USING PRINCETON PRECIPITATION CLIMATOLOGY TO PREDICT FUTURE PRECIPITATION EVENTS

Tyrone Zhang

A SENIOR THESIS

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF BACHELOR OF ARTS

RECOMMENDED FOR ACCEPTANCE

BY THE

DEPARTMENT OF GEOSCIENCES

Adviser: Frederik J. Simons

January 15, 2021

This paper represents my own work in accordance with University regulations,

Tyrone Zhang

Abstract

Princeton's climate has four seasons, with strong temperature variations, and precipitation occurring throughout the year. Statistically, precipitation event sequences can be characterized as drawn from exponential distributions in the three variables *duration*, *intensity* (the total precipitation divided by the duration), and interevent *separation*. The shortest and least intense precipitation events are the most frequent. Analyzing the precipitation measured from 2017 to the present day by a Vaisala WXT530 weather station located on the roof of Guyot Hall, I first summarize the data in terms of exponential distributions and their parameters, by season and by year. Subsequently, I evaluate the skill in predicting the arrival, duration and intensity of precipitation events solely based on this local "climatology", before including other variables logged by the weather station.

Acknowledgements

I would like to acknowledge my senior thesis advisor Frederik J. Simons for giving me constant feedback on my work as well as providing me with the data that he is collecting on top of Guyot Hall.

Contents

List of Figures

List of Tables

Introduction

The climatology of Princeton is one that belongs to the mid-latitudes, which is characterized by having four seasons that results in a large variation in temperature throughout a year. In terms of the average calculated between 1981 and 2010, Princeton gets an average of 1227 mm of precipitation annually, and the precipitation distribution throughout the year is fairly even, with less precipitation in the winter (PRISM, 2020). According to the Koppen-Geiger Climate Classification, Princeton, NJ lies in the classification Cfa, which denotes a temperate climate, with no dry season, and hot summers defined as reaching 22°C or higher (Peel et al., 2007). Princeton having no dry season means that precipitation is well spread out throughout the year.

Our weather station on the top of Guyot Hall is Vaisala weather transmitter WXT530 series. It measures six weather parameters of air pressure, temperature, humidity, rainfall, wind speed, and wind direction. The rainfall is measured using an acoustic Vaisala RAINCAP Sensor, which helps avoid the complications of flooding, wetting, and evaporation losses (Vaisala, 2017). By analyzing the precipitation bwteen 2017 and today, I first summarize the data using this climatology, in terms of exponential distributions and their parameters, by season and by year, before attepting to predict precipitation events based on other variables that are observed by the weather station.

Methods

I shall define the following terms. The time series of **precipitation** as recorded by the instrument is denoted e_i, where i indexes the measurement intervals, each 60 s long. I define a precipitation **event** E_j^τ as a sequence of **duration** $d_j \geq \tau$ containing contiguous nonzero precipitation measurements $e_i > 0$, flanked left and right by zeros, $e_i = 0$, and where τ is in minutes.

Furthermore, I define a precipitation **non-event** N_j^τ, as having a contiguous set of zeros, $e_i = 0$, whose combined duration exceeds τ, flanked left and right by non-zero values, $e_i = 0$.

One more term to define is **precipitation intensity**, which for a precipitation event E_j^τ is the total amount of precipitation divided by its duration, i.e.,

$$I_j^\tau = \frac{\sum_i e_i}{d_j}, \quad \text{for } i \text{ belonging to the event } E_j^\tau. \tag{1}$$

For further analysis, I breaking down the year into seasons, as different seasons may have different characteristics with regards to precipitation. I will define the seasons as follows: Winter will be December, January, and February. "Winter" of a certain year contains December of the previous year. Spring will be March, April, and May. Summer is June, July, and August. Fall is September, October, November.

Figure 1 shows the distribution of durations of 3198 precipitation events E_j^1, i.e. E_j^τ where $\tau = 1$ min for the year 2019, broken down by season. EXPLAIN THE X-BIN SPACING. I used an exponential fit to the frequency-duration histograms for all 1253 events E_j^2, i.e. E_j^τ where $\tau = 2$ min. For all the other years, as shown in Figures 2, 3 and 4, I used a similar procedure.

Excluding the first interval shown, focusing on events of duration greater than or equal to 2 min, we propose an exponential model for the histogram, with the following equation:

$$F = \beta e^{\alpha d}, \tag{2}$$

where F is the frequency and d the duration, and with β the unitless frequency coefficient and α is the exponential coefficient (in units of min^{-1}). Table 1 shows the coefficients β and the exponential coefficients α from looking at the yearly frequency of precipitation duration.

We shall also propose the following equations which will also describe an exponential model for the histogram regarding non-precipitation events, which is described by the following equation:

$$F_{np} = \gamma e^{\delta D}, \tag{3}$$

where F_{np} is the frequency of non-precipitation events, D being duration, and γ being the unitless frequency coefficient and δ is the exponential coefficient. Table 3 shows the coefficients γ and exponential coefficients δ from looking at yearly frequency of non-precipitation event durations.

Another similar equation for precipitation intensity can be described by the following equation:

$$F_{inten} = \varepsilon e^{\zeta I}, \tag{4}$$

where F_{inten} is the frequency of intensity of precipitation events, I is the intensity of the precipitation events, ε is the unitless frequency coefficient, and ζ is the exponential coefficient. Table 4 shows the coefficients ε and the exponential coefficients ζ from looking at yearly frequency of intensity of precipitation events.

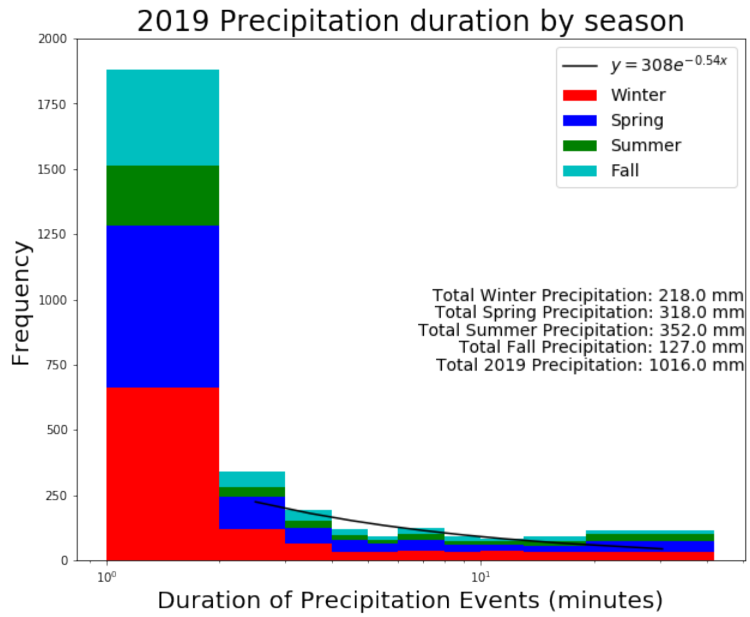

Figure 1: Distribution of precipitation duration in 2019, with a breakdown by seasons. The minimum duration, $\tau = 1$ min, and the maximum duration over the data set is 368 min, but the horizontal axis was limited to the 98th percentile of the durations, 42 min, for clarity. Superposed is a least-squares fit of a line in log-log space of duration versus frequency for the entire year, excluding the 1–2 min bin, quoted as the equivalent exponential in this space.

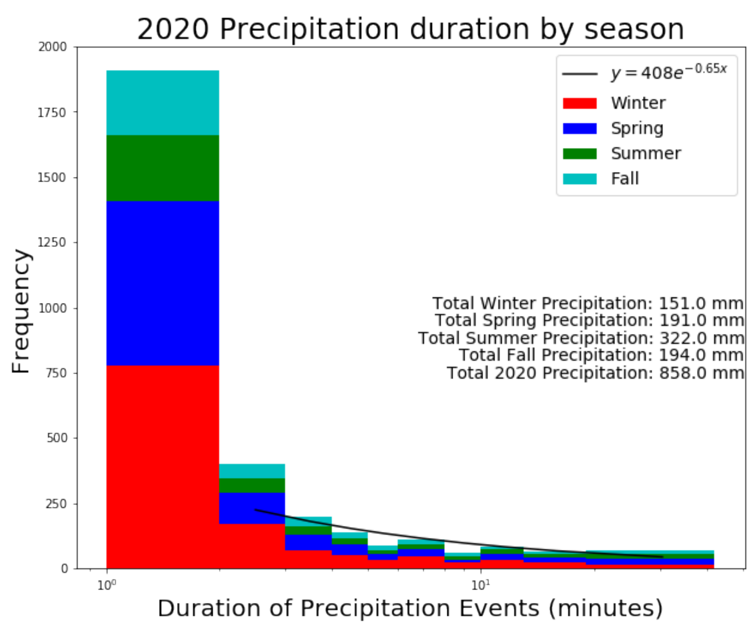

Figure 2: Distribution of precipitation duration in 2020, with a breakdown by seasons. 2020 is also incomplete as collection still ongoing. The layout is as in Figure 1.

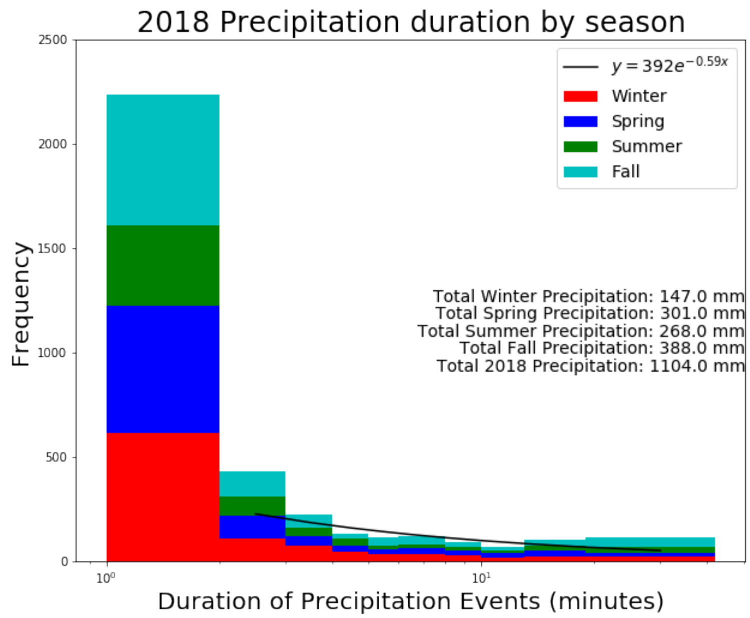

Figure 3: Distribution of precipitation duration in 2018, with a breakdown by seasons. The layout is as in Figure 1.

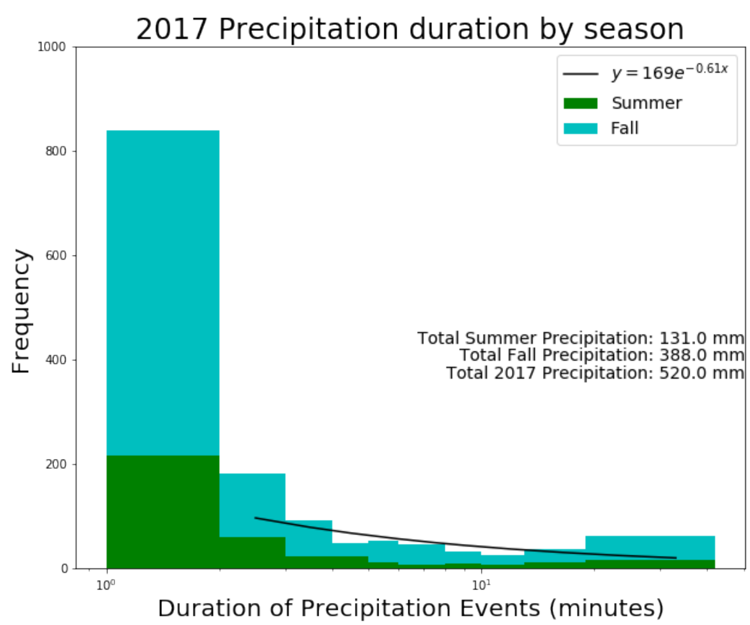

Figure 4: Distribution of precipitation duration in 2017, with a breakdown by seasons. The layout is as in Figure 1. Note that data collection began on 16 July 2017.

Season	Annual		Winter		Spring		Summer		Fall	
Year	β	α	β	α	β	α	β	α	β	α
2017	*169*	*-0.61*	NaN	NaN	NaN	NaN	*57*	*-0.70*	108	-0.57
2018	392	-0.59	148	-0.74	125	-0.76	65	-0.52	91	-0.50
2019	308	-0.54	118	-0.62	107	-0.58	32	-0.31	57	-0.60
2020	408	-0.65	201	-0.80	112	-0.66	48	-0.42	*64*	*-0.66*

Table 1: Coefficients found from the yearly distribution of precipitation duration (as in equation 2) as well as the seasonal distribution of precipitation duration. Italics refer to values obtained using incomplete information. NaN means there was no information.

Based on Table 1, the β values are similar to each other with the exception of 2017, which only had partial data starting in the Summer. Even, with the partial data we have from 2017 and 2020, it is clear that the α values are similar to each other. Focusing on the years with complete data, 2018 and 2019, it is clear that yearly variations exists between them. All α values are negative.

In the seasonal variations, we see that Summer has α values that are less negative compared to the other seasons as well as having a lower β compared to the other seasons. However looking at Table 2, we see that the summers also have a lot of precipitation. The fewer amounts of precipitation events in the Summer, but with a lot of precipitation gives a higher intensity for the Summer. For the Winter, the α are the furthest from 0 and the β are large. However, Winter tends to have the lowest values for total precipitation and the combination of lots of preciptation events and low precipitation totals results in the lowest intensities among the four seasons.

The Spring appears to mimic the Winter in that there are a lot of preciptiation events, but also shares the quality of Summer in having fairly high precipitation totals. The Fall shares the Summer quality of having relatively few preciptation events, but tends to have smaller precipitation totals, so its intensity is less than the Summer, but greater than the winter precipitation intensity.

Looking at trying to extend the exponential fit towards non-precipitation event durations, the exponentials for such duration, δ is less negative overall compared to the exponentials from the precipitation event duration, α. Some of this is explained from the fact that the duration

Season	Annual		Winter		Spring		Summer		Fall	
Year	T	I	T	I	T	I	T	I	T	I
2017	*520*	*0.07*	NaN	NaN	NaN	NaN	*131*	*0.07*	388	0.07
2018	1104	0.06	147	0.03	301	0.07	268	0.08	388	0.07
2019	1016	0.06	218	0.04	318	0.06	352	0.11	127	0.05
2020	*858*	*0.06*	151	0.03	191	0.04	322	0.12	*194*	*0.06*

Table 2: Total precipitation (in mm) and average intensity (in mm/min) of precipitation for each year and season. Italics refer to values obtained using incomplete information. NaN means there was no information.

of non-precipitation events range from 1 minute to over 2000 minutes, whereas the duration of precipitation events range from 1 minute to just over 300 minutes. At the same time how we bin the durations for both precipitation and non-precipitation events will affect how we get the fits.

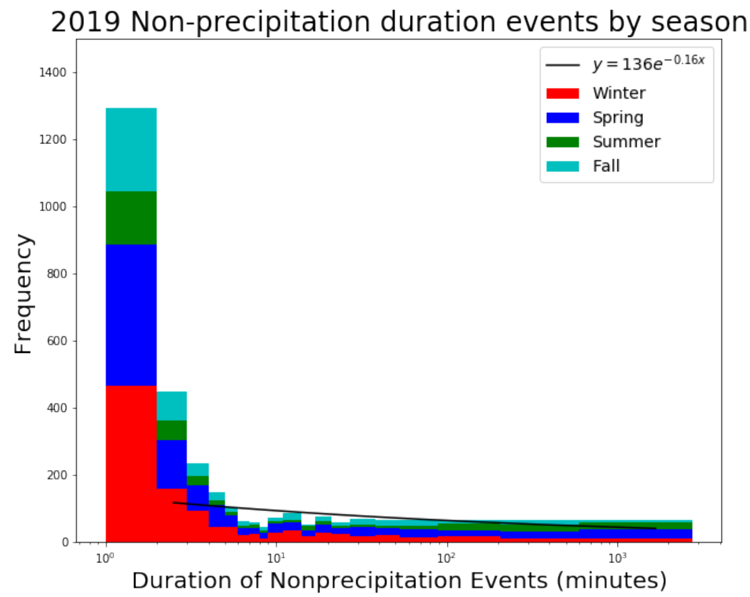

Figure 5: Distribution of non-precipitation event duration in 2019, with a breakdown by seasons. The minimum duration, $\tau = 1$ min, and the maximum duration over the data set is 2613 min, but the horizontal axis was limited to the 98th percentile of 2350 min, for clarity. Superposed is a least-squares fit of a line in log-log space of duration versus frequency for the entire year, excluding the 1–2 min bin, quoted as the equivalent exponential in this space.

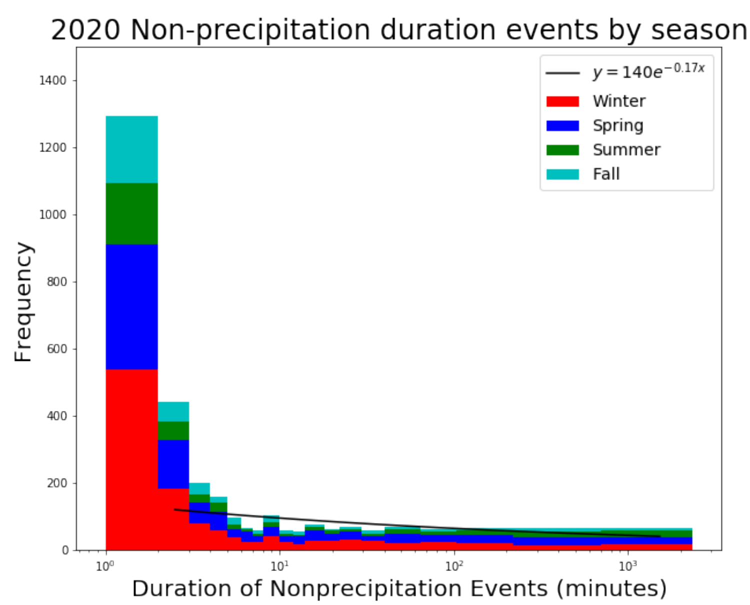

Figure 6: Distribution of non precipitation event duration in 2020, with a breakdown by seasons. 2020 is also incomplete as collection still ongoing. The layout is as in Figure 5.

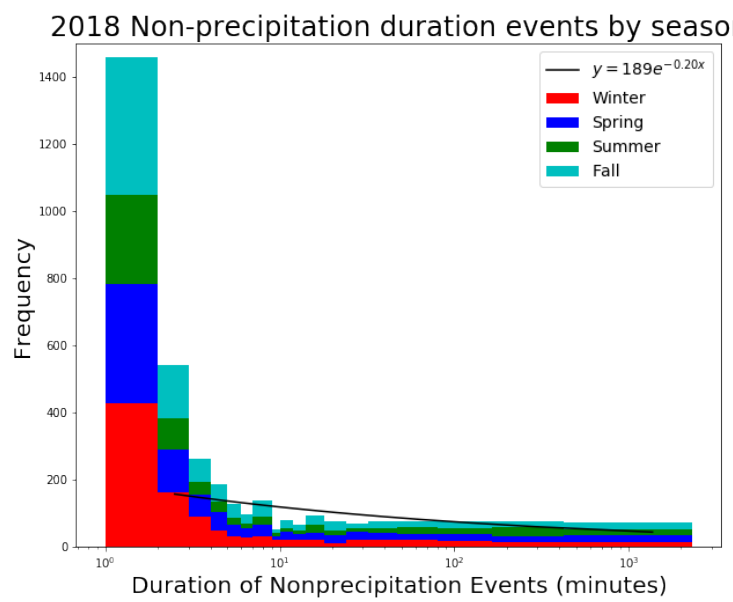

Figure 7: Distribution of precipitation duration in 2018, with a breakdown by seasons. The layout is as in Figure 5.

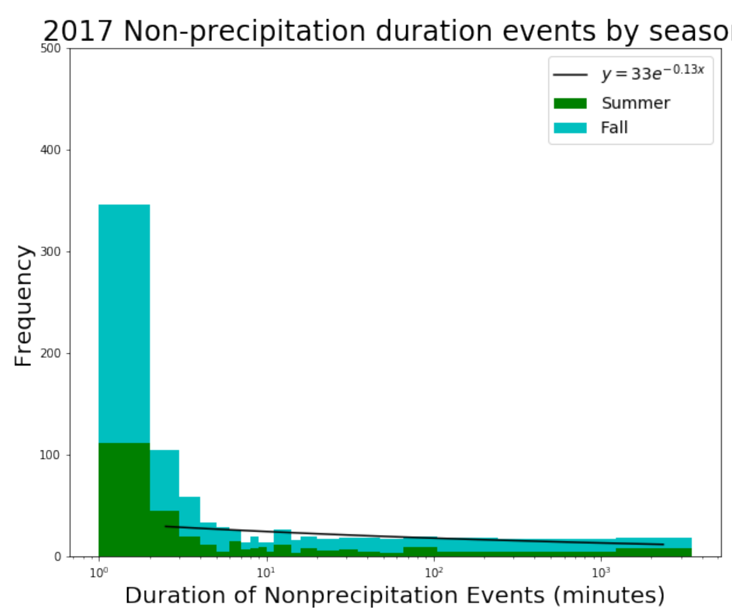

Figure 8: Distribution of non-precipitation event duration in 2017, with a breakdown by seasons. The layout is as in Figure 5. Note that data collection began on 16 July 2017.

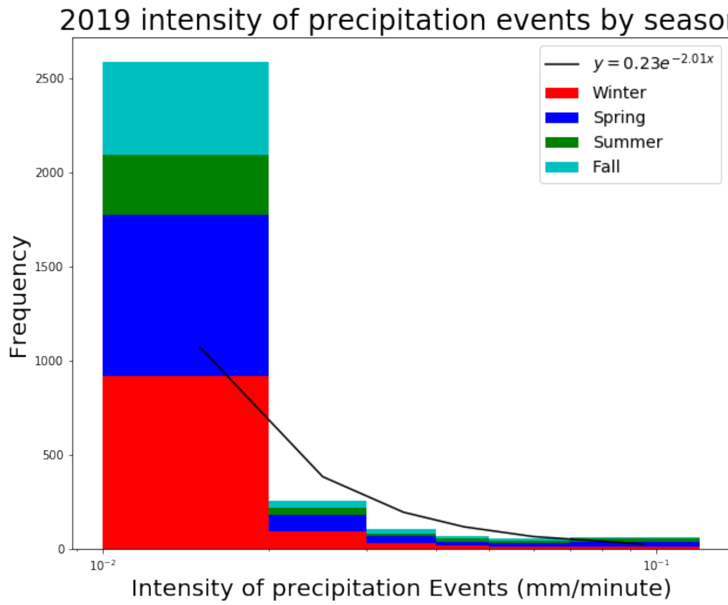

Figure 9: Distribution of precipitation intensity in 2019, with a breakdown by seasons. The minimum intensity, $I = 0.01$ mm/min, and the maximum intensity over the data set is 0.4 mm/min, but the horizontal axis was limited to the 98th percentile of the intensity, 0.1 mm/min, for clarity. Superposed is a least-squares fit of a line in log-log space of duration versus frequency for the entire year,

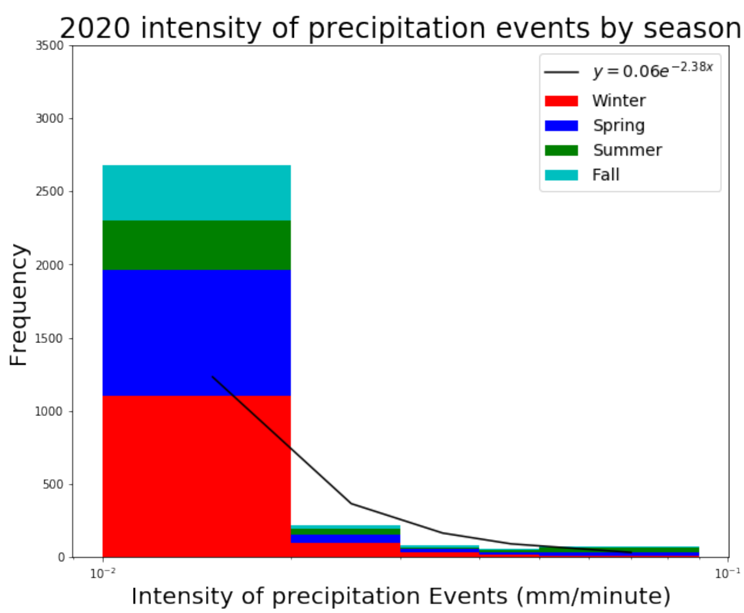

Figure 10: Distribution of precipitation intensity in 2020, with a breakdown by seasons. 2020 is also incomplete as collection still ongoing. The layout is as in Figure 9.

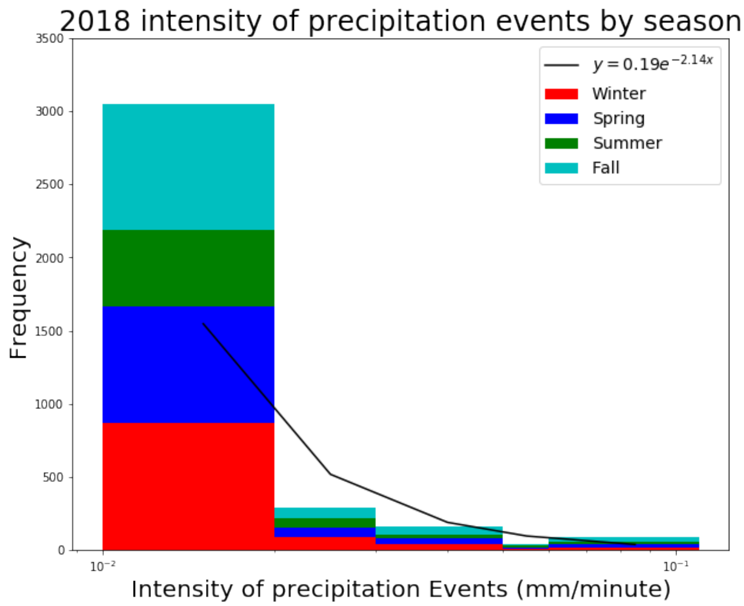

Figure 11: Distribution of precipitation intensity in 2018, with a breakdown by seasons. The layout is as in Figure 9.

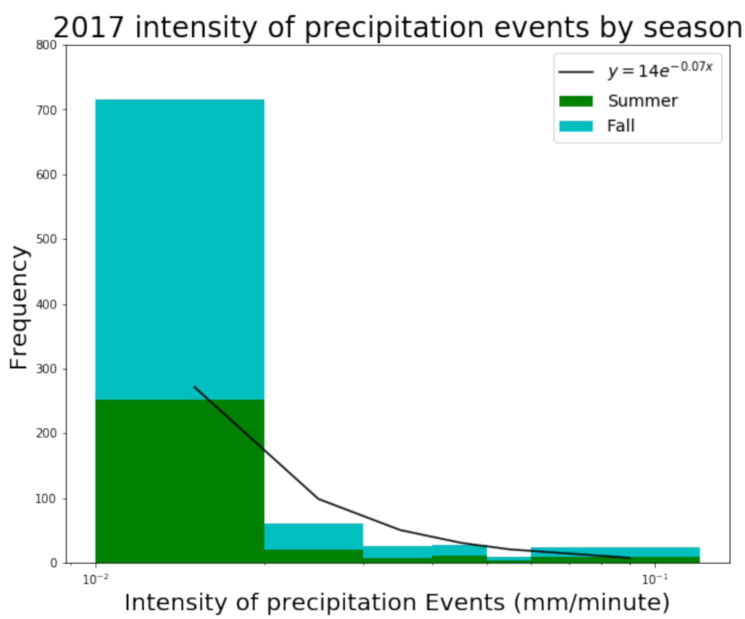

Figure 12: Distribution of precipitation intensity in 2017, with a breakdown by seasons. The layout is as in Figure 9. Note that data collection began on 16 July 2017.

Season	Annual		Winter		Spring		Summer		Fall	
Year	γ	δ	γ	δ	γ	δ	γ	δ	γ	δ
2017	*33*	*-0.13*	NaN	NaN	NaN	NaN	*12*	*-0.15*	19	-0.12
2018	189	-0.20	59	-0.28	50	-0.29	26	-0.10	55	-0.22
2019	136	-0.16	63	-0.30	46	-0.16	11	0.03	24	-0.16
2020	140	-0.17	64	-0.23	46	-0.16	13	-0.02	20	-0.21

Table 3: Coefficients found from the yearly distribution of non-precipitation event duration (as in equation 3) as well as the seasonal distribution of non-precipitation event duration. Italics refer to values obtained using incomplete information. NaN means there was no information.

Season	Annual		Winter		Spring		Summer		Fall	
Year	ε	ζ	ε	ζ	ε	ζ	ε	ζ	ε	ζ
2017	*0.06*	*-1.98*	NaN	NaN	NaN	NaN	*0.03*	*-1.94*	0.04	-2.00
2018	0.19	-2.14	0.02	-2.41	0.03	-2.24	0.10	-1.87	0.08	-2.05
2019	0.23	-2.01	0.03	-2.29	0.07	-2.02	0.25	-1.47	0.03	-2.16
2020	0.06	-2.38	0.002	-3.00	0.02	-2.38	0.10	-1.74	0.02	-2.18

Table 4: Coefficients found from the yearly distribution of precipitation intensity (as in equation 4) as well as the seasonal distribution of precipitation intensity. Italics refer to values obtained using incomplete information. NaN means there was no information.

In contrast, Table 4 for intensity of precipitation events show a more negative exponential compared to either precipitation or non-precipitation event durations. This is partially due to the fact that there are less bins in the intensity of preciptiation events as seen that the vast majority of precipitation events are not particularly intense on average.

Results

Running the model one hundred times let us look at what the average precipitation total the model yields, it yields 195 mm compared to the Summer 2018 total of 268 mm that we used. Furthermore, the accuracy of model precipitation matching model precipitation is only 6%. Looking at 13, we see that the model precipitation total is lower than the actual precipitation observed in the Summer of 2018. Furthermore, the precipitation seem to be more frequent compared to the observed 2018 Summer data.

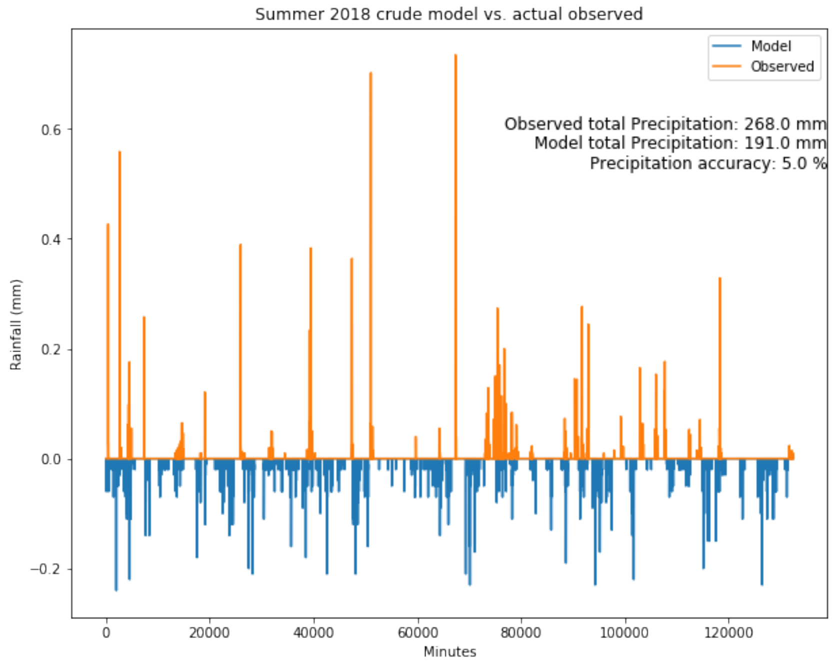

Figure 13: One model run using the exponentials calculated from Summer 2018.

Discussion

This section is blank for now.

Conclusions

This section is blank for now.

References

Peel, M., Finlayson, B., & McMahon, T., 2007. Updated world map of the Koppen-Geiger climate classification, *Hydrol. Earth Syst. Sci.*, **11**, 1633–1644.

PRISM, 2020. PRISM climate data, http://prism.oregonstate.edu.

Vaisala, 2017. Vaisala weather transmitter WXT530 series, `https://www.vaisala.com/sites/default/files/documents/WEA-MET-G-WXT530\%20Datasheet__Low.pdf`.