A Dry Jacob's Well?

Visualizing Springflow Declines using Time-series Data and Random Variables

Weston Kirk

December 2024

Jacob's Well, April 20th, 2019, Aleksomber, CC BY-SA 4.0, via Wikimedia Commons [1]

Jacob's Well, December 1st 2024, Weston Kirk

1 Introduction

Jacob's Well is a spring which emerges in the bed of Cypress Creek, fed by the Middle Trinity Aquifer, near Wimberley, Texas [21]. Today, Jacob's Well is a popular swimming hole (when flows are high enough), and a significant cultural monument for the community of Wimberley and the surrounding Hill County, also acting as a major economic driver via feeding the downstream flow of Cypress Creek and how central water is for tourism in the area [24] [17], furthermore, it supports an extensive cave system far beyond what meets the eye [15]. The first people to know of Jacob's Well were Indigenous groups such as the Tonkawa, Jumano, and Commanche [10], current archeological evidence is sparse due to the recurrent flooding of the creek, but Jacob's Well was almost certainly an important locale for these groups [24]. In past, present, and future, Jacob's Well "carries an aura of mysticism" [24] which represents a place of immense importance for those in the Texas Hill Country.

2 Hydrogeologic Background

Central Texas experiences long periods of drought, interspersed with floods, especially, and is known as "Flash Flood Alley" for its extreme rain events. In the future, longer droughts, interspersed with extreme rainfall events of greater magnitude are projected [20]. The limestone geology of this region supports large karstic aquifer systems such as the Edwards and Trinity Aquifers. Unlike the Edwards, the Trinity Aquifer has a very slow recharge rate, and has seen aquifer-wide declines in water levels [7]. This means that the Trinity Aquifer is far more vulnerable to predicted climatic shifts, as it cannot quickly recharge to capacity like the Edwards Aquifer in major flood events, so it must be managed carefully. It should be noted that declines in the Edwards-Trinity Aquifer [1], which presumably has similar recharge characteristics to the Trinity Aquifer, has experienced the drying of several large West Texas springs, after "large irrigation withdrawals" according to a 1994 study employing a finite-element model [12]. Furthermore, this study found that the Trinity is "hydraulically connected" to the Edwards, providing underground recharge at a rate of 500 cfs/s [12].

3 Recent Developments and Water Laws

In addition to climatic shifts, Central Texas is experiencing massive strain from development along the I-35 corridor, resulting in land clearing and increased demand of limited water resources. In fact, Hays County (where Jacobs Well is located), is the fastest growing county in Texas [9]. More residents and developments reduce the ability for land to recharge aquifers, and increases demand for residential water. This leads us to 2022, where Aqua Texas, the "primary water provider for Hays County" [23] overpumped its legal limit of 90 million gallons, for a total of 162.1 million gallons. In 2023, Aqua Texas pumped 156.4 gallons of water [19]. The only penalty for these actions are fines, which are disputed by Aqua Texas–and a number of legal proceedings regarding this case are ensuing [19]. As current water laws in Texas stand, specifically the Rule of Capture, there is nothing preventing private landowners from taking as much water from a well as that entity pleases, so long as there is no "malicious" intent [18]. Essentially, the deepest well has the most sure access to water, causing existing wells of longtime landowners to run dry, and look to other sources for drinking water. This, coupled with the cultural and economic significance of Jacob's Well to Wimberley and the surrounding Hill Country, makes the springflow of Jacob's Well an environmental justice issue. In addition, the karstic caves of the Edwards and Trinity Aquifers are home to many endemic, some endangered, cave- and spring-dwelling species which are very sensitive to changes in habitat, which are affected by water level declines [6]. While downloadable continuous data is only available online since 2005, records go back much further. Before the year 2000, Jacob's Well had not run dry–but zero flow conditions have been recorded in 6 periods since then [17].

4 Statistical Modeling

Subsurface groundwater flows are extremely complex, and our understanding of recharge, subsurface flow dynamics, and discharge are constantly evolving. The aforementioned 1994 finite-element modeling yielded conclusions regarding the connectivity of the Trinity and Edwards Aquifers [12], which is useful for groundwater management decisions. Physics-based simulations have proven effective for groundwater modeling, but cannot completely describe systems of this complexity and magnitude. In many cases, relying on observed statistical data, especially for management decisions and predictions, is an effective tool. In Sepúlveda 2009, it was shown that Artificial Neural Networks outperformed hydraulically-based methods such as the Theis, Hantush-Jacob, and Darcy-Weisbach equations when predicting spring flows in the Upper Floridan aquifer [16]. Not only have statistical methods proven effective in springflow prediction, but have even been able to demonstrate and quantify the effects of agricultural practices and groundwater pumping in the Minqin Oasis in China [8]. In fact, using ANNs for groundwater modeling represents a "superior alternative to traditional numerical modeling approaches," especially in karstic aquifer systems which requires "less development" than numerical simulations–though both methods are necessary for effective management [5]. Creating a logistic regression classifier to predict zero-flow days is a tangible idea for future work. However, similar to the pedagogy of CS 109, contextualizing the system using first principles is important, so the scope of this study will focus on the MLE of the Poisson and Exponential Distributions across time-series data. [2]

5 Datasets

To conduct our time series analyses of random variables, we obtained a spring discharge dataset from USGS which contained daily minimum, maximum, and mean spring discharge values for Jacob's Well from 2005-04-23 to 2024-11-26, henceforth referred to as time period. For these same dates, we obtained daily rainfall data from NOAA from the WIMBERLEY 1 NW station (GHCND:USC00419815), which represented the closest rain gauge with data matching this times-

[1]The Edwards-Trinity Aquifer is distinct from the Edwards and Trinity Aquifers, yet hydraulically connected. For clarification, see the Texas Water Development Board's website: https://www.twdb.texas.gov/groundwater/aquifer/major.asp

[2]Note to reader: Although I did not have the data resources to create a fully functional ANN within the timeframe of the 109 challenge, I plan on creating one over winter break/in future quarters. If you know of any professors working on geospatial or hydrogeologic ML, please send me an email at `wkirk@stanford.edu`.

pan. We stored this information in a dictionary with the keys as dates, each having a value of a tuple holding the spring discharge as its first entry, and rainfall as its second entry.

6 Methods and Results

6.1 PMF of Spring Discharge

The first step in data analysis was to create a PMF of daily mean spring discharge from Jacob's Well for our time period (Figure 1). The maximum daily discharge for this time period was 101 cfs, which occurred on 2015-05-28 after the 2015 Memorial Day Floods [2]. We discretized the data into 103 categories, incrementing by 1 cfs, including the minimum value and excluding the maximum, $[\min, \max)$, where days with a mean value of zero being put in a separate category. The colors in the graph correspond to flow indicators set by Hays Trinity GCD Rule 15 [22], seen in Supplementary Table 1.

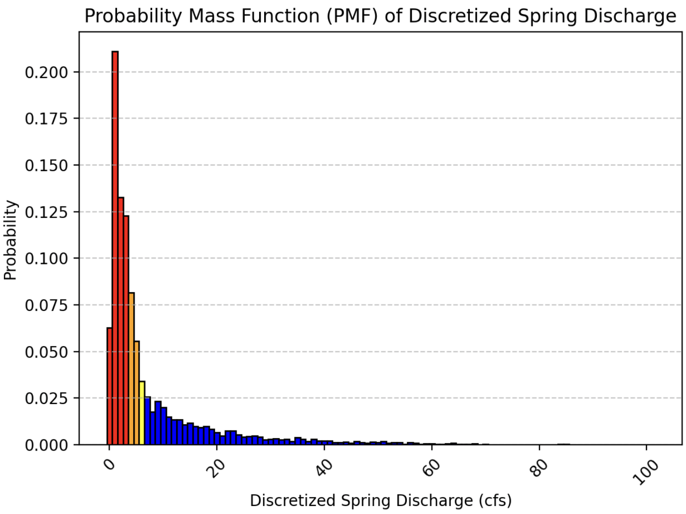

Figure 1: PMF of daily mean spring discharge for Jacob's Well

6.2 Discharge to Rainfall Ratio

While aquifer water level is more commonly referenced in relation to spring discharge, we use rainfall as a proxy. While this is not the most direct measure, we know that rainfall directly influences aquifer water level, which directly influences spring discharge. To formalize this, for each date, we summed the daily rainfall over the previous 365 days to obtain a cumulative value for each date. To check that this relationship was relevant, we looked at the correlation between $R \sim$ Cumulative Rainfall Over the Prior Year and $D \sim$ Spring Discharge. This is the derivation used to calculate correlation, where all R_i's and D_i's were looped over using double summation,

$$\begin{aligned}\rho(R, D) &= \frac{\mathrm{Cov}(R, D)}{\sqrt{\mathrm{Cov}(R, R) \cdot \mathrm{Cov}(D, D)}} \\ &= \frac{E[RD] - E[D]E[R]}{\sqrt{E[R^2] - E[R]^2 \cdot E[D^2] - E[D]^2}} \\ &= 0.447\end{aligned}$$

demonstrating a moderate positive correlation between R and D. This makes sense based on known physical phenomena, but we are not aiming to isolate this correlated fluctuation, so there must be some attempt to normalize the spring discharge measurements to the recent rainfall, hence why the ratio was implemented. Further refinements are made in **Section 6.3**. The original discharge to cumulative rainfall ratio can be seen in Figure 6.

6.3 Log Transform and Percentiles

In order to best visualize the time-series data, we performed a logarithmic transform and scaling procedure to the original spring discharge to the cumulative rainfall ratio. We are concerned with the relative relationship, meaning that logarithmic transforms and scaling do not impact this aside from easier visualization. We apply this formula to each date in the time period,

$$\log\left(\frac{D_i \cdot 100}{R_i}\right)$$

yielding the result seen in Figure 2. Note, because $\log(0)$ is undefined, we assigned all zero-flow days to have a value of -1.5. The lowest non-zero log-transformed value was -1.367 from 2023-06-23. The reason -1.5 was used as opposed to a more incremental value is due to the increasing steepness of the log function, as well as the fact that "zero-flow" days represent a wide variety of subsurface conditions and water levels. For instance, the aquifer could be completely empty, or just barely below a detectable flow–these are two vastly different situations, accounted for by the slight departure from the next lowest value. Notice that within the dataset, we labeled the 10th, 25th, 50th, 75th, and 90th percentile of points.

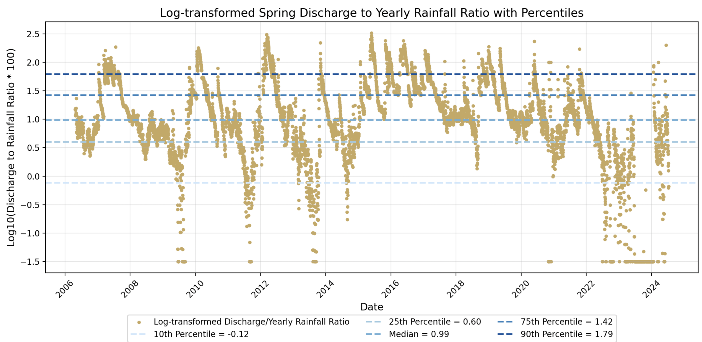

Figure 2: Spring Discharge to Prior Yearly Rainfall Ratio Across Time.

6.4 Poisson MLE Across Percentiles

To visualize temporal shifts in the frequencies of spring discharges of a certain percentile due to anthropogenic influences, we employed the Poisson Distribution. We defined $X_{\text{threshold}}$ to be the number of days in a 365 day (1 year) time period, above a certain threshold, so $X_{\text{threshold}} \sim \mathrm{Poi}(\lambda)$. Assume events occur independently. This was done for all 5 average thresholds or percentiles, (used interchangeably), defined in **Section 6.3**. For each date with **a)** 365 days of prior year rainfall data and **b)** 365 days of spring flow after, (2006-04-26 to 2023-07-11), we estimated the MLE for λ for the next year. For instance, let $c =$ 2007-01-01, $\hat{\lambda}_c$ represents the MLE value for λ for the time period until 2008-01-01, resulting in rolling 1 year timeframes across the dataset. Following the derivation obtained from [14], we can see that the MLE for a Poisson of a particular day c is, where x_c is the number of days above a certain threshold 1 year from that day:

$$\hat{\lambda_c} = \sum_{i=1}^{n} \frac{x_c}{n}.$$

Considering that our time period is one year, this simplifies to allow us to count the number of days above a certain threshold for one year to estimate $\hat{\lambda_c}$.

$$\hat{\lambda_c} = x_c.$$

In addition to daily estimates for $\hat{\lambda_c}$, the average of each percentile is graphed in the same color, alongside daily rainfall totals plotted in tan for context, in Figure 3.

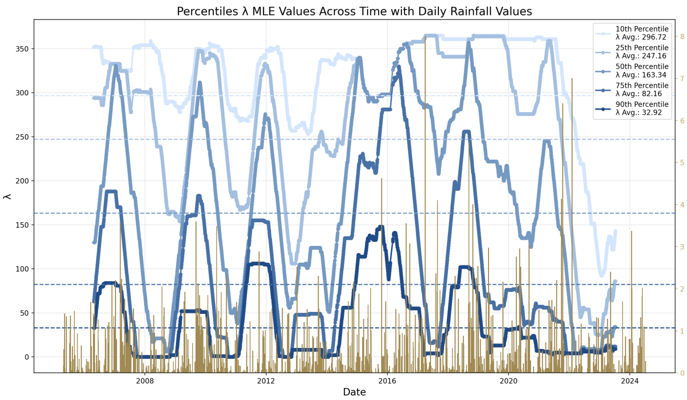

Figure 3: $\hat{\lambda_c}$ for Log data Percentiles Over Time with Daily Rainfall. Dashed line is Time Period Average.

These results demonstrate a significant reduction in spring discharge over time across all percentiles. However, the most pronounced declines come with the two trendlines for days above the 10th and 25th percentiles. Assuming values plateau near the current level, we will have seen decreases of around 325 days per year to 125 days per year, and around 250 days per year to 75 days per year, respectively. By looking at recent data with reference to the relative distribution of data, we can see how recent human actions such as overpumping have diminished springflows rapidly since 2022. Furthermore, daily rainfall data is not extremely skewed towards the first half of dates. One may ask, what is so convincing about this most recent dip? Look at 2011 or 2013, with nearly equivalent declines that eventually rebounded? The 10th and 25th percentile declines compared to their previous lows, which were ~ 250 and ~ 120 before 2021/2022, are magnitudes lower in the last 10 years, demonstrating the severity of the spring flow[3] reduction in addition to its duration.

6.5 Exponential MLE for Zero-flow and High-flow

One of the key concerns regarding Jacobs Well is the increased frequency of zero-flow days. While the article referenced in the **Background** section mentions 6 occurrences of zero-flow [17], you can see the frequency and duration zero-flow days in Figure 4, with a singular thin, tan line representing one zero flow day. We can see that over time, the duration of zero-flow periods has increased, especially since mid-2022. For those who are curious, all of the days have been plotted in this style (Supplementary Figure 8) based on their classification in Supplementary Table 1. The exponential distribution

[3]Spring flow and spring discharge are used interchangeably in this paper, though spring discharge refers to the exact measurement

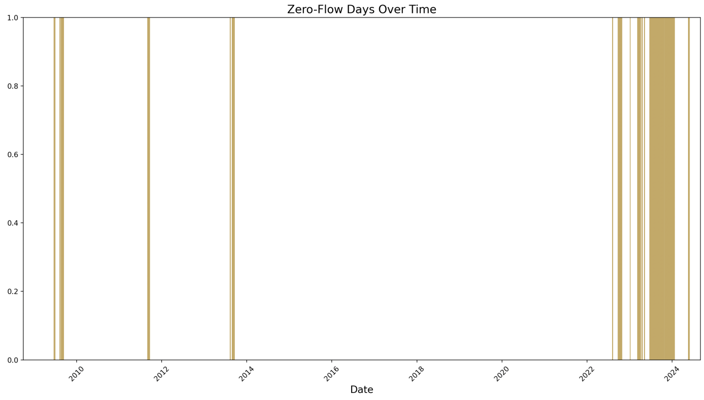

Figure 4: Zero-Flow Days Across Available Data Timespan

is an excellent tool to visualize this phenomena, as the event is now a day with zero-flow, and we are measuring the days between events. We assume each event is independent. Similar to Section 6.4, we calculate the MLE, $\hat{\lambda_c}$, which requires a correction [4] using the 1-year rolling window technique once more. Within this one year time frame, n represents the number of instances of zero-flow days, and the number of days between each instance represents an x_i, where x $\hat{\lambda_c}$, following the derivation obtained from [4] citing [11][13], we can estimate $\hat{\lambda}_{cu}$, the final unbiased estimator, by finding $\hat{\lambda}_{cb}$ the biased estimator. We can find the biased estimator by taking the reciprocal of the average days between our events:

$$\hat{\lambda}_{cb} = \frac{1}{\bar{x}}$$

with n representing the number of gaps, and $\bar{x}$ representing the average duration of a gap between zero-flow days over a one year period:

$$\begin{aligned}\hat{\lambda}_{cu} &= \hat{\lambda}_{cb} - \text{Bias} \\ &= \hat{\lambda}_{cb} - E\left[\hat{\lambda}_{cu} - \hat{\lambda}_{cb}\right] \\ &= \hat{\lambda}_{cb} - \frac{\hat{\lambda}_{cb}}{n-1}\end{aligned}$$

The results of the MLE estimation across time can be seen in Figure 5, using rolling windows of 1 year and 10 years. An identical process was repeated, but the event is days with a spring discharge above 6 cfs (no drought) in Figure 6.

In Figure 5, looking at how λ changes over a one year period does not reflect the longer durations of zero-flow days we see in Figure 4, why is this? The reason for this is because our time window for the exponential distribution is far too small. Compare this to the 10 year window. The smaller gaps in 2009, 2011, and 2013 (see Figure 4) still account for a large share over a one year period, but their frequency in comparison to events since 2022 is of lesser magnitude, which is more accurately reflected over a long timescale. The emergence of a clear trend for high flows is also evident in Figure 7, identifying that groundwater pumping and land use change have contributed not only to more frequent zero-flow days, but less-frequent high flow days–a trend not glaring in the Poisson data (Section 6.4).

7 Supplementary Figures/Notes

PMF Criteria and table obtained from Watershed Association [22]. Note, we only used 1 day of data for the PMF, not 10 day averages.

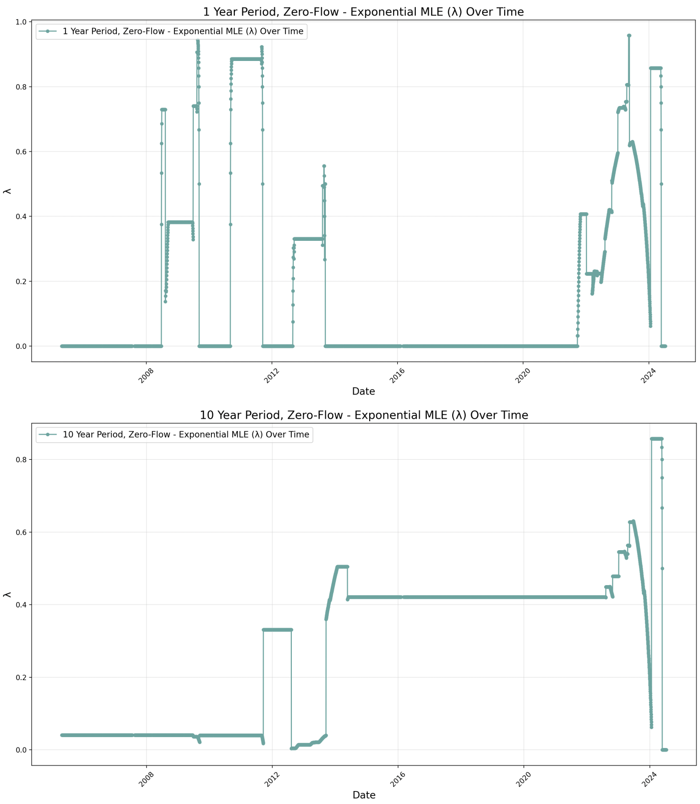

Figure 5: Comparison of MLEs for Zero Flow Days: (Top) 1-Year Rolling Period, (Bottom) 10-Year Rolling Period.

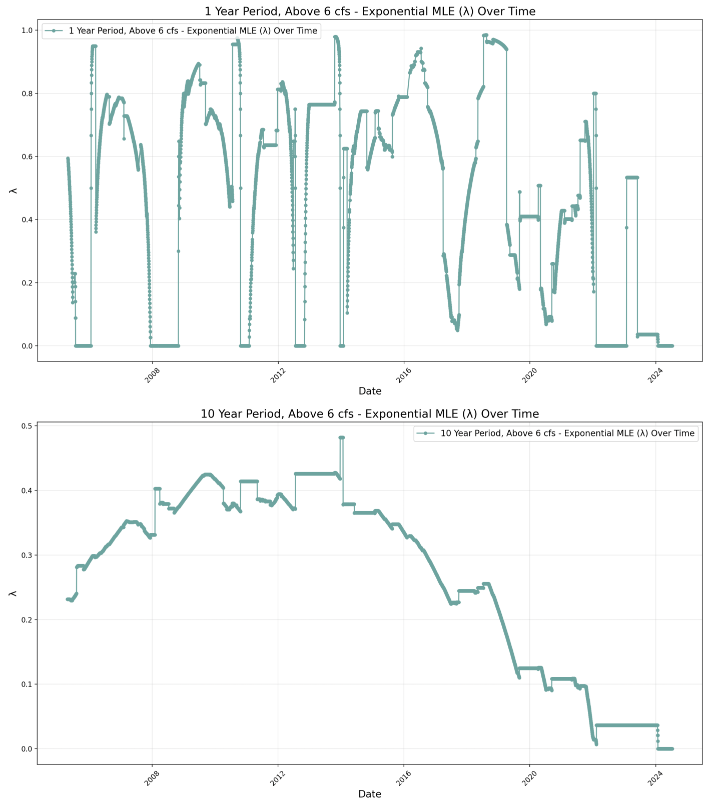

Figure 7: Comparison of MLEs for Above 6 cfs: (Top) 1-Year Rolling Period, (Bottom) 10-Year Rolling Period.

Table 1: Flow Thresholds and Corresponding Graph Colors

Graph Color	**Flow Threshold**	**Explanation**
Blue	> 6 cfs	No drought condition.
Yellow	6 cfs	10% Pumpage Curtailment Level.
Orange	5 cfs	20% Pumpage Curtailment Level.
Red	< 3 cfs	30% Pumpage Curtailment Level.

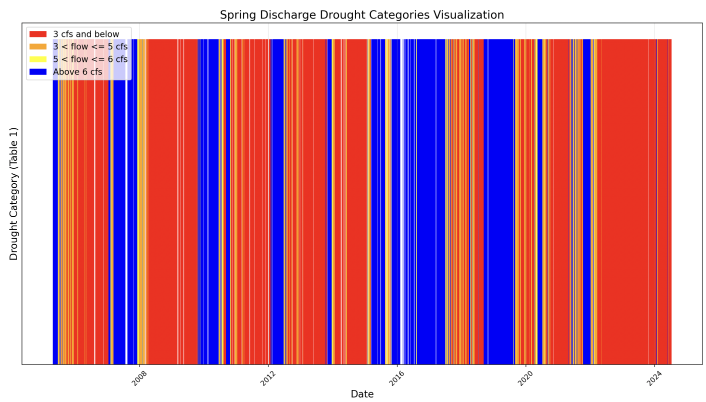

Figure 8: Drought Categories and Their Corresponding Thresholds.

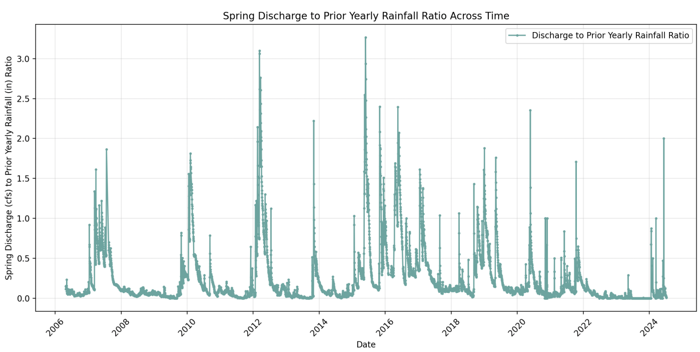

Figure 6: Spring Discharge (cfs) to Prior Yearly Rainfall (in) Ratio Across Time

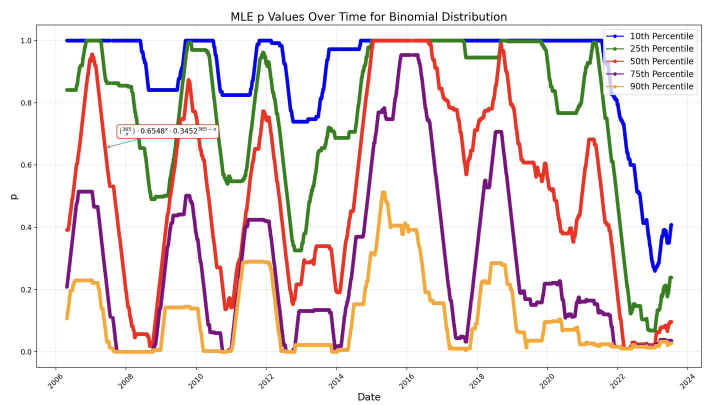

Figure 9: Binomial Distribution Analysis for Spring Discharge Events.

7.1 Use of Generative AI

I leveraged ChatGPT translate big ideas into large code chunks and debug. Sometimes methods were tweaked based on feedback on overlooked details. ChatGPT was also used for LaTex help. I asked ChatGPT to cite itself in BibTex for its help in this project [3].

8 Bibliography

References

[1] Aleksomber. *Creative Commons Attribution-ShareAlike 4.0 International License.* https://creativecommons.org/licenses/by-sa/4.0. Image via Wikimedia Commons, licensed under CC BY-SA 4.0. 2024.

[2] Watershed Association. *A Reminder About the Likelihood of Tragic Flood.* https://watershedassociation.org/a-reminder-about-the-likelihood-of-tragic-flood/. Accessed: December 2, 2024. 2024.

[3] OpenAI's ChatGPT. *Interactive Assistance with Data Visualization and LaTeX Formatting.* Provided comprehensive assistance in Python data visualization, including plotting Poisson distributions and rainfall data with interactive features, as well as LaTeX formatting for mathematical equations and dynamic annotations. 2024.

[4] Wikipedia contributors. *Exponential distribution — Wikipedia, The Free Encyclopedia.* https://en.wikipedia.org/wiki/Exponential_distribution. Accessed: 2024-12-02. 2024.

[5] Emery Coppola Jr et al. "Artificial neural network approach for predicting transient water levels in a multilayered groundwater system under variable state, pumping, and climate conditions". In: *Journal of hydrologic engineering* 8.6 (2003), pp. 348–360.

[6] Edwards Aquifer Website. *Endangered Species in the Edwards Aquifer.* Accessed: 2024-12-01. 2024. URL: https://www.edwardsaquifer.net/species.html.

[7] Edwards Aquifer Website. *The Trinity Aquifer.* Accessed: 2024-12-01. 2024. URL: https://www.edwardsaquifer.net/trinity.html.

[8] Feng ShaoYuan Feng ShaoYuan et al. "Neural networks to simulate regional ground water levels affected by human activities." In: (2008).

[9] Hays County. *Financial Transparency.* Accessed: 2024-12-01. 2024. URL: https://www.hayscountytx.gov/financial-transparency#:~:text=Hays%20County%20topped%20the%20list,53%25%20in%20the%20past%20decade..

[10] Jacob's Well Natural Area. *Informational Infographic Signs at Jacob's Well Natural Area.* Observed in person. Specific date unknown. n.d.

[11] R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis.* Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 2007. ISBN: 9780131877153. URL: https://books.google.com/books?id=gFWcQgAACAAJ.

[12] Eve L Kuniansky and Kelly Q Holligan. *Simulations of flow in the Edwards-Trinity aquifer system and contiguous hydraulically connected units, west-central Texas.* Vol. 93. 4039. US Geological Survey, 1994.

[13] National Institute of Standards and Technology (NIST). *Exponential Distribution.* https://www.itl.nist.gov/div898/handbook/eda/section3/eda3667.htm. Accessed: 2024-12-02. 2024.

[14] Pennsylvania State University. *Lesson 1.5: Overview of Statistical Concepts.* Accessed: 2024-12-02. 2024. URL: https://online.stat.psu.edu/stat504/lesson/1/1.5.

[15] Jacob's Well Exploration Project. *Jacob's Well Exploration Project.* https://www.jacobswellexplorationproject.org/jwep.html. Accessed: December 2, 2024. 2024.

[16] Nicasio Sepúlveda. "Analysis of methods to estimate spring flows in a karst aquifer". In: *Groundwater* 47.3 (2009), pp. 337–349.

[17] KUT Staff. *Jacob's Well swimming hole remains dry as water flow dwindles in Wimberley and Texas Hill Country.* Accessed: 2024-12-02. Aug. 2023. URL: https://www.kut.org/energy-environment/2023-08-02/jacobs-well-swimming-hole-spring-water-flow-wimberley-texas-hill-country.

[18] Texas Legislature. *Texas Civil Practice and Remedies Code Chapter 75: Limitation of Landowners' Liability.* Accessed: 2024-12-01. URL: https://statutes.capitol.texas.gov/Docs/CP/htm/CP.75.HTM.

[19] Texas Monthly. *Jacob's Well Is Dry—And Overpumping Is to Blame.* Accessed: 2024-12-01. 2024. URL: https://www.texasmonthly.com/news-politics/jacobs-well-dry-overpumping-aqua-texas/.

[20] The Meadows Center for Water and the Environment. *Climate Dashboard: Precipitation.* Accessed: 2024-12-01. 2024. URL: https://www.meadowscenter.txst.edu/climatechange/climatedashboard/precipitation.html.

[21] Wimberley Valley Watershed Association. *Jacob's Well Natural Area.* Accessed: 2024-12-01. 2024. URL: https://watershedassociation.org/impactareas/landconservation/jacobs-well-natural-area/.

[22] Wimberley Valley Watershed Association. *Jacob's Well Flow Monitoring.* Accessed: 2024-12-01. 2021. URL: https://watershedassociation.org/2021/02/03/jacobs-well-flow-monitoring/.

[23] Wimberley Valley Watershed Association. *Take Back Our Water.* Accessed: 2024-12-01. 2024. URL: https://watershedassociation.org/take-back-our-water/.

[24] Wimberley View. *Diving into Jacob's Well.* Accessed: 2024-12-01. 2024. URL: https://www.wimberleyview.com/article/508,diving-into-jacob-s-well.