EDA: Analyzing the Relationship Between Snowpack Variation, Precipitation, and Water Levels in the Great Salt Lake

Julia Duffy
04-30-2025
University of Utah
Link to EDA: [FINAL PROJECT - Colab]

Abstract

This project seeks to compare the impact of various factors on water levels in the Great Salt Lake. In this analysis, daily snow water equivalent (SWE) measurements from the mountains and daily precipitation records from Salt Lake County are paired with lake-level data from 01-01-2000 through 01-01-2025 to identify which variable has the strongest effect on lake elevation in the Great Salt Lake. This EDA uses several different analysis techniques, such as a lagged correlation analysis, linear regression analysis, normality test, and Pearson correlation test. It was found that Snowfall in the mountains has a very weak but significant relationship to water levels in the Great Salt Lake. Introducing a time delay of 70 days maximized the correlation between SWE levels and water levels in the lake. Precipitation and SWE seemed to have no immediate impact on water levels in the lake.

Introduction

It is no secret that water levels in the Great Salt Lake have been steadily declining over the past few decades. As the shoreline retreats, more and more of the lakebed has been exposed, leaving fine dust, minerals, and salt on the surface. When the winds kick up, that fine dust is kicked up into the air, and tiny particles drift into nearby towns. This has posed a serious health threat to the general public, and the situation is only worsening. Because the lake collects runoff from across Utah, those dust clouds often carry pesticides, arsenic, selenium, and other toxins as well.

Central Question

The goal of this analysis is to better understand what factors affect water levels in the Great Salt Lake and how long it takes for snow melt in the mountains to replenish the lake.

This analysis asks two central questions:

1. ***Which variable has the stronger relationship to Great Salt Lake elevation—mountain snowpack (SWE) or Salt Lake County precipitation?***

2. ***Is there a clear time lag between increases in snowpack and subsequent rises in the Great Salt Lake?***

Data Sources

Data for this analysis was sourced from the U.S. Department of Agriculture NRCS SNOTEL Data as well as the Great Salt Lake Hydro Mapper (USGS). All three datasets in this report included daily data spanning the years of January 1st, 2000 to January 1st, 2025.

The three datasets used were:

- **Brighton SNOTEL SWE**: The USDA NRCS SNOTEL system provides snow water equivalent (SWE), precipitation, and temperature data from automated stations called “SNOTELS” which are located in mountains all across the state. SWE measures how much water is contained in the snowpack if it were to melt, helping to predict how much water will be available in the coming months. Higher SWE means more available water for reservoirs and groundwater recharge, while lower SWE signals potential shortages.

- **Salt Lake County Precipitation**: Daily rainfall accumulation (inches) from a USDA NRCS SNOTEL station near the lake’s southern shore.

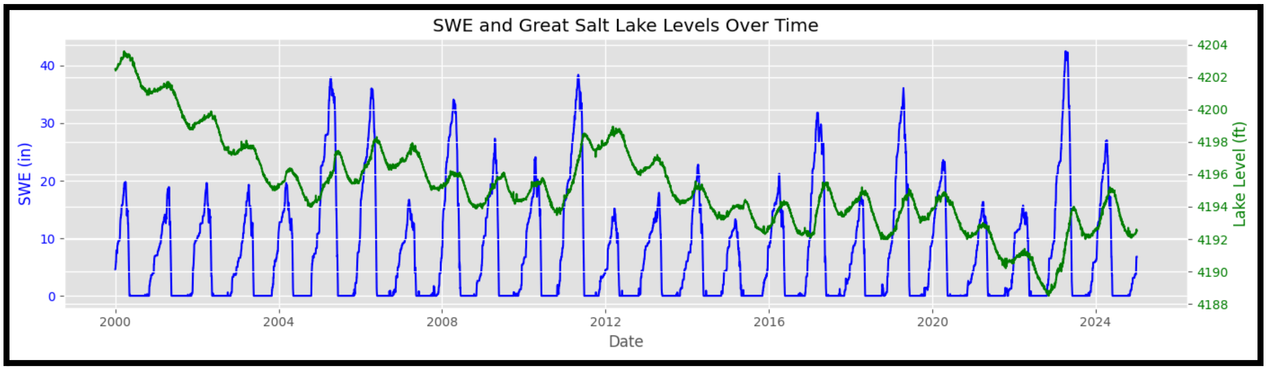
- **Great Salt Lake Elevation**: The Great Salt Lake Hydro Mapper tracks long-term water level fluctuations in the lake. Like the other datasets in the project, this database is convenient as it allows users to download data directly into a .csv file from the site.

Data Retrieval and Cleaning

All three datasets in this experiment were downloaded from their respective organizations directly as .csv files and uploaded to github for easy access. The data cleaning process was fairly simple. No interpolation was necessary as there were no missing values. To combine all three datasets into one dataset, all three datasets had their "Date" column converted to be of datetime type. Then, important rows were renamed to be shorter for easier access and unecessary columns were dropped from each dataset. After these steps, all three datasets were merged on the 'Date' column via an inner join. No missing values remained after this process, and no imputation was necessary.

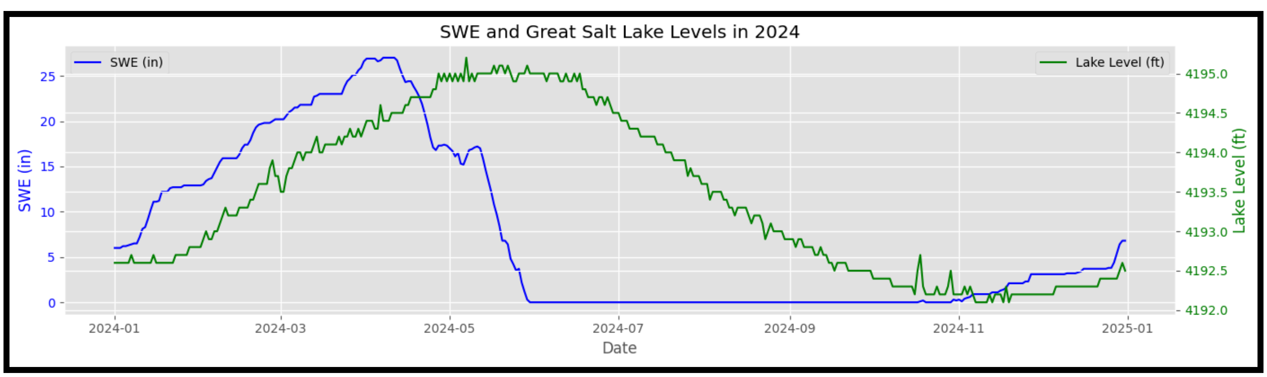
Visualization 1: Full-Period Time Series Line Graph

The goal of this visualization was to see long-term trends and seasonality. The graph shows that there is an obvious pattern in both the water level and SWE data. SWE data tends to peak at the end of each Spring, and water level data tends to peak in the middle of the summer. There is a very obvious decline in water levels over the past 25 years. Also, interestingly, SWE seems to decline at a much more rapid rate than water levels, which exhibit a steadier decline at the end of each cycle.



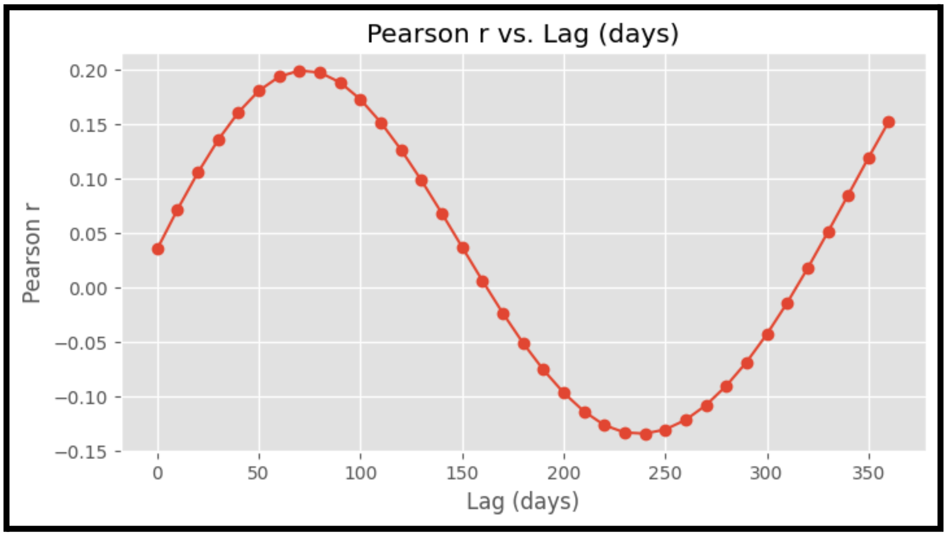
Visualization 2: 2024 Time Series Line Graph

The purpose of this visualization was to quantify the exact lag that maximizes correlation between SWE measurements and water levels in the Great Salt Lake. The intuition behind this graph was to investigate how introducing a time delay changed the relationship between snowpack levels and GSL water levels.



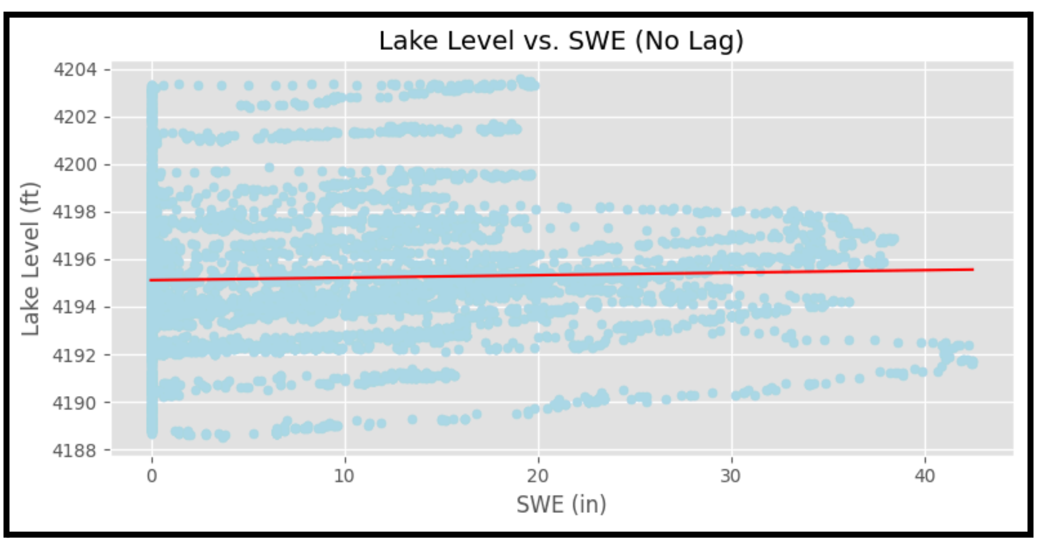
Visualization 3: Lag-Correlation Graph

This graph shows Pearson r values for time delays ranging between 0–360 days, with an increment of 10 days. The highest Pearson correlation values were exhibited at a 70-day delay.



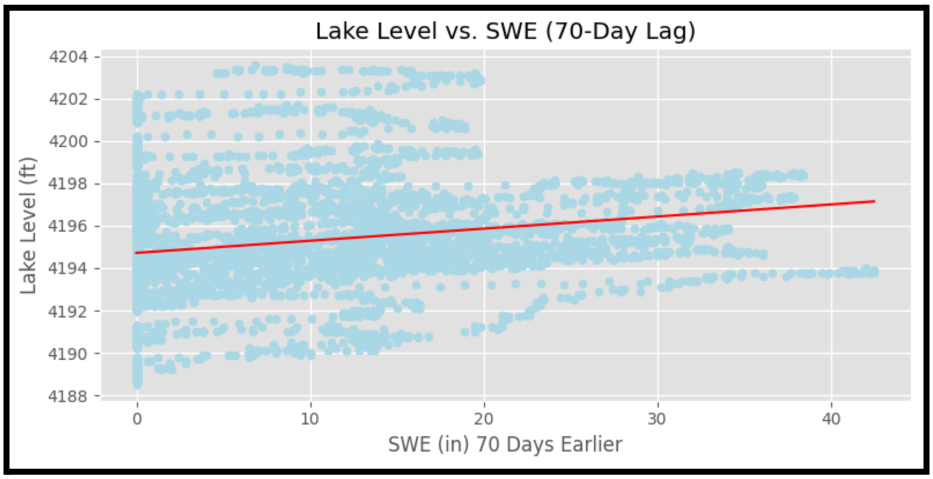
Visualization 4: No-Lag Scatter Plot and Regression Line

The goal of this visualization was to investigate the immediate (no time delay) relationship between snowpack levels and GSL water levels. The visual shows a scatter plot of SWE vs. lake elevation for the same date. The graph includes an OLS fit line that shows the general trend of the data. The cloud of points has a lot of variance, and the slope of the line is very slight and suggests a weak or non-existent immediate correlation.



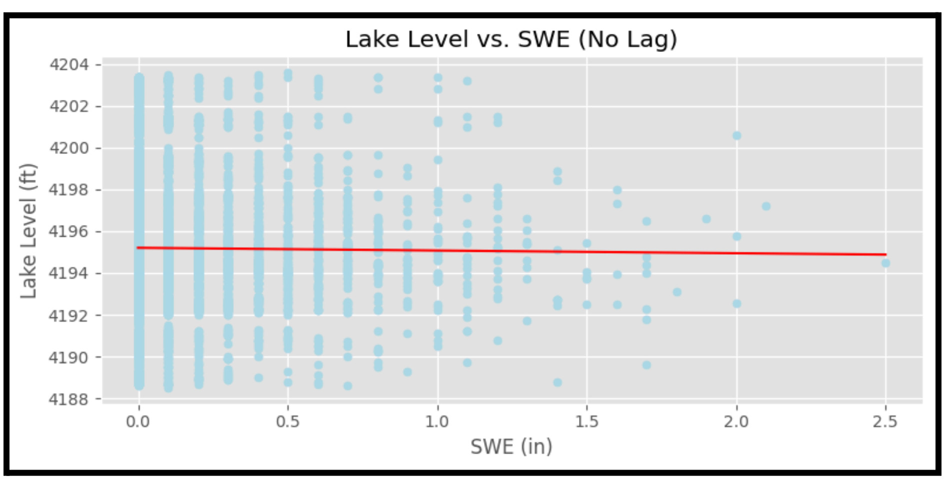
Visualization 5: 70-Day Lag Scatter Plot and Regression Line

The goal was to investigate the relationship between snowpack levels and GSL water levels with a 70-day time delay. The final graph shows scatter plot of SWE shifted 70 days earlier against today's lake elevation with a red fit line. The points seem to cluster more tightly, and the slope is noticeably steeper. This confirms that introducing a 70-day delay increases the correlation between SWE levels and lake levels, though the correlation is still very weak.



Visualization 6 – Precipitation versus Lake Level Scatter Plot and Regression Line

This visualization is similar to Visualization #4 and Visualization #5, but instead of showing a scatter of daily SWE measurements, it shows a scatter of daily precipitation in Salt Lake County and GSL levels. The line of fit is very slightly sloped downward, suggesting against the possibility of a positive correlation between precipitation in Salt Lake County and same-day water levels in the Great Salt Lake.



Statistical Tests

Correlation Testing to Determine Time Delay

```
[18] # Find the highest correlation
     best_r   = max(correlations)
     best_lag = lags[correlations.index(best_r)]
     print(f"Highest correlation: r = {best_r:.3f} at a {best_lag}-day lag")

     Highest correlation: r = 0.199 at a 70-day lag
```

A series of Pearson correlation tests determined which time lag between snowpack and lake elevation yielded the strongest linear relationship. Pearson's method was chosen because the lagged SWE data approximated a normal distribution, and the test provides a straightforward measure of linear association. The highest correlation occurred at a 70-day lag, with an R value of approximately 0.199 and a p value of nearly zero. This indicates a weak but statistically significant effect: snowpack levels have a modest positive influence on Great Salt Lake elevation roughly two months later.

Regression Modeling To Determine Relationships Between Variables

```
SWE vs Lake Level (no lag)
Slope     : 0.01
Intercept: 4195.12
R-squared: 0.001
```

Using today's snowpack (SWE) to predict today's lake level yields a slope of 0.01 and an intercept of 4,195.12 ft. Each additional inch of snow today corresponds to a predicted 0.01-ft rise in the lake, and an R-squared of 0.001 indicates that today's snowpack accounts for almost none of the lake's daily variation. Therefore, there is little significant correlation between these variables.

```
SWE vs Lake Level (70-day lag)
Slope     : 0.06
Intercept: 4194.72
R-squared: 0.040
```

Using snowpack from 70 days earlier produces a slope of 0.06 and an intercept of 4,194.72 ft. In this model, each extra inch of snowpack two months prior predicts a 0.06-ft increase in lake elevation today. An R-squared of 0.04 shows that snowpack from 70 days ago explains about 4 percent of daily fluctuations, which is still quite weak.

```
Precipitation vs Lake Level
Slope    : -0.13
Intercept: 4195.20
R-squared: 0.000
```

For same-day precipitation, the slope is –0.13 with an intercept of 4,195.20 ft. The negative slope lacks physical meaning, and an R-squared effectively equal to zero confirms that today's rainfall provides no predictive value for today's lake level.

Limitations

There were several limitations of this analysis. First, SWE measurements were obtained from only one SNOTEL station (Brighton), limiting the ability to generalize results. Snowmelt from many areas other than Brighton makes its way into the Great Salt Lake. Second, on a similar note, precipitation data were sourced from just one location in Salt Lake County, potentially missing rainfall variability throughout the larger region. It also means that rainfall runoff from areas outside of Salt Lake County was neglected. Third, factors such as evaporation rates, water use, groundwater infiltration, and runoff diversions were not considered, even though more than likely they significantly influence lake water levels. Finally, predictive modeling based purely on historical data assumes stability in climate and water management practices, which may not hold true in the future. Climate change is real, and this study does not account for climate change-related factors such as overall temperature increases in the state.

This study may have been improved if a different reservoir were studied, as the Great Salt Lake is surrounded ban y urban landscape and is used for many purposes. Both of these factors could potentially disrupt the natural order of how snow and rain fall replenish the lake.

Conclusion

The analysis found a statistically significant but weak relationship between mountain snowpack (SWE) and Great Salt Lake elevations. Snowpack measured 70 days earlier produced the highest correlation (r equals 0.199) of all time delays tested, which potentially suggests that snowmelt has a delayed and modest influence on lake levels. Immediate snowpack measurements showed weaker correlations, while local precipitation showed pretty much no relationship with lake elevation. There were many limitations with this study, but it still provided some very interesting results and demonstrated some nuance as to how water enters the Great Salt Lake.

Sources:

https://wcc.sc.egov.usda.gov/nwcc/sensors

https://www.sltrib.com/news/environment/2019/02/16/utah-snowpack-packed-with/

https://www.nrcs.usda.gov/utah/snow-survey

https://water.utah.gov/reservoirlevels/

https://webapps.usgs.gov/gsl/data.html