Trend Analysis of Priest Lake Water Temperature Data

Ephraim Romesberg

The Data

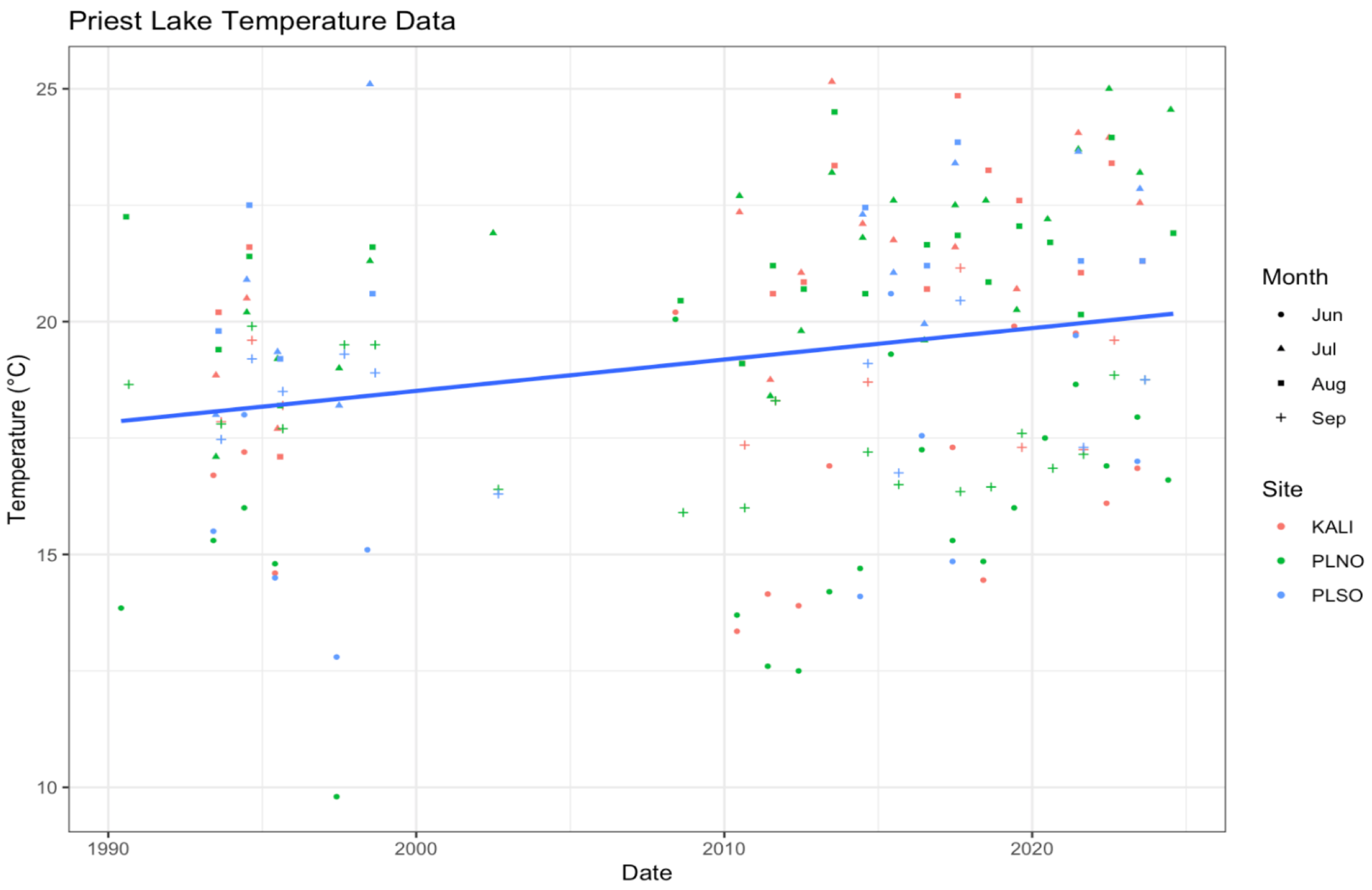
- Data consists of temperature measurements from Priest Lake
- Only Data from top meter is used
- Data is only used from the sites KALI, PLNO, and PLSO and months June, July, August, and September
- Other sites and months contain too little data or have too large of gaps between observations

Is temperature increasing over time?

Two ways to test this:

- Fit a parametric function to model the central tendency (i.e. mean or median) of the data over time:
  - Simplest example of this is linear Regression (i.e. modeling temperature as a linear function of time)
  - If the regression line has a positive slope this is evidence of an increasing trend
- Look at the pairs of observations:
  - For each pair of successive observations y_i, y_j measured at times $i < j$,how many increase over time (i.e. $y_i < y_j$) and how many decrease or stay the same?
  - If more of the pairs increase over time than decrease this is evidence of an increasing trend

Can we use Linear Regression?



Problems With Using Linear Regression:

- Assumes data is normally distributed
- Assumes Relationship between x and y is linear
- Doesn't account for the Seasonal and Regional nature of the data
- Doesn't account for spatial or temporal dependence between observations (i.e. correlation between Y_i and Y_j is assumed to be zero for $i \neq j$)

The Second Approach:

- The approach that looks at pairs of successive observations is known as the Mann-Kendall Test
- This test looks at all possible pairs of sample data y_i, y_j where $i < j$
- For each pair we calculate $sign(y_j - y_i)$, where $sign(y_j - y_i) = 1$ if $y_j > y_i$, 0 if $y_j = y_i$ and -1 if $y_j < y_i$
- We take the average of these values
- This average is a value between -1 and 1 and is known as Kendall's Tau
- If Kendall's Tau is positive this means that more of the pairs increase over time than decrease

The Mann-Kendall Test:

- Null Hypothesis: Observations are equally likely to increase or decrease over time (no trend)
- Alternative Hypothesis: Observations are more likely to increase than decrease over time (positive trend)
- The Mann-Kendall Test tells us if Kendall’s Tau is different enough from zero to conclude there is a trend
- If the null hypothesis is true then (the sample) Kendall’s Tau is approximately normal (for large enough sample sizes) with an expected value of zero
- Hypothesis test is constructed using the approximate normality of Kendall’s Tau
- The approximate normality of Kendall’s Tau holds regardless of whether the data is normal
- Mann-Kendall test makes no assumption that the relationship between x and y is linear
- Only depends on the ranks of the observations

Issues with the Mann-Kendall Test:

Seasonal/Regional effects:

- While the MK test can deal with non-normal and non-linear data it is not designed to deal with seasonal or regional effects
- This can lead to misleading results
- When observations are compared across seasons or regions seasonal and regional differences can be mistaken for trend

Spatial/Temporal dependence:

- MK test also assumes that observations are independent
- This is often an unrealistic assumption
- It is natural for observations close in space and time to have some type of dependence

Dealing with Seasonal/Regional Effects:

- Fortunately there are seasonal and regional versions of the MK test
- Rather than looking at all pairs of sample data we can only look at pairs within the same season and region
- This is known as a Regional-Seasonal Kendall test
- We only compare pairs within each Region-Season combination
- These pairwise comparisons are then put together and averaged like before to calculate Kendall's Tau
- A positive R-S Kendall's Tau tells us that there is an increasing trend in one or more Region-Season combination

Dealing with Spatial and Temporal Dependence:

- R-S Kendall test still assumes that the observations are independent
- Using a test on correlated observations that assumes independence will often lead to misleading results
- A correlated sample provides less information than an independent sample
- We need to adjust for this loss of information when performing our hypothesis test

Dependent Sample=Less Information

- There is a version of the Seasonal Kendall test that is adjusted for serial dependence over time
- For technical details see Hirsch and Slack (1984)
- This test is easily extended to also account for spatial dependence in regional data

Which test is best for our data?

- We have no reason to assume our data is normal or that if there is a trend it is linear
- Our data is from multiple seasons (June, July, August, and September) and Regions (KALI, PLNO, and PLSO)
- More than likely there is spatial and temporal dependence between observations
- This makes a Regional-Seasonal Kendall test adjusted for Spatiotemporal dependence a better choice than a more conventional method such as linear regression

Regional-Seasonal Test Results:

- R-S Kendall test adjusted for Spatiotemporal dependence was performed using `rkt` package in R
- Here regions are sites (KALI, PLNO, and PLSO) and seasons are months (June, July, August, and September)
- Based on this test we got a Tau value of approximately `.27` and p-value of approximately `0.005`
- Using a standard significance level of `.05` we reject the null hypothesis and conclude that there is an increasing trend; later observations are likely to be larger than earlier ones

Trend Within Seasons:

- The R-S Kendall test tells us that there is a trend in at least some of the region-season combinations
- It does not tell us which specific regions and seasons have a trend
- It seems plausible that warmer months could have more of a trend than cooler months
- We can investigate this further by running individual Regional Kendall tests for data from each month

Results From By-Month tests:

- Greatest trend is in July (Tau value of `0.49` and p-value of `0.002`)
- Weakly significant positive trends in June and August
- No significant trend in September

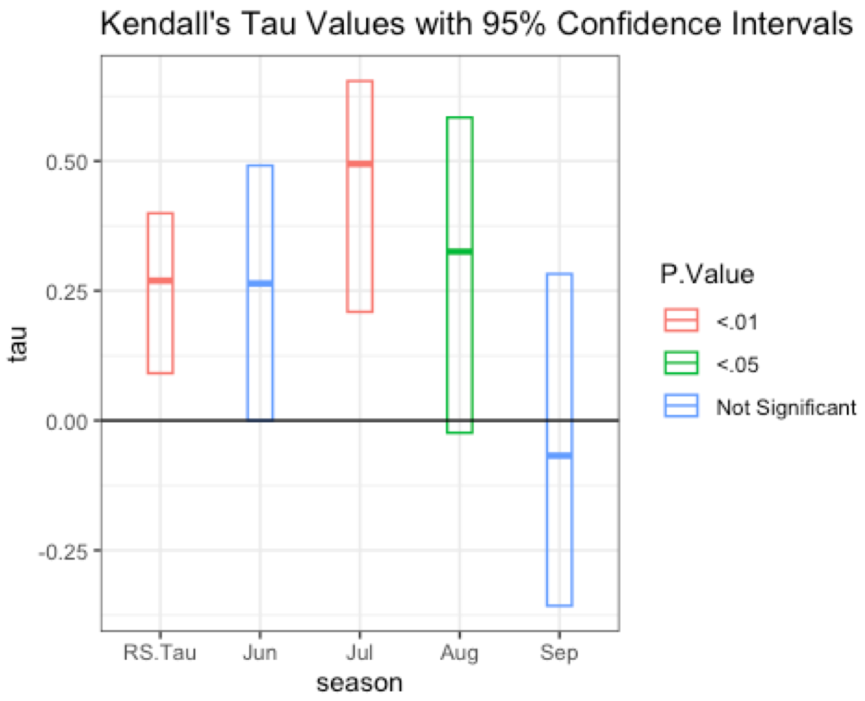
Quantifying the Trend:

- A positive Kendall's Tau tells us that observations are more likely to increase over time
- It does not tell us how much they increase (i.e. °C/Year)
- To estimate the magnitude of the trend we have another statistic known as the Theil-Sen's slope
- Theil-Sen's slope also uses pairs of successive observations
- To calculate Theil-Sen's slope we calculate the slope of the line connecting each pair of observations, we then take the median of these slopes
- This approach can be adjusted for regional/seasonal data just as with Kendall's Tau
- Theil-Sen slope does not assume normality and is much less sensitive to outliers than ordinary regression slope
- It does however assume the trend is linear, so interpret with caution

Theil-Sen Slopes:

- The Regional-Seasonal Theil-Sen Slope is approximately .08
- As a (very rough) approximation we can say that Temperature is increasing at about .08 °C/Year
- Since some months have a more significant trend than others it is better to look at the slopes by month
- We get the largest slope in July at approximately 0.14 °C/Year

Confidence Intervals for Tau:



Confidence Intervals for Theil-Sen’s Slope:

Interpreting the Confidence Intervals:

- The confidence intervals give us a range of plausible values for Tau and TS-Slope values
- If a confidence interval contains zero this indicates that the trend is not significant
- We get the same results from the CI’s as we did from the hypothesis tests
- Overall Regional Seasonal trend is positive
- By season we have the greatest trend and largest slope in July

Conclusion:

- The Regional-Seasonal Kendall Test relies on much less restrictive assumptions than linear regression
- Data does not need to be normal and relationship does not need to be linear
- Test accounts for the regional and seasonal nature of the data and spatial/temporal dependence
- Based on this test we can conclude that the water temperature is increasing over time in Priest Lake
- The most significant trend and largest rate of increase is in July

Refrences:

Design Trend Seasonal Kendall. (n.d.). Vsp.pnnl.gov. https://vsp.pnnl.gov/help/Vsample/Design_Trend_Seasonal_Kendall.htm

Helsel, D. R., & Hirsch, R. M. (2020). Statistical methods in water resources (pp. 342–349). U.S. Geological Survey.

Hirsch, R. M., & Slack, J. R. (1984). A Nonparametric Trend Test for Seasonal Data With Serial Dependence. Water Resources Research, 20(6), 727–732. https://doi.org/10.1029/wr020i006p00727

Mann, H. B. (1945). Nonparametric Tests Against Trend. Econometrica, 13(3), 245. https://doi.org/10.2307/1907187

Marchetto, A. (2024). rkt: Mann-Kendall Test, Seasonal and Regional Kendall Tests. R-Project.org. https://cran.r-project.org/package=rkt