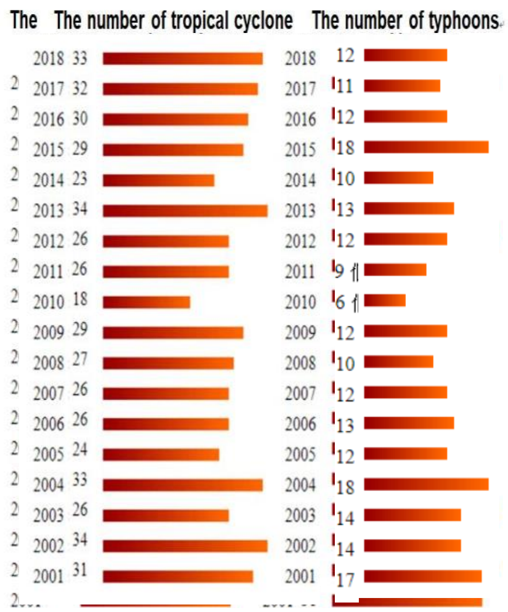
Analysis of the relationship between tropical cyclone warning signals and meteorological data in Hong Kong

1. Overview

1.1 Introduction

A typhoon refers to a mature tropical cyclone which generates and develops from the area between 180° and 100°E in the Northern Hemisphere. Locating in one of the most typhoon-affected area, Hong Kong, a Special Administrative Region of the People's Republic of China, can be geographically regarded as a coastal city and major port in Southern China. Due to its unique geographical environment, it is seasonally influenced by tropical cyclones which occasionally lead to severe damage to Hong Kong. Therefore, making a research on typhoons is important for us to have a deeper understanding and learn to prevent mitigate the loss caused by them.



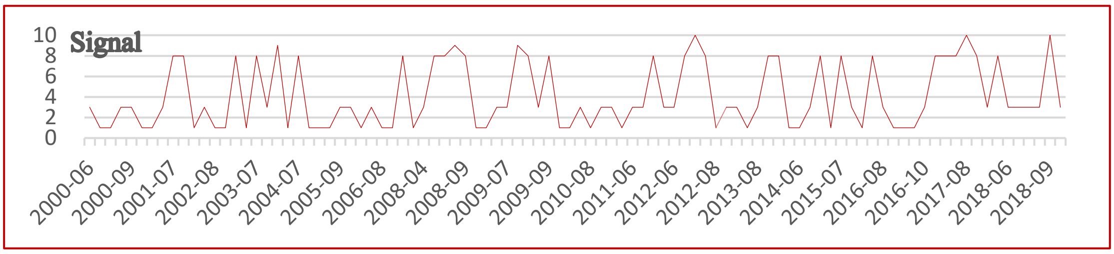
Graph 1: The number of tropical cyclones and typhoons each year from 2001 to 2018.

According to the report from Hong Kong Observatory, during 2001 – 2018, it is reported that the total numbers of tropical cyclones and typhoons are respectively 507 and 255, and the average numbers are 29 per year and 13 per year. The graph (Graph 1) on the left demonstrates the number of tropical cyclones and typhoons.

The meteorological dataset was selected from Hong Kong Observatory. After initial processing and analysis, it is found that the relationship between Typhoon signals and other meteorological data including dew point, temperature, pressure, humidity, cloud and rainfall should be analyzed further. In the following part of this report, the processing and virtualizing measures to find out the potential effects and relationships involve correlation, clustering, network, association rules, mutual information and some other statistical methods.

1.2 Typhoons in HK

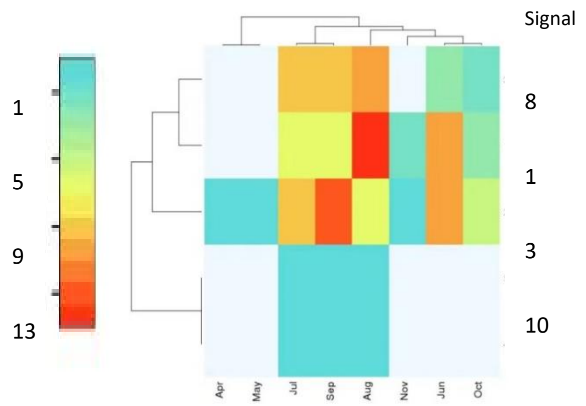
Hong Kong's climate is sub-tropical, tending towards temperate for nearly half the year. Summers in HK are hot and humid while winters are cool and dry, and weather in other months is mild and comfortable. Typhoon seasons (July to September) are associated with the period where Hong Kong is most likely to be influenced by tropical cyclones. The following graph (Graph 2) shows the official released Typhoon signals from 2000 to 2018, and it can be concluded that the time trend graph shows seasonality.



Graph 2: Official released Typhoon signals during 2000 – 2018 by HK observatory

Before the report analyzes the seasonality of typhoons, the typhoon warning signals system in Hong Kong should be mentioned. The Hong Kong's tropical cyclone warning signals are a system of signals utilized to notify the influences or threats of typhoons, which are demonstrated in the following table (Table 1).

In order to analyze the relationship between typhoon seasonality and signals, a heat map has been prepared to compare the frequencies of typhoons with various levels in different months from 2000 to 2018 (Graph 3).



Graph 3: Heat map and clustering analysis of typhoon seasonality and signals frequency.

Notes： 1) x-axis only lists the months with typhoons data. 2) y-axis lists the signal levels. 3) Different colors demonstrate different frequencies of typhoons signals.

It can be concluded that compared with the higher signals, lower signals are released with a higher frequency (There are more signals 1, 3, 8 than signals 9, 10). Meanwhile, from July to Sep, there are more released signals than other months, and higher signals are concentrated during this period.

2. Association and frequent item analysis

2.1 Data Processing

To eliminate the effect of different seasons, monthly data is used as the reference substance. In detail, the daily data is compared with the monthly average data. Then "1" is marked if the daily data is larger than the corresponding monthly average, otherwise "0". Below is table (Table 3) of the count of different warning signals and the count of outcomes on different attributes.

Signal		1	3	8	above 8
Count of Signal		142	96	53	9
Count of	Presure: 1	33	4	1	0
	Temp: 1	75	27	10	0
	Dew point: 1	70	45	18	1
	Relative humidity: 1	66	66	42	8
	Amount of cloud: 1	97	89	51	9
	Rainfall: 1	66	66	42	8

Table 2

In the following table (Table 3), rows represent signals and columns represent attributes. The confidence in cells indicates rules representing the columns are derived from corresponding rows (e.g. If signals equals to 9 or 10, and pressure equals to 0, then confidence equals to 1.00).

Signal	Pressure: 0	Temperature: 0	Dew point: 0	Relative humidity: 1	Amount of cloud: 1	Rainfall: 1
9 and 10	1.00	1.00	0.89	0.89	1.00	0.89
8	0.98	0.81	0.66	0.79	0.96	0.79
3	0.96	0.72	0.53	0.69	0.93	0.69
1	0.77	0.47	0.51	0.46	0.68	0.46

Table 3: Association rules representing the columns are derived from corresponding rows

If there is a strong tendency that tropical cyclones can affect meteorological data, the confidence of corresponding rules will be far away from 0.5, otherwise the daily data would be randomly higher or lower than the monthly average.

2.2 Conclusion

Several conclusions. First, tropical cyclones are usually related to lower pressure and higher amount of cloud. As for temperature, relative humidity and rainfall, the tendencies are not so strong, especially when the signal is 3 or 1. For one specific attribute, higher the signal was, stronger the relationship was. So we can conclude that high signal typhoons are more likely to affect climate data. Next, most of tropical cyclones are not likely to affect dew point because the confidence is closer to 0.5 which is a normal level. Meanwhile, when there is a signal 8 typhoon, higher relative humidity is usually accompanied with lower temperature, comparing to monthly average. Finally, rainfall and relative humidity are of high correlation, which fits with our daily physical experience.

3. Correlation Analysis

3.1 Pre-processing data

Considering the effect of the month, I use raised rain and raised temperature as our study objects, which is calculated by equation as follows:

$$raised\ object = daily\ mean - monthly\ mean$$

Also, the signal level is classified into two categories later as follows: $weak = \{1,3\}$,$strong = \{8,9,10\}$.

3.3 Correlation Analysis

3.3.1 Correlation between signal and rainfall, temperature

I drew the scatter plot and calculated the Kendall Coefficient below (Appendix 2,3) and found that the distributions of two factors, rainfall and temperature are quite similar within groups and different between groups when I divide them by signal levels as in 3.2. It can also be proved by Kendall Coefficient since the overall Tau is way larger than each group's.

Kendall Tau Correlation			
	Weak	Strong	Overall
Signal , Rain	0.2776622772	0.2467966676	0.38594076
Signal , Temp	-0.2210166007	-0.1659626365	-0.29036906

Table 4

Note: (1) See 3.2.Pre-processing data. (2) Kendall's Tau Formula is used here because there are large overlaps on signal levels and typhoons in the same level should not be compared.

It is shown that tropical cyclones do bring rain to Hong Kong and cool down the temperature here. However, due to the low correlation coefficient, it is hard to conclude that stronger tropical cyclones have more positive effects.

3.3.2 Correlation between rainfall and temperature

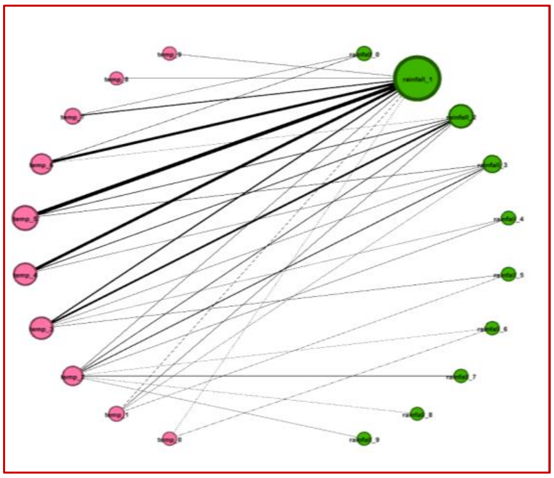
	Pearson	Spearman	Kendall
Rainfall, Temp	-0.593415717	-0.7193271706	-0.5274390578

Table 5

In this part, the 3 correlation coefficients all indicate that rainfall and temperature are significantly related negatively. The reason behind it may be the fact that rainfall can cushion the hot weather in summer.

3.4 Network Analysis

With the inspiration of 3.1, I categorize tropical cyclones into 2 groups, weak and strong based on the signal level, to further discuss their influence to the rainfall and temperature. Below (Appendix 4,5) is the scatter plot and it seems that under strong tropical cyclone, rainfall and temperature have good linear correlation while clustering happens under weak tropical cyclones.



Graph 4: Bipartite network
Note:
(1)The size of each node represents "frequency". The thickness of the edge represents the times that tropical cyclones happen.
(2)The red nodes stand for temperature, green for rainfall.
(3)From top to down, the lowered temperature and raised rainfall get bigger.

I use a bipartite network above to further visualize the cluster. It suggests the combinations of low rainfall and raised temperature happened relatively frequently under the mild tropical cyclone, which may be worth further research.

3.4 Conclusion

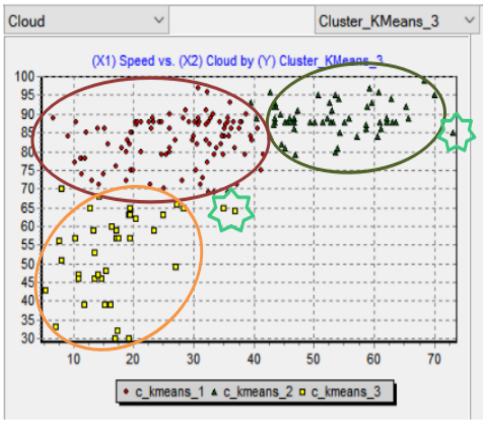
Tropical cyclones generally bring rain and lower temperature in Hong Kong, and there may be two different affecting patterns in signal {1, 3} and {8, 9, 10}. Also, raised rainfall is highly negatively related to raised temperature. Under weak tropical cyclones, the combination of little rainfall and raised temperature happens relatively frequently.

4. Clustering

4.1 Mean Wind Speed VS Mean Amount of Cloud

4.1.1 K-means method

The figure shows the centroids the K-means method is used. From the data, it can be concluded that the faster the wind speed, the more amount the cloud.



Cluster centroids

Attribute	Cluster n?	Cluster n?	Cluster n?
Cloud	82.937984	89.196581	53.465116
Speed	28.042636	54.371795	17.983721

Use GROUP CHARACTERIZATION for detailed comparisons

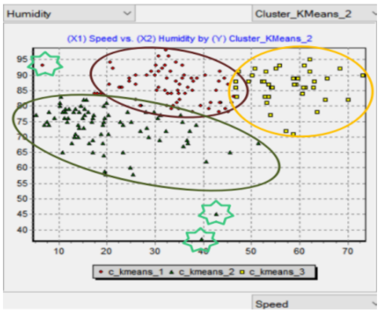
The graph above shows the relation between the mean wind speed and the mean amount of the cloud. The yellow squares indicate that the low wind speed corresponds to the small quantity of cloud. From green triangles, it can be found that the high wind speed wind results in a large quantity of cloud. These two kinds of points reflect the mean amount of the cloud is in proportion to the mean wind speed. However, the red dots show the relation of the low wind speed and a large amount of cloud. It is because that the weak low-speed wind usually comes before the strong typhoon and in this situation, the amount of the cloud brought by low-speed wind will be larger. In general, higher wind speed results in a larger amount of cloud and lower wind speed results in fewer amounts of cloud, and the abnormalities are derived from that some weak winds come before the strong high-speed tropical cyclone.

4.1.2 Outliers

From the scatter diagram, there are three outliers. The first outlier (from left to right) is called Doksuri. Its signal is 8 and it greatly influenced Hong Kong and Guangdong Province. The second one is called Wukong. It greatly influenced Hai Nan province. The third outlier is Vicente, which is a huge typhoon. Its signal is high and the speed of movement is fast, but the mean amount of the cloud is lower than expected. The reasons why these three Typhoons become outliers are inconclusive and remain to be studied.

4.2 Mean Wind Speed VS Mean Relative Humidity.

4.2.1 K-means



The right graph shows the relationship between the mean wind speed and the mean relative humidity. By the K-means method, the cluster centroids can be obtained, which shows there is a slight positive correlation between humidity and wind speed.

4.2.2 Outliers

Cluster centroids

Attribute	Cluster n?	Cluster n?	Cluster n?
Humidity	86.272727	72.648936	87.023529
Speed	34.286364	21.694681	58.135294

Use GROUP CHARACTERIZATION for detailed comparisons

There are three outliers in the scatter graph. The first and second outlier is the same wind called Yutu. It had great impacts on the Philippines. The third one is a severe Tropical Storm with no name because its signal is too small. The reasons for these 3 outliers remain to be studied.

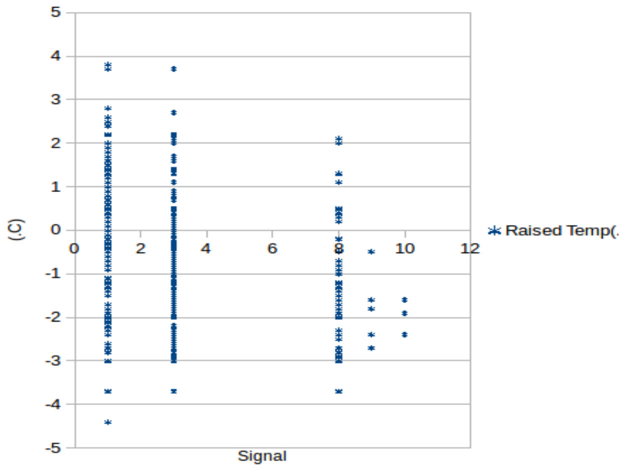
5. Conclusion and future work

Through the analysis with multiple methods, a conclusion can be drawn that tropical cyclones are likely to bring lower pressure, lower temperature, more rain and higher humidity, while the directions of No.8 typhoon have little effect on those meteorological data. Multivariable regression models can also be tried to further investigate relationships between different attributes. What's needed to be improved is to refer to related typhoon formation theories in order to give theoretical reasons on our analysis.

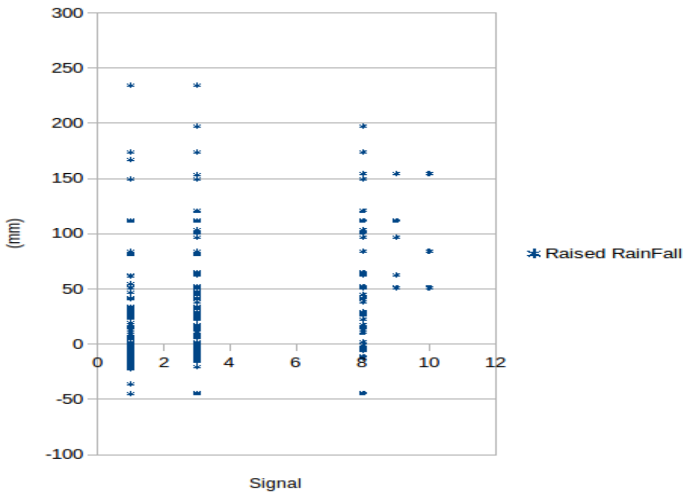
Appendix

1. Hong Kong's tropical cyclone warning signals by HK observatory

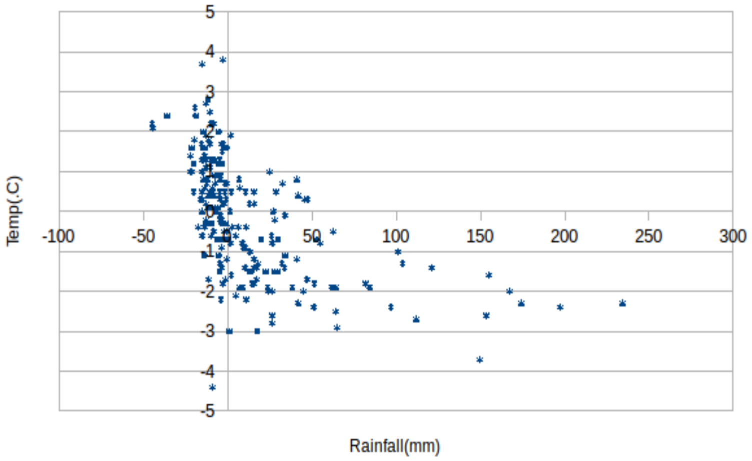
2. Scatter plot between signal and raised temperature



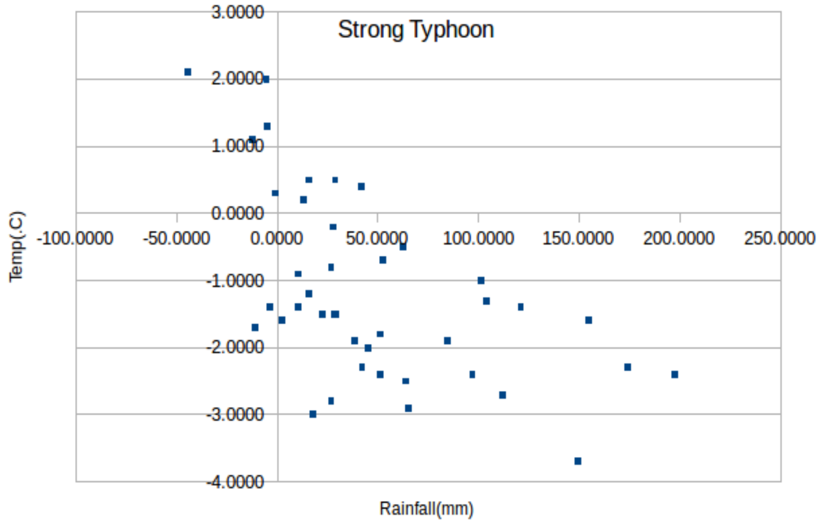
3. Scatter plot between signal and raised rainfall



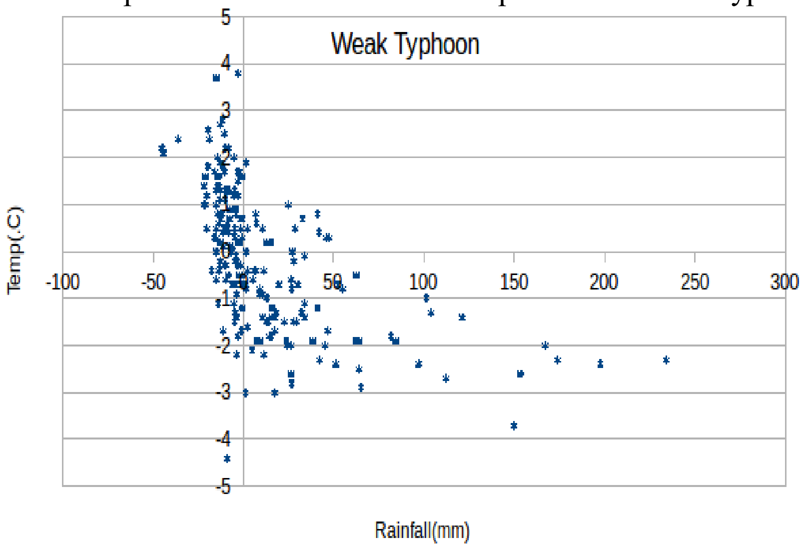
4. Scatter plot between rainfall and temperature



5. Scatter plot between rainfall and temperature of strong typhoon



6. Scatter plot between rainfall and temperature of weak typhoon



7. The calculation of mutual information

	Pressure:	Temperature:	Dew point:	Relative humidity:	Amount of cloud:
	TRUE	TRUE	TRUE	TRUE	TRUE
NE p(x1,y1)	0	4/53	5/53	10/53	13/53
NW p(x2,y1)	0	4/53	1/53	6/53	12/53
SE p(x3,y1)	1/53	1/53	10/53	16/53	15/53
SW p(x4,y1)	0	1/53	2/53	10/53	11/53
	FALSE	FALSE	FALSE	FALSE	FALSE
NE p(x1,y2)	14/53	10/53	9/53	4/53	1/53
NW p(x2,y2)	12/53	8/53	11/53	6/53	0
SE p(x3,y2)	15/53	15/53	6/53	0	1/53
SW p(x4,y2)	11/53	10/53	9/53	1/53	0
p(x1)	14/53	14/53	14/53	14/53	14/53
p(x2)	12/53	12/53	12/53	12/53	12/53
p(x3)	16/53	16/53	16/53	16/53	16/53
p(x4)	11/53	11/53	11/53	11/53	11/53
TRUE Total p(y1)	1/53	10/53	18/53	42/53	51/53
FALSE Total p(y2)	52/53	43/53	35/53	11/53	2/53

Pressure:				
P(x1,y1)	0	P(x1,y2)	0.005031599	Sum all boxes:
P(x2,y1)	0	P(x2,y2)	0.004312799	0.023021349
P(x3,y1)	0.022598173	P(x3,y2)	-0.012874621	Normalize by ln(2)
P(x4,y1)	0	P(x4,y2)	0.003953399	0.033212786
Temperature:				
P(x1,y1)	0.031316517	P(x1,y2)	-0.024034045	Sum all boxes:
P(x2,y1)	0.042950531	P(x2,y2)	-0.029641254	0.048347314
P(x3,y1)	-0.02084683	P(x3,y2)	0.040911305	Normalize by ln(2)
P(x4,y1)	-0.01377714	P(x4,y2)	0.02146823	0.06975043
Dew point:				
P(x1,y1)	0.004745353	P(x1,y2)	-0.00456604	Sum all boxes:
P(x2,y1)	-0.02650918	P(x2,y2)	0.068061457	0.105555278
P(x3,y1)	0.11507859	P(x3,y2)	-0.064062498	Normalize by ln(2)
P(x4,y1)	-0.02357841	P(x4,y2)	0.036386008	0.152284076
Relative humidity:				
P(x1,y1)	-0.01959433	P(x1,y2)	0.024123296	Sum all boxes:
P(x2,y1)	-0.05213489	P(x2,y2)	0.099537675	0.132489849
P(x3,y1)	0.070225599	P(x3,y2)	0	Normalize by ln(2)
P(x4,y1)	0.025907946	P(x4,y2)	-0.015575446	0.191142447
Amount of cloud:				
P(x1,y1)	-0.0087423	P(x1,y2)	0.012039385	Sum all boxes:
P(x2,y1)	0.008709347	P(x2,y2)	0	0.022130986
P(x3,y1)	-0.00737894	P(x3,y2)	0.009519925	Normalize by ln(2)
P(x4,y1)	0.007983568	P(x4,y2)	0	0.031928264