chest hight, from contact rescuers to making death arrangement. At least 2.4 million people were displaced and 71 deaths, those includes 14 of those who were drowned in Zhengzhou subway (reported by [Los Angeles Times](https://www.latimes.com/world-nation/story/2021-07-28/china-henan-floods-control),2021).

Urban flood is global concerns. From Australia to German, regulator around the globe is concerned with flooding and natural disaster alike. In financial sector, insurers may not prepare adequate capital requirement in time (Actuaries, 2021).

IPCC (Field et al., 2021) predicts that global temperature change awaits higher risks of extreme climate events. While Chinese State media refer to Henan flood an extreme rare event, figuratively “once in a thousand years” (Gan & Wang, 2021), western journal networks frequently blame urbanisation and global warming (BBC, 2021; Guardian, 2021; Gan & Wang, 2021). Indeed Zhengzhou, the capital and largest city of Henan province, has been expansion to fit the growing population and demand. This is marked by opening of its first metro line in 2014 ([Smith, 2014](https://universityofexeteruk-my.sharepoint.com/personal/fl339_exeter_ac_uk/Documents/Assignment/where%20the%20vulnerable%20congregate)) with air conditioning and Wi-Fi.

Is Zhengzhou flood really a karma of urbanisation for Chinese officials? Or has Zhengzhou finally fell victims of global warming? This report attempts to access flooding risks of Zhengzhou city and explore the linkages with concerns of raising global temperature.

The first part of the report seeks the trace of global warming in Zhengzhou. This could be achieved by monitor temperature change over the years and apply linear regression.

In the second part exam if risks of extreme weather events have changed via parametrising probability density function for precipitation volume. The risk of rain exceeds certain level can then be estimated via exponential distribution.

Finally, had flooding maps of Zhengzhou been published, regional risks of flood in Zhengzhou could have been assessed. So instead, this report propose methodology that could be applied when required data are collected.

The evidence of global warming is presented in Zhengzhou using some models. However, the estimation for extreme weather event, as will be demonstrated in the second section, behaves abnormally. This report is unable to conclude if risk of extreme climate event is increasing in Zhengzhou.

# Data

NOAA supplies historical to precent climate data from around the world. Data used for analysis were accessed via NOAA api (2021a). Data related to Zhengzhou are collected by Zhengzhou weather station. The abbreviation used in the data can be found at.

Zhengzhou is notoriously known for its high temperature, one of the five “hotpot” cities of China. Big city with heavy urbanisation tends to suffer from what is called “urban heat island” effect (Howard). Where excessive energy was produced, greenhouse gas and concrete construct forged to trap heat from losing. As a result, urban area has higher temperature than that of surroundings.

There is only one weather station in Zhengzhou which is in the centre of city, the centre of the “heat island”. While having only one station is convenient, it is less informative regarding monitor different part of city. London, Sydney, have more than 5 weather stations. Likely the station would record higher temperature.

A picture containing diagram

Description automatically generated

# Temperature

### *I. Is there any trace of global warming in Zhengzhou?*

**Exploration: Daily Temperature Change Has Shinked**

Chart

Description automatically generatedChart, scatter chart

Description automatically generated Chart, histogram

Description automatically generated

The change of maximum temperature is not obvious by just looking at the heatmap. However, it appears the daily temperature change (maximum day temperature minus minimum day temperature) is shrinking. This likely to be the case of “urban island effect”. From 1960s to 2010 Zhengzhou has undergone large scale construction and urbanisation, the population density has grown and the urban island effect, thus, became more prominent.

**Methodology: Different Model shows different results**

A period from 1970 to 2000 (31 years) were selected to feed three models. Maximum temperatures were selected. Three base models were used to reveal global warming: simple linear model (LM), time series model (TS), linear mixed effect model (LME).

Special cares were taken when feed data to LM and TS. For LM, daily maximum temperature was aggregated by computing annual mean. For TS, monthly mean of daily max temperature was computed for each 12 month of each 31 years.

As will be demonstrated below, each model yields different results which make it difficult to conclude if global warming manifests in Zhengzhou.

**Result and Analysis: Different model different conclusion**

In simple linear model (LM), day to day variance and seasonality were ignored by simply compute annual average. Controverse to common sense, the model actual estimated a negative result of - 0.19 per year at 89% confidence interval, that is, hardly significant. Further the model has very poor fit. An adjusted R square of only 0.014. That is, only able to explain around 1.4% variation.

A picture containing text, receipt, screenshot

Description automatically generated

TS model is performed via Mitchell O'Hara-Wild and Rob J Hyndman’s (2021) R package “*forecast”*. The model showed a good fit with more than 95% of variation explained (adjusted R-Square = 0.9678, p = 2.2 e-16). Standard error of residual is around 17, that is, around 66% of the time prediction would not exceed 1 °C degree variation. The model has demonstrated good prediction capability too. This is validated by letting the model project for the next 10 years. The projection line covers actual temperature in years 2001 to 2010. However, the model does not capture effect of global warming. Estimation for trendline is negative and not significant (p = 0.063).

Chart, line chart, histogram

Description automatically generatedTable

Description automatically generated

Chart

Description automatically generatedText

Description automatically generated

The LME method is recommended by Michael Crawley (2013), the author of *the R Book*. Crawley (2013, p.798) argues that the linear methods contain too much *pseudoreplication* hence trend should be tested by compare two models with or without fitting trend lines. Mixed effect model concerns what is called “fixed effect” and “random effect” (Eisenhart, 1947). For this model, the Fixed effect is assuming time and maximum temperature is seasonal cycle equation; the random effect, as instructed by Crawley (2013, p.798). Over the model showed a better fit the linear model but much poor fit than TS model by comparing residual. Comparing two models (one assumes global warming one do not), the result is significant (p = 0.02, at 98 % confidence. A positive effect where estimated (7.225e-04). That means, Zhengzhou is warming about 0.026 °C each year (7.225 \* 365 \* 104 \* 0.1 = 0.263), twice the rate of NASA’s () global warming rate.

The trace of global warming is only obvious when applying linear model and **trigonometric** mixed effect model. When applying r’s building time series function, trend is not statistically significant. The latter (TS model) have a higher adjusted r square and lower residual value compares to previous two model, but it also contains much more parameters (intercept parameter, plus 12 parameters, one for each month). In this sense, results from linear model and mixed model were more trustworthy. That concludes

# Hydrology

### *II. Has the Risks of Extreme Climate Event Changed?*

**Exploration: Extreme Rainfall Has Happened 4 Decades Ago**

Graphical user interface, chart

Description automatically generated

A glance at precipitation of Zhengzhou throughout the years from 1960 to 2020, it appears that extreme precipitation happens more frequently than has been claimed. The highest value traces back to 1978, and most recently in 2008 (South China flooding).

Analysis for precipitation data were split into two parts. First, the frequency of raining days. It evaluates how raining frequency has evolved through the years via simple linear regression. Second, the volume of precipitation and risks of extreme value. This is achieved by parametrize probability density function of precipitation distribution curves.

This report chooses not to model with **trigonometric** functions as did with temperatures. Of course, one may be attempted and can always coerce a trigonometric model by aggregating daily precipitation volume into monthly one. This approach ignores day to day variation of rain. It does not hurt to have a few drizzling months; it does hurts to have a few high-frequency, high-volume storms that exceeds a city’s capacity to discharge water. For example, in 2018 the Zhengzhou has a lot more raining days than that of 2021, but there is not recorded flood in 2018. Arguably, as far as the concerns of urban flood risks assessment, monthly precipitation is not as helpful daily precipitation.

Chart, scatter chart

Description automatically generatedChart, histogram

Description automatically generated

Majority of the days in Zhengzhou are dry days and Zhengzhou is becoming drier. Frequency of raining days has mean of 22.7% and standard deviation of 3.21%. That is in a year, 80 days are averages are raining days. The frequency of rain appears to be normally distributed which satisfies assumption of linear regression. Estimation result shows that the frequency of rain is decreasing (slope parameter is negative value, model significant at p = 0.005523). Every ten year, raining frequency of raining is decreasing by 0.7% per year. That is, in ten year- time, Zhengzhou will experience 2 days fewer raining days but with around 11 days variation.

**Estimate Risks: Estimate Probability Distribution Density Function**

The distribution of precipitation volume appears to be exponentially distributed. Exponential distribution has the property of mean equals standard deviation. However, the calculation shows that mean of precipitation volume is not equal to standard deviation (mean = 0.2217114; sd = 0.03215555). However, standard deviation and mean appears to be linearly related. Further, this relation is highly significant, verified by apply linear model (p-value = 1.42e – 11). Hence this report attempts to fit probability density curve from scratch. The data used to fit probability density curve can be retrieved from same algorithm that was used to compute histogram. Overall, 1000 breaking data were retrieved from 5025 entry of daily precipitation volume from year 1960 to 2010 whenever raining volume is not 0; the density value can be approximated by density.

Chart, histogram

Description automatically generatedChart, scatter chart

Description automatically generatedText, letter

Description automatically generated

Two parameter Exponential Decay function (Ritz & OECD, 2006) was used to fit probability density function. Initially, poor result was yield. As shown in the diagram, actual value (shown in dots) does not align with fitted value (shown in blue line) at all. This problem is solved by accidently apply logarithm scale to precipitation volume which is the x axis in the chart. As can be seen model appears to be a much better fit to the model.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Two parameter exponential decay function were given as:

*(1)*

Probability density function of exponential distribution is:

*(2)*

Let , set new parameter :

*(3)*

≤

Replace equation *(3)* into equation *(1)*:

*(4)*

Here by equation (4) have resemblance to equation *(2)* but is equation *(1)*. Now one can use parameter as if , but instead of use x, use *k \* x.* Because the logarithm of x fits into function (1) one. *k\*log(x)* is exponential distribution at . That is:

*(5)*

This allows estimate probability distribution via exponential distribution can use exponential distribution to estimate probability of extreme events but, instead, through . Calculation of are given in *(3).*

One statistical common sense is shape of histogram will be affected if different bin were taking. So just in case, this report has considered using different bin numbers. In common sense, bin number should not exceed total number of sampled. It turns out that there are procedures for choosing bin numbers. The simplest one is simply square root total amount of data. Three methods were considered in addition to square-foot method: Sturges’ formula (1926), Rice’s Rule, Duane’s formula. Results are in tables below. Note that year indicate beginning year of ten years interval.

Table

Description automatically generated

*\*Length indicates how many rain days in a ten-year interval.*

A problem with Sturges (1926) is it is originally designed for data that are normally distributed. The only method that was designed for abnormally distributed data is Donna’s (1976), which takes consideration of the Skewness. Those choses result in dramatically different results, as demonstrated in below chart:

Chart, line chart

Description automatically generated

Above chart illustrate the probability of rain exceeds 10 cm (prcp > 1000). The process used to calculate probability is the same as previously introduced in *formula (5).* The results are most dramatic when took Strugar’s approach (using break = 11), and less when using Rice or Donnas (using break = 30).

**What has gone wrong?**

It is difficult to trust results like this. To solve this problem, this report has run following experiment that using different break sizes every 10 breaks, from 20 to 80 breaks and see how this affect the results:

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

This experiment reveals models starts to behave irregularly when applies different levels of breaks. It appears that greater the number of bins (breaks) took from histogram, the smaller the estimation for probabilities. In exponential function, parameter d controls the height of the curve while parameter e controls the skewness of curve. From the above chart (top left) parameter e is decreasing, while parameter d changes randomly. This indicate the curve of is less skewed, which explain why the higher the break level, the lower the estimated probability.

An explanation for this may be because when higher break level (number of bin) was selected, the width of bin is lower. In a histogram, while height of that bin is maintained, the position of bin (indicated by x axis) shifts to left as width of bin decrease, hence density curve became less screwed.

So far, a larger bins number is negatively associated span of density curve and will result in a smaller estimation of probability. However, it is not clear if this effect will reach a climax or rock bottom. In 1960s, value changes drastically between bin size 40 and 50, while in 1970s, the value changes drastically between 60 and 70. There is no clear pattern as how this is happening. It is difficult to determine the right bin numbers without being backed by in-depth statistical research. Since Donna’s (1976) took consideration of the skewness, preferably, only Donna’s method (bin size = 19) is kept for the next step of analysis.

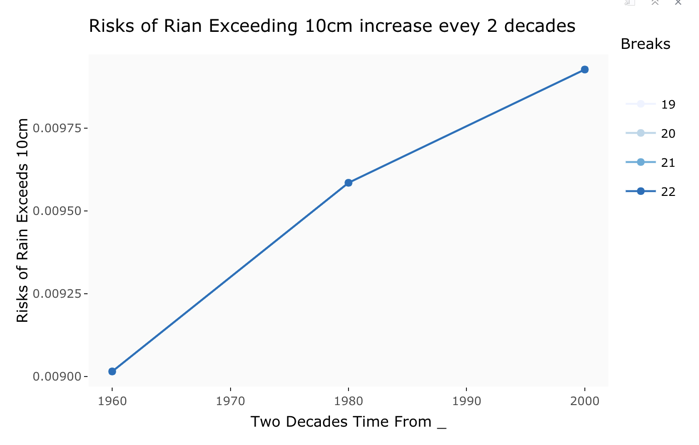
Even correct bin size was selected, the results is still hard to interpreted. The probability of extreme weather fluctuates ups and downs. To solve this, the window of analysis has been increased from 10 years to 20 years and estimation of parameter are shown as below:

Chart, scatter chart

Description automatically generated

In twenty years, window, the shape of curve is becoming flatter. This is indicated by parameter e became larger and parameter d became smaller. As the result, the probability is of extreme precipitation value is increasing in a more linear way. However, not so when a higher level of breaks was selected.

Chart, line chart

Description automatically generated

To put things into perspective, the probability of a single rain level exceeds 10 cm is about 0.9%; the probability of a single raining day exceeds 15cm is 0.75% between 1980 to 2020. At a linear rate of 0.05% increasing. Note, because of limited data, no model can be fitted yet. At this rate, one can expect at least one rain days will exceeds 15cm in Zhengzhou every 125(1/ (0.75% + 0.005%)) rainy or snowy days. Given that the about 20% of days in Zhengzhou are wet days, every 2 years there would be storm where precipitation level exceeds 15 cm. This is highly controverses to intuition and common sense. In 62 years recorded data, there is only 4 recorded precipitation volume that exceeds 15cm. Graphical user interface, chart

Description automatically generatedGraphical user interface, application

Description automatically generated

This suggests the model is far from faithfully represent true risks of extreme precipitation. This may be because incorrect level of breaks is selected. One may be harsh to conclude that perhaps true break level is much higher than Donna’s. However, this is not the case. As shown in below chart, the trendline became distorted again when a higher value Breaks is selected.

Chart, line chart

Description automatically generated

**What else can be done?**

In terms of predicting storm, one approach is called Neyman-Scott Rectangular Pulses (NSRP). This is a more complicated approach that requires estimation of several parameters: The original process starts with knowing the occurrence each storm and ends with a total intensity at a given time. However, this requires specialist knowledge in hydrology and potentially consumes more detailed data. For example, know how to identify a storm intensity of rain in an hourly basis.

If flooding map in Zhengzhou were given, one can predict when neighbourhood are subject to higher risks of flooding. A range of technique can be applied to predict flood. The simplest one is logistic regression. But in practice, researchers often use more advanced model machine learning l or even neural network work. (Duncan et al, 2011). Applications includes prediction of flood in India (Bano, Singh & Aggarwal, 2021) and in German (Rozer et al, 2021). The latter, provides 5 minutes before peak time of extreme rainfall events. More recently, Motta, de Castro & Salento (2021) has used mixed approach to predict flooding in an area. The best performing model, they have found, is what is called random forest. In the future, researchers and climate investigators alike can spatial and temporal data of flooding in Zhengzhou city to facilitate this analysis. For data mining, one can sample NASA satellite photos. This process could take days, which would be too ambitious for the short duration of this project. In addition, only one weather station in Zhengzhou would prone to be a problem because it could be raining in one place but not raining in other places.

# Conclusion

Although this report is unable to conclude if climate risk is increasing because of global warming, it has indeed revealed a manipulative statistic. Of the three models used to test significance of global warming, the best performing one showed lower significances.

It has been discovered that, for unknown reasons, the logarithm of precipitation volume is exponentially distributed. The density function can be described by two parameter exponential decay. This knowledge has been used with an attempt to find out probability of extreme events. Much to disappoint, the model has failed to yield stable results. Result shrinks drastically when a higher level of breaks was selected, and shape pattern of trend became distorted.

This report is able to replicate the cases when risks for extreme climate events became higher. The case happens only under computational circumstances. That is when using breaks of about 19 to 20 in a two-decade window. This result can be highly convincing given 20 breaks level is computed via Donna’s Formula. Even so, there is only three two-decade (only about six decades of recorded data) not to mention the result is completely detached from common sense. More advanced way is needed to estimate density of the curve.

Insurance companies may be concerned with if flooding risks in a specific area within the city. The report has proposed methodologies in the last section and flooding map are needed to complete his method. This could have great values for insurers whose portfolio backs a large proportion of immobile assets (housing, factory ect).

Two significant climate effects are found by exploring climate data in Zhengzhou city. First, the daily temperature is shrinking. Seconds, frequency of raining days is decreasing. This serves as an evidence of urban “heat island” effect in Zhengzhou. If this is the case, Zhengzhou authorities may be concerned with setting up more weather stations to improve quality of climate data.

It turns out that the “yellow river” is oblivious. Once about every 20 years, rainfall will exceed 15cm. In 2021, it falls within that expectation, even it has breaks historical record. IPCC (Field et al., 2021) framework suggest there are two conditions for an extreme event to constitute disasters: being “exposed” to the disaster and vulnerable because of lack of capability to predict, adapt, and recover. Most recent flooding in Henan has happened just ten years ago. But apparently Zhengzhou is not prepared for this. Lack of “capacity”, community where the vulnerable congregatein exposure of extreme climate events are prone to harm of disasters such as flooding.

# Reference

Actuaries.org.uk. 2021. Climate Change for Actuaries: An Introduction. [online] Available at: <https://www.actuaries.org.uk/system/files/field/document/Climate-change-report-29072020.pdf> [Accessed 10 August 2021].

Ball, P. (2021). Xi Jinping should take the Zhengzhou floods as a warning from China’s history. The Gardian.

Bano, P., Singh, R., & Aggarwal, G. (2021). Forecasting of Flood in Upper Yamuna Basin by Using Artificial Neural Network and Geoinformatics Techniques & Learning. Ilkogretim Online, 20(5), 3008–3021.https://doi-org.uoelibrary.idm.oclc.org/10.17051/ilkonline.2021.05.326

David P. Doane. (1976). Aesthetic Frequency Classifications. The American Statistician, 30(4), 181–183. https://doi-org.uoelibrary.idm.oclc.org/10.2307/2683757

Duncan, A., Chen, A. S., Keedwell, E., Djordjevic, S., & Savic, D. (2011). Urban flood prediction in real-time from weather radar and rainfall data using artificial neural networks.

Field, C. B., Barros, V., Stocker, T. F., & Dahe, Q. (Eds.). (2012). Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change. Cambridge University Press.

Gan, N., Wang, Z. (2021). Death toll rises as passengers recount horror of China subway floods. CNN

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2021). forecast: Forecasting functions for time series and linear models. R package version 8.15, https://pkg.robjhyndman.com/forecast/.

Lin, L., Wu, Z., & Liang, Q. (2019). Urban flood susceptibility analysis using a GIS-based multi-criteria analysis framework. Natural Hazards, 97(2), 455–475. https://doi.org/10.1007/s11069-019-03615-2

Motta, M., de Castro Neto, M., & Sarmento, P. (2021). A mixed approach for urban flood prediction using Machine Learning and GIS. International Journal of Disaster Risk Reduction, 56, 102154.

Organisation for Economic Co-operation and Development (OECD) (2006) Current approaches in the statistical analysis of ecotoxicity data: A guidance to application - annexes, Paris: OECD (p. 80).

Rözer, V., Peche, A., Berkhahn, S., Feng, Y., Fuchs, L., Graf, T., Haberlandt, U., Kreibich, H., Sämann, R., Sester, M., Shehu, B., Wahl, J., & Neuweiler, I. (2021). Impact‐Based Forecasting for Pluvial Floods. Earth’s Future, 9(2), 1–18. https://doi-org.uoelibrary.idm.oclc.org/10.1029/2020EF001851

Sturges, H. A. (1926). "The choice of a class interval". Journal of the American Statistical Association. 21 (153): 65–66. doi:10.1080/01621459.1926.10502161. JSTOR 2965501.

Su, A. (2021). Amid crushing floods in China, officials focus not on climate change, but on control. Los Angeles Times.

Wu, Z., Shen, Y., Wang, H., & Wu, M. (2020a). Quantitative assessment of urban flood disaster vulnerability based on text data: Case study in Zhengzhou. Water Supply, 20(2), 408–415. https://doi.org/10.2166/ws.2019.171