NAME: Matthew Thompson
DATE: 05/09/22
Final Project: Scientific Article
STAT/CS 387

Abstract

Streamflow percentile values within the Colorado River Basin (CRB) may be correlated to large-scale teleconnection patterns like El Niño–Southern Oscillation (ENSO) and sun spots. This interrelation may also extend to a regional scale within the basin. The CRB's streamflow values' tie to teleconnection patterns has yet to be explored. Streamflow percentile and teleconnection data were obtained from the United States Geological Survey (USGS) and analyzed by using autocorrelation plots, time series decomposition, Spearman's correlation and cross correlation coefficients, continuous wavelet transformation (CWT), and wavelet coherence (WTC). The autocorrelation and decomposition hinted at overlaps in seasonal cycles, and the correlation and wavelet analyses confirmed those observations. For a signal period range of 9-15 years, strong correlation between each region of streamflow percentiles and ENSO, northern hemisphere sun spots, and southern hemisphere sun spots was discovered. This period range changes depending on the region, but usually stays within a 3 year period size for each ones. It was also found that specific teleconnection patterns were interrelated strongly or barely at all depending on the region, on top of statistically significant patches of high correlation being found for periods of 16 and 64 months spread inconsistently along the time axis for each region.

Introduction

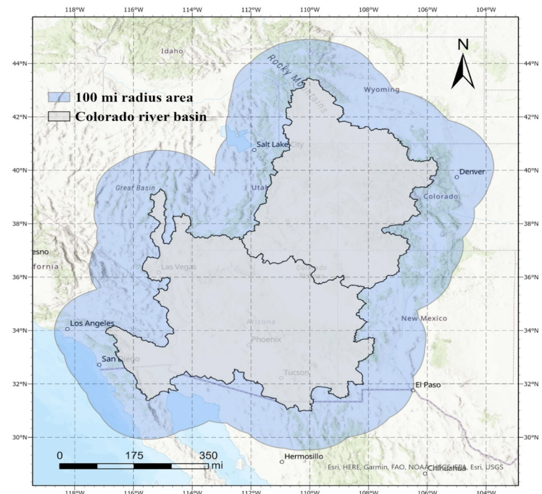
Understanding where the variability of streamflow is derived from is fundamental to forecasting future streamflow values. These could be smaller scale events like weather or droughts, or they could be on a much larger scale like changes in climate. Regardless, the impact that these phenomena have on streamflow is what has allowed the industry to support communities by estimating future hydrological values. In between the scales of general changes in climate and local weather, lies what are known as teleconnection patterns. These are relatively large-scale processes that have been proven to impact other environmental systems around the world, although on a scale that is difficult to observe. The goal of this research is to use time series analysis techniques to explore the relationship that streamflow percentile values from within the Colorado River Basin (CRB) have with these teleconnection patterns.

Streamflow percentile and teleconnection datasets were obtained from the United States Geological Survey (USGS) for aiding in this research.

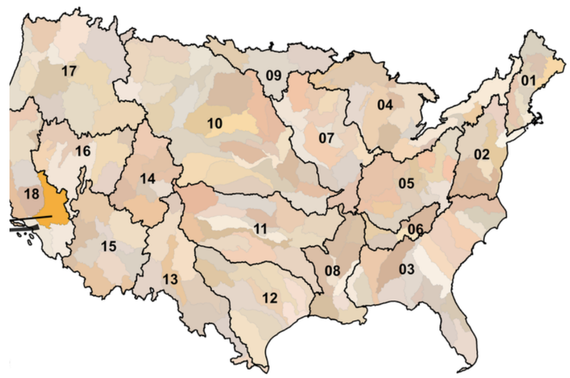
Methods

Study Area

The study area includes the CRB as well as the area within a 100 mile radius of that, as shown in Figure 1a. The Colorado river is about 1,450 miles long flowing from Colorado and Wyoming to the Mexico border. The drainage basin area is about 246,000 square miles and resides in parts of California, Colorado, New Mexico, Nevada, Utah, and Wyoming. According to the "Colorado River Basin Water Supply and Demand Study", the river and its tributaries provide water to nearly 40 million people, both within and outside of the basin, and irrigates nearly 5.5 million acres of agricultural lands [29]. The USGS has divided the United States into hydrological regions, as shown in Figure 1b.



(a)



(b)

Figure 1: 1a Colorado River Basin study area, all gauges located at least within 100 mile radius area. 1b Map of regions relating to USGS 2-digit Hydrological Unit Codes.

Streamflow Data

The streamflow dataset contains daily data from 425 gauges across the CRB, each with differing data collection time ranges. Within these, 70 are considered to have unregulated streamflows. In this case, this means that there are 70 gauges that do not have upstream dams whose operations would be impacting the streamflow values.

These datasets contain multiple columns including the raw streamflow values, as well as variables known as "weibull_jd" and "weibull_site". The "jd" stands for Julian Date, which is the number of days into the current year a day is. The "weibull_jd" feature was constructed by combining the streamflow values of the current day, 3-days before, and 3-days after into a 7-day rolling mean, and then comparing that value to previous values from that same Julian Date for each year of data. This was used to define its percentile using the Weibull plotting position function. The "weibull_site" variable was created in a similar manner, however, the values of the previous year were compared to create the percentile rather than the past values from the same Julian Date. Using the Julian Dates essentially detrends the streamflow percentile data, allowing for a deeper investigation of what makes the data collected by each gauge unique. Due to this, "weibull_jd" was chosen as the primary variable of research and will be referred to within this paper as "streamflow percentile". All streamflow percentile data was downsampled to monthly averaged observations for analysis against teleconnection pattern data.

From all the gauges, six were selected to represent each region within the basin. These were chosen by USGS because they were from drainages of similar sizes, as well as all having unregulated flows. The regions included along with their 2-digit Hydrological Unit Codes and each gauge's site ID are as follows:

- 06224000 = 10U = Upper Missouri
- 07207000 = 11 = Arkansas-White-Red
- 08324000 = 13 = Rio Grande
- 09223000 = 14 = Upper Colorado
- 09404450 = 15 = Lower Colorado

- 13023000 = 17 = Pacific Northwest

A time range of 1980-05-01 to 2016-12-31 was identified as containing overlapping, uninterrupted data for all six.

Teleconnection Pattern Data

Six total teleconnection pattern datasets were explored.

- Atlantic Multidecadal Oscillation (AMO) index
- El Niño–Southern Oscillation (ENSO) 3.4 index
- Pacific Decadal Oscillation (PDO) index
- Pacific-North America (PNA) index
- Number of Sun Spots in the Sun's Northern Hemisphere
- Number of Sun Spots in the Sun's Southern Hemisphere

The AMO index is monthly data represents the variability of sea-surface temperature (SST) within the North Atlantic Ocean. This index has an estimated period, or length of time it takes for oscillation to occur, of about 60-80 years [29]. This dataset was differenced in order to make it stationary.

The ENSO 3.4 index summarizes the periodic change in winds and SST within the tropical Pacific Ocean [4]. "El Niño" is the phase of eastern Pacific SST warming and western Pacific high air surface pressure, and "La Niña" is the phase of eastern Pacific SST cooling and western Pacific low air surface pressure. This oscillation occurs every 3-7 years [19, 22]. This index uses a 5-month running mean with each value representing one month and an ENSO event is defined as when this index crosses the threshold of either -/+ 0.4 .

The PDO index incorporates monthly SST and sea-surface height data to define the phases of the 20-30 year fluctuations within the Pacific Ocean [1, 21].

The PNA monthly index describes the pressure oscillation that occurs over North America. This pattern is extremely variable, so there is no defined period of phase shifting [5].

The daily number of sun spots is divided into the Sun's northern and southern hemispheres. Sun spots are areas about the size of Earth on the surface of the sun where the magnetic field is about 2,500 times as strong as Earth's [20]. Due to this magnetic field strength, the magnetic pressure increases while the atmospheric pressure decreases, resulting in reduced temperatures in that location. These areas have proven to play a role in our planet's atmospheric behavior on a cycle of 11-years [8]. This data was downsampled to monthly average observations in order to analyze along with the other monthly teleconnection pattern data. Only the Northern hemisphere data was analyzed with the large-scale analysis tools (autocorrelation and time series decomposition) because the results across both hemispheres were essentially identical. They were split for the correlation and wavelet analyses.

Each of these teleconnection pattern datasets were sliced to the same time range as the streamflow gauge data of 1980-05-01 to 2016-12-31.

Autocorrelation

The autocorrelation was explored for each streamflow gauge and each teleconnection pattern dataset. This was done to see the way the data correlates with itself at different lengths between observations.

Time Series Decomposition

Each dataset was decomposed into seasonal, trend, and residual curves. Separating these components allows for analysis on the seasonality of time series data independent of its trend. This can be extremely useful for learning the pattern of value fluctuations over time.

Spearman's Correlation Coefficient

Spearman's correlation coefficient was used to explore the multicollinearity of streamflow percentiles across each region, as well as the correlation between streamflow percentiles and each teleconnection pattern variable. Each correlation value is supported by a p-value to prove statistical significance. Insights gained from this analysis can provide a base level understanding of the most prominent teleconnection relationships across the regions.

Spearman's Cross Correlation Coefficient

Spearman's correlation coefficient was also used for cross correlation analysis. This analysis was explored to see if teleconnection pattern data changes correlated with daily streamflow percentiles months in the future. The "lag" defines how many months, and lags of 1 to 12 were used for. Considering the differences in scale between, it is expected that there will be a lag time in streamflow values being effected by these climatic fluctuations. Each correlation value is supported by a p-value to prove statistical significance.

Continuous Wavelet Transformation (CWT)

CWT is a signal analysis technique similar to Fourier analysis, except this technique is built off of wavelets rather sine waves [6]. Sine waves have an infinite length, whereas wavelets are self-contained signals that can be applied over a larger signal at different scales and frequencies for analysis just like Fourier analysis. However, due to the wavelets being a set length it is possible to analyze the signal within a local time space, something that isn't possible with traditional Fourier analysis [18]. This step was done to begin gathering an understanding of how the frequencies within the data of each time series variable vary in time. The lined space around the edge is known as the Cone of Influence (COI). This is where the transformations outputs are biased by not having enough data in all directions, hence why it appears near the edges. This values are worth including, but the bias should be noted [21, 23].

Wavelet Coherence (WTC)

WTC is a signal analysis technique that is essentially another tool to measure correlation between two signals, however, due to it being built off of wavelets it can observe local correlation within the time-frequency space [2, 13]. This can uncover correlations that were hidden to the over-arching perspective of Spearman's correlation coefficient. Statistical significance for this tool's outputs was tested using Monte Carlo randomizations, and can be found on the plots by the regions enclosed by a black polygon. Within those regions there are phase relation arrows. When the arrows point to the right, it means that the two variables are "in phase", meaning their values move in the same direction. This means the opposite for when arrows point to the left. If the arrows point right-down or left-up this means that the first variable is leading the change, if they point right-up or left-down this means that the second variable is leading the change [25, 27, 28]. In the case of this paper, streamflow percentile is always the first variable.

Results

Autocorrelation

Both a 3-year and 30-year autocorrelation analysis was run for each streamflow guage. For the 3-year, each gauge's values dropped very quickly during the first few lags. Afterwards, Upper Missouri and Arkansas-White-Red continued to drop, reaching near 0 by a lag of 100. Upper Colorado and the Pacific Northwest gauges shifted to dropping at a milder pace, reaching near 0 at a lag of 400. The outlier regions were Lower Colorado, whose autocorrelation decreased at a steady pace from start to finish, not reaching 0 until a lag of 700, and Rio Grande. The Rio Grande gauge decreased at a similar rate as Lower Colorado, however, it began increasing again between the lags of 400 and 600. The autocorrelation for it never dropped below 0.2, even after a lag of 1000. For the 30-year analysis, Upper Missouri and Arkansas-White-Red had no cycle, Upper Colorado and the Pacific Northwest gauges had a weak 14-year cycle, Lower Colorado seemed to be on a variable amplitude 6-year cycle, and Rio Grande displayed very weak evidence of the potential existence of a 40-year cycle.

The autocorrelation analysis for teleconnection patterns was run at 5-year and 30 year lengths in order to look for smaller and larger scale fluctuations. The AMO displayed a weak annual cycle of about 0.2, however, displayed evidence of an estimated 60 year fluctuation within the 30-year analysis. ENSO showed a similar, but stronger annual cycle of 0.4, but the interannual oscillation was relatively weak and appeared about every 4 years. PDO did not show any autocorrelation seasonality at all in the smaller scale, it was a close to linear decrease as the lag increased. There was also a weak 10-year cycle observed. The PNA had an extremely weak yearly cycle, but it is likely small enough to be negligible, as well as no interannual cycle. The sun spot data showed no fluctuation at an annual scale, but when increased to interannual displayed a very strong 11-year season.

Upper Missouri and Arkansas-White-Red dropping as quickly as they did could be due to a super variable streamflow climate, because the data consistently held no correlation to itself more than 100 days out. The Upper Colorado and the Pacific Northwest gauges represent a more consistent hydrological climate. Lower Colorado and Rio Grande's extreme autocorrelation may have to do with reduced precipitation when compared to other regions, making for fewer large shifts in streamflow at the daily scale. At a larger scale, it is super interesting to see clusters appear of regions that behave similarly when it comes to streamflow autocorrelation. Upper Colorado and the Pacific Northwest gauges provide potential evidence of sun spot impacts or perhaps a lower period teleconnection pattern. Lower Colorado's 6-year cycle points towards ENSO impacts.

Time Series Decomposition

The seasonal decomposition plots with a period of 10-years displays a overall seasonality of about 10-years for each gauge. This fluctuation is very noisy for the Upper Missouri and Arkansas-White-Red gauges. The trend plots with a period of 5-years display that each region experienced the same drop in streamflow percentile values from about 1995-2003, as shown in Figure 2. Outside of this, clusters of trend shapes can be found. Upper Missouri and Arkansas-White-Red have similar trend shapes to Upper Colorado and the Pacific Northwest, however, with a lower amplitude. Lower Colorado and Rio Grande are the outlier gauges again, with their separating behavior being between 1980-1990.

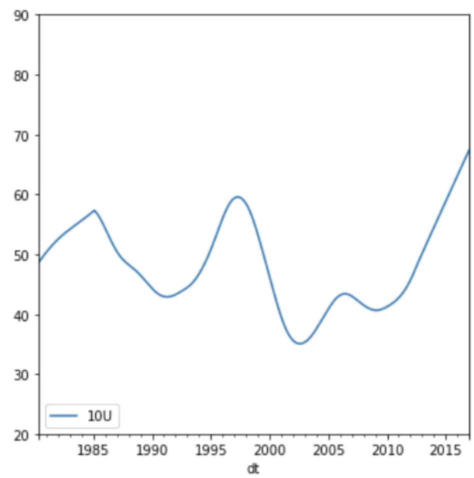
The 10-year seasonality displayed by this decomposition lines up well with the sun spots 11-year cycle. This is a spot to look further into. What is also interesting about these trend plots is that each region seems to be at a pretty low value for their streamflow percentiles, except for Upper Missouri, and yet, the amplitude of the seasonal plots seems to be increasing for all regions. This can be interpreted as overall lower streamflow percentiles, with larger value swings at smaller time scales.

For the AMO and for each of the following teleconnection patterns, a 10-year period was used for the seasonal plots and a 5-year period was used for the trend. These were chosen because they provided the most clear representation of the changes within the datasets. The seasonal plot displayed a very consistent seasonal change of about 5-years. The trend data dropped from 1940 to 1970, and has been rising since.

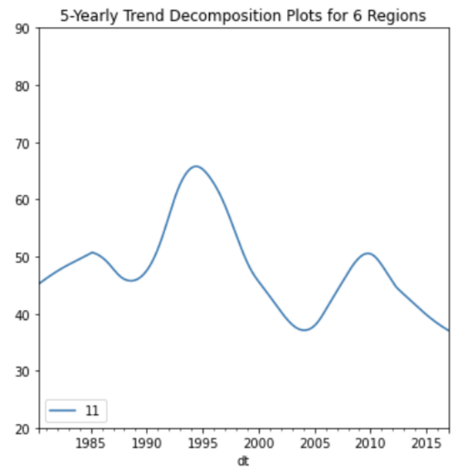
The seasonal data for ENSO from the decomposed time series shows a clear seasonal fluctuation about every 5 years. The trend data seemed to be slowly increasing as time went on, however, it took an interesting dive in the years 2019 and 2020.

The seasonal component of the decomposed PDO time series does not contain a clear seasonal change for a period of 1-year, but when that is extended to 5 years the about 5-year seasonal changes this dataset holds become clear. When looking at the trend section it is also easy to see that the values have been sharply trending downwards for the past 25 years.

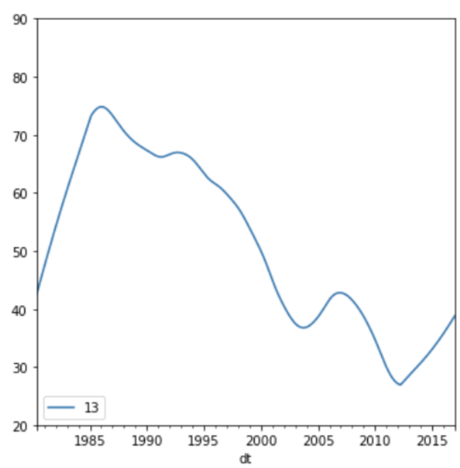
For PNA, the seasonal trend for the decomposed time series shows a slowly decreasing variance from 1950 to about 1995, and since then the variance has been slowly increasing. The trend displays a large dip around 1970, and since then the values have risen sharply and then stayed consistent.



(a)

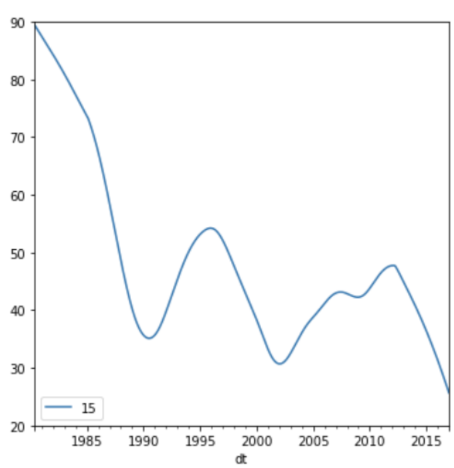


(b)



(c)

(d)



(e)

(f)

Figure 2: Time series decomposition trend plots for each region are shown. The plot for Rio Grande (13) and Lower Colorado (15) are clear outliers, however, that being mostly dependent on the behavior from 1980 to 1990.

The seasonal plot for sun spot data displays a 10-year fluctuation the begins suddenly and tapers until the next jump around 10 years later. The trend plots display this explosion and tapering as well, also showing the 10-year seasonal change.

The seasonal and trend plots for ENSO and PDO look very similar, which makes sense considering they are both indices from the Pacific Ocean. The recurrence of the about 10-year fluctuations within the sun spots dataset is solidifying the goal of investigating this relationship to streamflow. The 5-year fluctuations seen in the AMO, ENSO, and PDO, may be errors, it seems odd that they each look quite similar.

Spearman's Correlation Coefficient

The multicollinearity exploration using Spearman's correlation coefficient displayed values equal to or above 0.5 for 3 pairs of regions. They can be observed in Figure 3a. The highest was Upper Colorado and the Pacific Northwest at 0.82, a pair that has appeared many times during this analysis. The second highest was 0.51 for Upper Missouri and the Pacific Northwest, and the third highest was 0.50 for Upper Missouri and Upper Colorado. Each of these relationships were statistically significant by at least 25 orders of magnitude.

This connection that Upper Missouri has to the clear pair of Upper Colorado and the Pacific Northwest is interesting because this is the first tool to pick up on it. Strangely enough, the other clear pair from previous techniques was Upper Missouri and Arkansas-White-Red, which displayed a correlation of 0.13 and was also statistically significant.

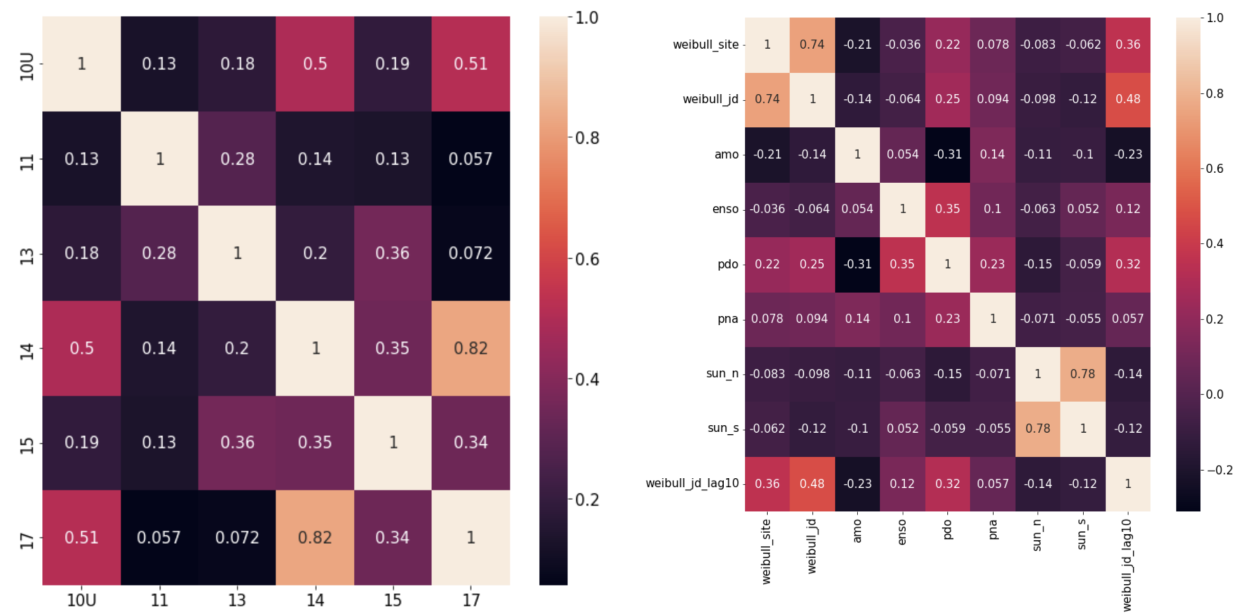
The correlation to the streamflow percentiles for each teleconnection pattern was low. However, there were a few interesting outliers. All the regions had correlation values with the PDO around 0.15, but Rio Grande had twice that at 0.31. Rio Grande displayed a large negative correlation of -0.36 to AMO as well. Also, almost all the regions had extremely low correlations to the ENSO at around 0.01, but Arkansas-White-Red had a significantly larger value of 0.17. For sun spots, most regions stayed around -0.15 for both northern and southern, but the southern sun spot correlation jumped to -0.2 and -0.27 while the northern sun spots were only -0.14 and -0.15 for Upper Colorado and Pacific Northwest respectively. Each of the identified relationships are statistically significant.

It is clear that there are region-specific streamflow percentile correlations to teleconnection patterns, however, they do seem to appear sporadically. There is promise in region hydrological behavior clustering given the number of times the same pairs emerge from analysis.

Spearman's Cross Correlation Coefficient

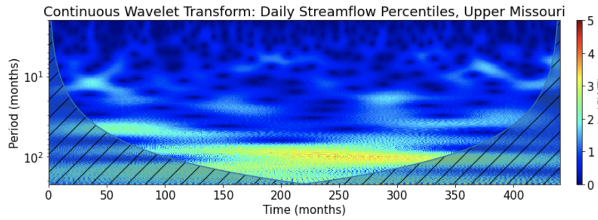
The streamflow percentile correlations to teleconnection pattern data most of the time showed the values for almost all of the variables decreasing as lag increased. Only the statistically significant relationships will be discussed. Upper Missouri's streamflow data correlation contained no outlier relationships and none of the variable pairs proved statistically significant. For Arkansas-White-Red, the value for AMO slowly decreased from -0.1 to a peak of -0.31 at the lags of 7 and 8 months. The gauge for Rio Grande had the value for AMO slowly decrease from -0.36 to reaching -0.50 at lags 9 and 10, while PDO stayed consistent as lags increased. Upper Colorado's values had PDO, northern sun spots, and southern sun spots all stay consistent, although PDO and southern sun spots only barely significant. The streamflow percentile data correlations for In Figure 3b, Lower Colorado showed that at a lag of 10 months AMO decreased to -0.23 from -0.13, ENSO increased to 0.12 from -0.06, PDO increased to 0.32 from 0.25, and northern sun spots decreased from -0.1 to -0.14. The ENSO and northern sun spots were only barely statistically significant. The Pacific Northwest gauge correlations showed an ENSO decrease from -0.01 to -0.14, a northern sun spot decrease from -0.15 to -0.25, and southern sun spot's value staying consistent. ENSO was again only barely statistically significant.

Given enough time, AMO has significant correlation to the gauges in the southern eastern part of the CRB given the correct time period. Stronger ENSO correlations only appeared for Lower Colorado and Pacific Northwest and only during certain lags. This may be due to streamflow percentile changes from snowmelt because ENSO is well known to impact snowfall [10, 19].

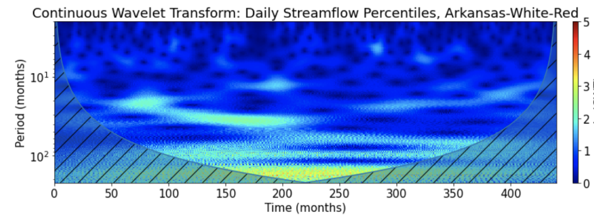


(a) (b)

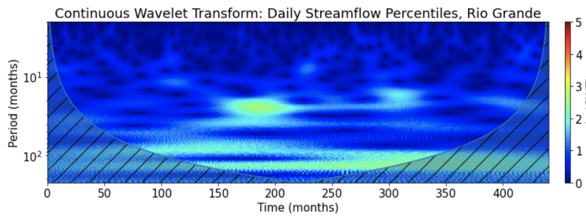
Figure 3: 3a Shows a correlation heatmap for checking multicollinearity. Upper Colorado (14) and Pacific Northwest (17) are strongly correlated, with Upper Missouri's (10U) 0.5 correlation to each of those regions also worth noting. 3b Displays the cross correlation heatmap for Lower Colorado at a lag of 10 months. It is clear here that the interrelations to AMO, ENSO, PDO, and northern sun spots have all strengthened when streamflow percentile values 10 months in the future were considered.



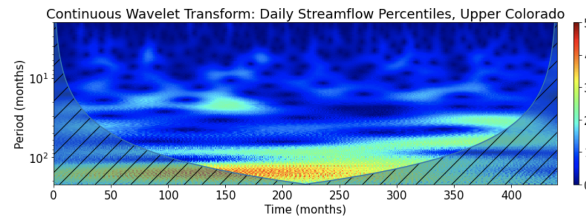
(a)



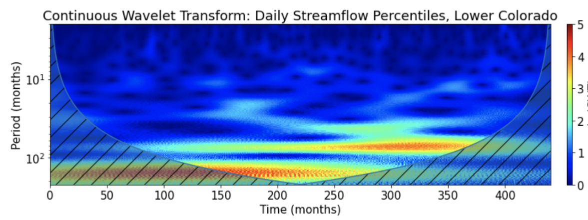
(b)



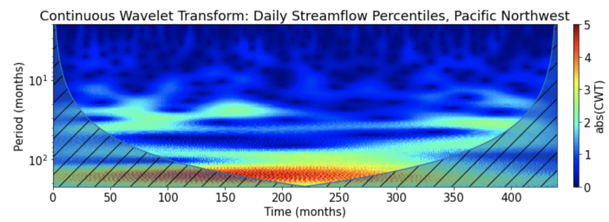
(c)



(d)



(e)



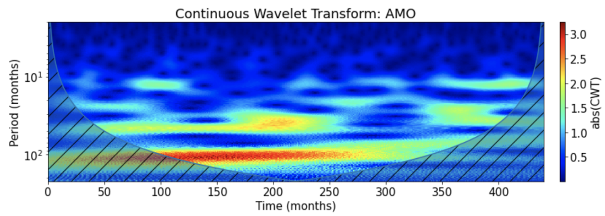
(f)

Figure 4: Continuous wavelet transformations for each streamflow percentile region. Upper Colorado, Lower Colorado, and Pacific Northwest gauges show clear strong frequencies in the band near a period value of 128. A band near 90 months is also visible in Upper Missouri and Lower Colorado.

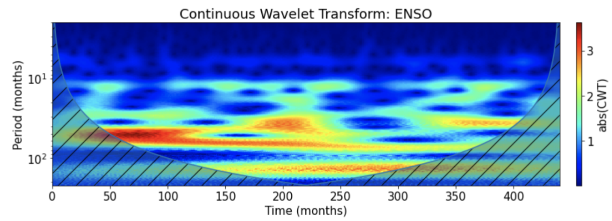
The correlation to AMO seems to be either decisively correlated or not at all, like there is a clear threshold of which regions are impacted and which are not. PNA and ENSO correlation values to streamflow percentiles show large variations in value as the lag increases. This can be interpreted as these teleconnection pattern correlations being very time dependent or maybe the correlation itself comes in bursts of short periods. The values for PDO and northern and southern sun spots remain fairly consistent. The correlation between them and streamflow does not change much as the lag changes. The Northern and southern sun spot datasets display interesting relationships to some regions where the northern sun spots could be correlated and the Southern would be not at all, and vice versa.

Continuous Wavelet Transformation (CWT)

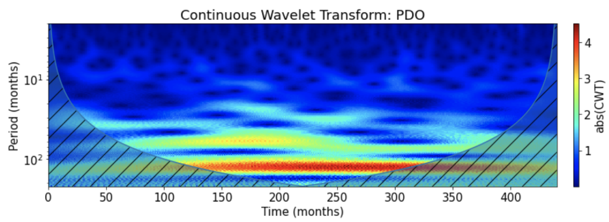
The CWT output for streamflow percentiles shows decent frequency strength at a period of around 10 months and a time scale of 200 to 300 months for Upper Missouri (**??**). For Arkansas-White-Red, it shows a mild strength at a period of about 160 months from 200 to 250 months on the time scale (4b). The CWT to Rio Grande has a couple weak frequencies at a period of around 120 months for the whole time scale (4c). For Upper Colorado, the CWT shows a strong frequency band at a period of around 110-200 months for the time scale up until about 350 months (4d). The Lower Colorado plot displays even stronger frequency strength, for the same plot location as Upper Colorado, but also includes a new band from periods of 80-90 that begins around 100 months and continues to the end of the plot (4e). The CWT for Pacific Northwest shows the strongest frequencies yet, but only for the band for period 110-220 and with the most intense part being from 50 to 300 months on the time scale (4f).



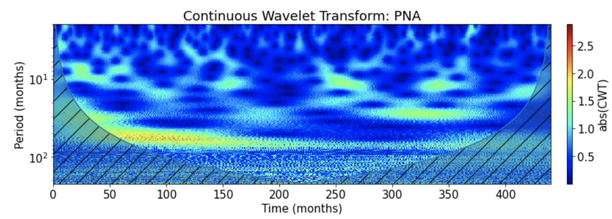
(a)



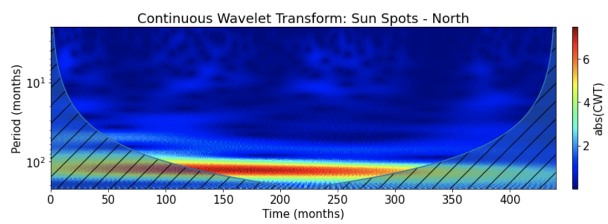
(b)



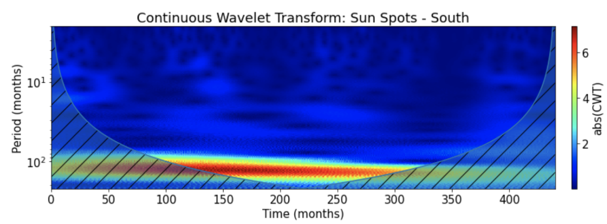
(c)



(d)



(e)



(f)

Figure 5: Continuous wavelet transformation plots for each teleconnection pattern. The strong band at a period of around 120 months appears for all patterns except PNA.

The CWT plot for AMO shows a strong band from periods 100-180 for time scale 50 to 275, but also a smaller band from periods 50-80 with its strongest frequency at 150-250 for time scale (5a). For ENSO, there is a strong frequency band from periods 50-90 for the whole time scale, with some gaps in between (5b). The CWT for PDO shows a super strong frequency within that same 100-160 period band for the whole time scale (5c). The PNA plot shows a very weak frequency for the period range 70-90, within the strongest portion existing from 50 to 150 months on the time scale (5d). The northern and southern sun spot CWT output are almost identical. Each of them have extremely strong frequency bands from the period range of 90-200, with it being the most intense from 50 to 300 months on the time scale(5e, 5e).

There is clear overlap between 100 and 200 month periods for Upper Colorado, Lower Colorado, and Pacific Northwest, as well as the AMO, PDO, northern sun spots, and southern sun spots. Each CWT frequency band appears at different time scales, but a frequency with that period is appearing within each of those. ENSO appear a bit as well, but it is less defined within that period range. The mild pairing between Upper Missouri and Arkansas-White-Red can be observed again within these plots. Interestingly enough, ENSO and PNA share the same frequency hotspot. The portion of ENSO from 70-90 on the period axis and 250-350 months on the time axis seems to appear on the Lower Colorado CWT, and very mildly on Upper Colorado and Pacific Northwest's.

Wavelet Coherence (WTC)

The WTC for Upper Missouri shows a correlation band for 128 to 200 months on the period axis and for the entire time range for every teleconnection pattern. AMO and PNA are not as strong as the others, and ENSO is less precise on the period axis and is only strongest from 50 to 300 months on the time axis, but regardless that band is there. There is also a period range from 32-64 months for the first 150 months on the time scale with strong interrelation for AMO, PNA, ENSO, and northern sun spots. AMO, ENSO, and northern sun spots also show strong correlation at the same period range but for 350-440 months on the time scale. The band around 16 months on the period scale contains statistically significant but inconsistent patches of high strength interrelation for all of the teleconnection patterns.

The WTC for Arkansas-White-Red is significantly less consistent for each teleconnection pattern. The band for 128-200 months on the period axis only weakly appears within the AMO, ENSO, northern sun spots, and southern sun spots plots. What is interesting is that a small patch of correlation within period range 32-128 and time range 300-440 months appears strongly within each WTC plot. A similar patch appears within the same period range but for the time ranges 0 to 75 months for ENSO, PNA, and PDO, and it appears from 0 to 250 months on the time scale in the AMO plot.

For Rio Grande, the WTC plots for ENSO, PDO, PNA, northern sun spots, and southern sun spots all have that band from 128-200 months on the period scale that stretches the whole times frame. For ENSO, PNA, and northern sun spots there is also a strong interrelation from period range 50-64 and within time range 0 to 200. This can be found in Figure 6.

The WTC for Upper Colorado shows the same band of correlation for a period of 128-200 months, however, PDO and ENSO have that interrelation extending into a period of 64 after the 200 month mark on the time axis. There is a strong pocket of correlation across AMO, ENSO, and PDO from periods 32-50 within time range 300 to 440 months. This patch extends almost the entire time scale for PDO.

For Lower Colorado, the WTC plots show AMO, ENSO, northern sun spots, and southern sun spots all containing the interrelation band from 128 to 200 months, although it is only strong within the sun spots plots. There is a strong patch of correlation within the period range 64-90 for time range 200 to 440 months for every teleconnection pattern, although it is a bit weaker for southern sun spots.

In Figure 7, the WTC for Pacific Northwest shows the same band of interrelation for periods of 128 to 200 months for every teleconnection pattern, although this time the band seems to be significantly stronger from 128 to about 160 months on the period axis. A path of high correlation appears for the period of about 64 for time range 250 to 400 months for each variable except southern sun spots.

Most plots displayed the same correlation along the period of 10 to 15 years, however, it was consistently stronger in the ENSO, northern sun spot, and southern sun spot WTC plots. AMO, PDO, and PNA sporadically included this strong band, but it was extremely region dependent. Almost every plot included

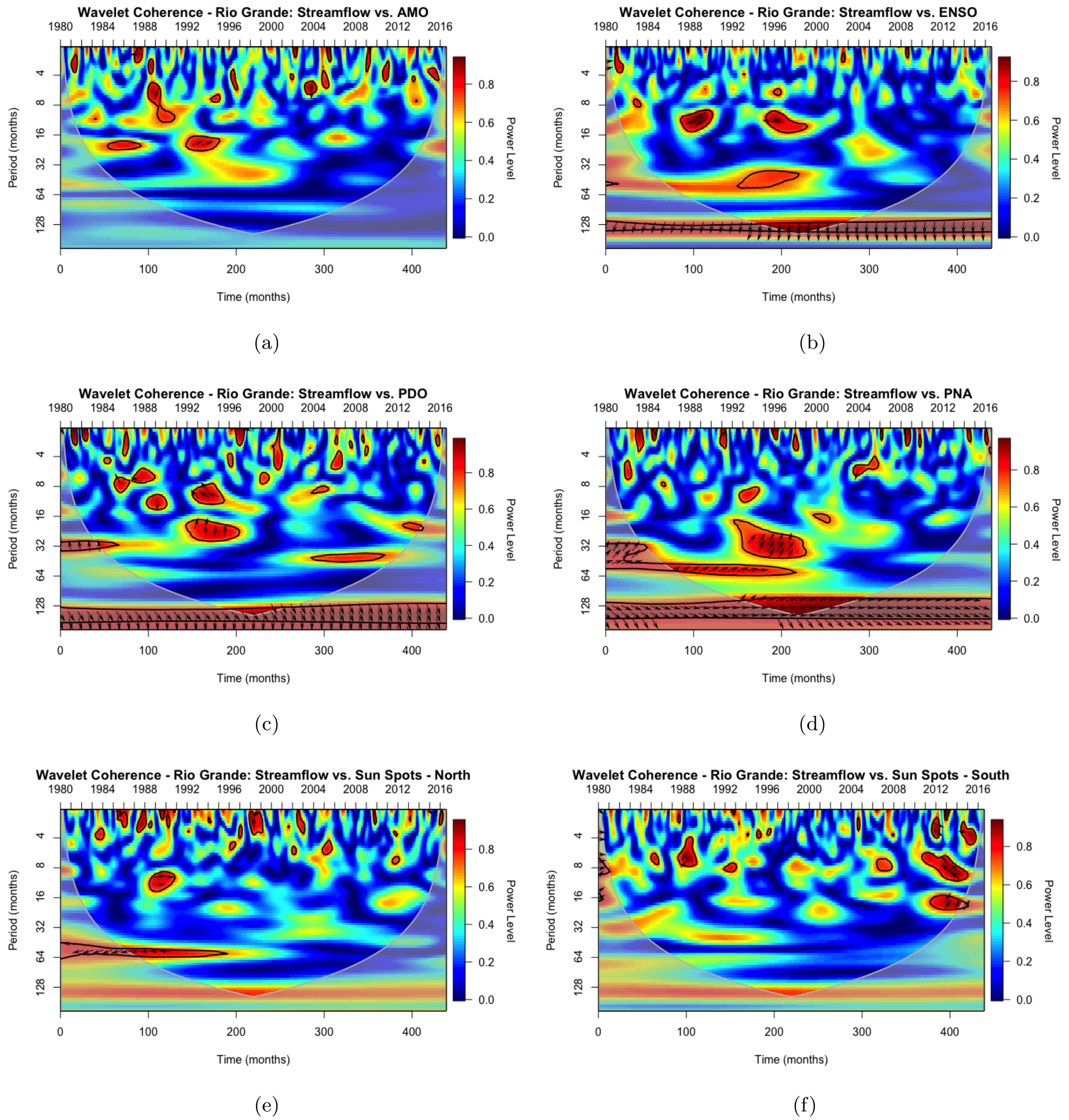


Figure 6: Wavelet Coherence plots between the Rio Grande region streamflow percentile data and all teleconnection pattern data. This region displayed the most intense correlations, yet oddly AMO showed almost no correlation at at in th expected period range of 128-220 months.

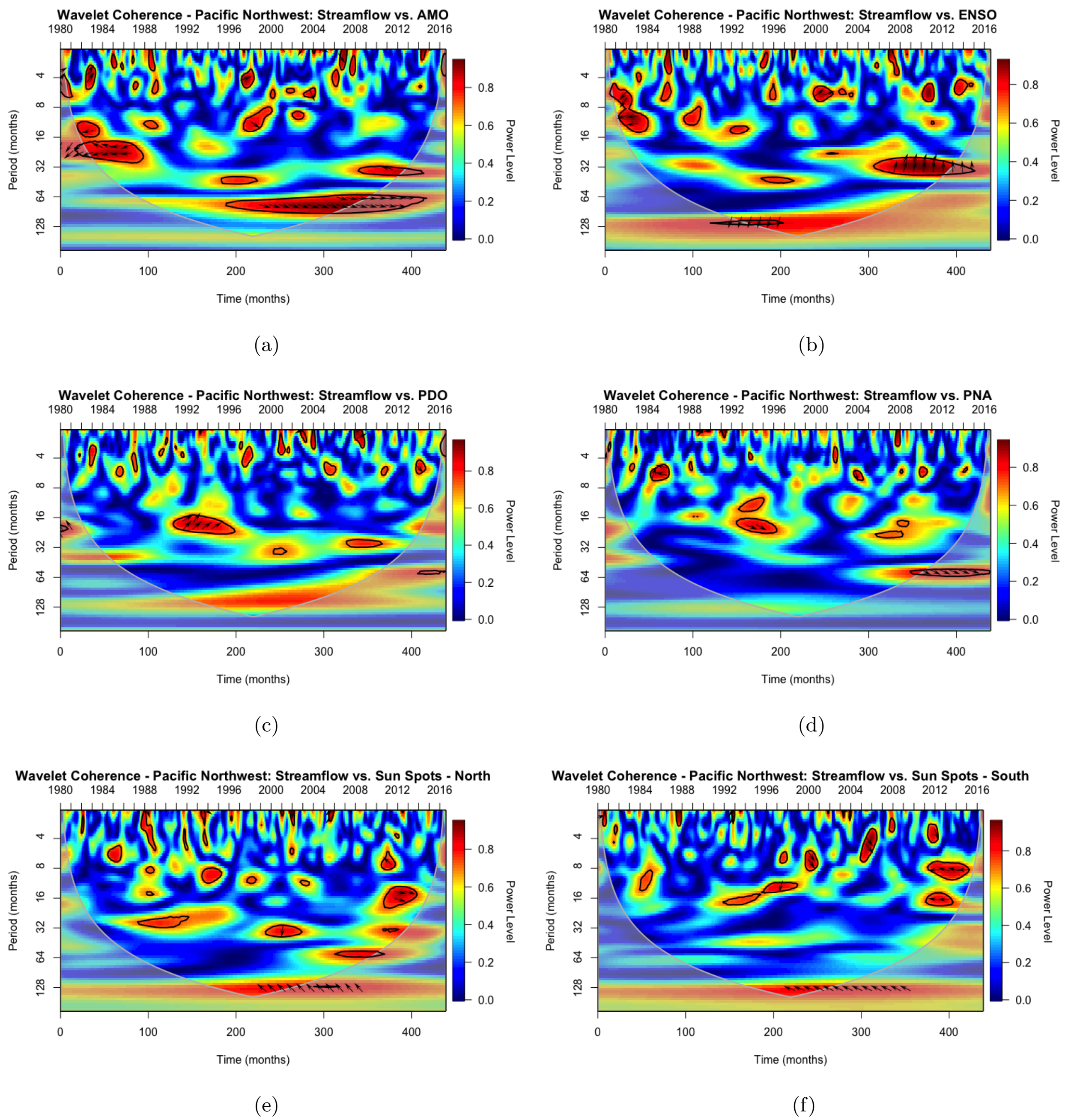


Figure 7: Wavelet Coherence plots between the Pacific Northwest region streamflow percentile data and all teleconnection pattern data. This region is geographically the furthest from the Rio Grande gauge (6), and it displays a very different collection WTC plots. Obvious differences are the lacking correlation to PNA and the very strong interrelation to AMO at a period of 64.

statistically significant patches of correlation in the space around periods of 16, however, they were always inconsistent on the time scale. This same phenomenon appeared around periods of 64 too, but those patches typically extended further along the time axis. What was extremely interesting was how strongly the Rio Grande region was correlated with every feature except AMO. This may be due to the dryness of this region.

Discussion

The most obvious takeaway from this research is the prevalence of the band of correlation for periods of 10 to 15 years for ENSO, northern sun spots, and southern sun spots for each region. This periodicity appeared in the autocorrelation, decomposition, CWT, and WTC analyses, which points a very strong finger towards the interrelation between streamflow percentiles and teleconnection patterns. This likely means that these three patterns are directly impacting streamflow within the CRB, those impacts are just on a very large scale and likely non-linear which explains why traditional correlation methods do not pick them up.

The regions have been shown to correlate to teleconnection patterns very differently hydrologically. Starting from the same band that appears in the period range of 10 to 15 years, this range changes slightly depending on the region. For Upper Colorado for example, the range is about 11-15 years, whereas for Upper Missouri it is more like 9-13 years.

It was shown the each region displayed statistically significant patches of interrelation along the periods of 16 and 64 months inconsistently spread along the time scale for most teleconnection patterns. This helps differentiate the way each region relates to these patterns.

On top of this, each region also correlated to teleconnection patterns different as lag was introduced. Lower Colorado is a great example of how much differently ENSO behaved as the lag increased when compared to other regions. In others, ENSO interrelation would typically decrease, however, in Lower Colorado it instead increased by 0.13 by the time a lag of 10 months was reached.

There is also the case of which patterns do not correlate well with each region. It is shocking how large of a difference in interrelation there can be when comparing streamflow percentile data at a regional scale. They are different regions, however, in the grand scheme of the atmospheric processes that are being examined, there are not very far away. Streamflow from one region can be highly correlated to AMO, while the region directly north could be barely correlated with AMO at all.

Clusters of regions could also be explored to determine regional difference. From the autocorrelation, multicollinearity, CWT, and WTC analyses, the identified clusters are as follows:

1. Arkansas-White-Red
2. Upper Missouri, Upper Colorado, and Pacific Northwest
3. Lower Colorado
4. Rio Grande

Upper Colorado and Pacific Northwest gauges produced extremely similar results for the entirety of the analysis. That pairing is set, and the multicollinearity and WTC analyses were convincing enough to include Upper Missouri within that group as well. Of the remaining three regions, Lower Colorado is definitely the most similar to that cluster, however, some of its results pointed it to being in its own category. For example, the strong correlation with ENSO and PDO at a period range of 120-180 months are defining factors of that cluster, whereas Lower Colorado only has a medium correlation with each. The Rio Grande gauge was a consistent outlier through the analysis, hence it not being clustered with any other regions.

Conclusion

The importance of understanding the way streamflow varies cannot be understated when it comes to those who rely on the Colorado River Basin's water. It is exciting to know that streamflow can be tied to teleconnection patterns both at a general scale of the entire CRB, but also at regional scales. The similarity across regions of strong correlation with ENSO, northern sun spots, and southern sun spots informs the scientific community that the streamflow percentiles within the CRB are partially described by changes in teleconnection patterns. This important understanding can be used for future prediction models, as well as knowing that the signals at periods of 10 to 15 years are where these interrelations can be found. ENSO is usually considered to be on a 3-7 year cycle, which end of that spectrum being constantly switching, so knowing that at somewhere around 11 to 13 years there is a consistent overlap could be helpful for excepting the next ENSO fluctuation.

At the regional scale, the inconsistent patches of correlation around periods of 16 and 64 months could be further examined to determine the specific ways each streamflow region differs from the other when it comes to interrelation to teleconnection patterns. The differences in which regions correlate with which patterns could also be incorporated into the research of hydrological behavior clustering. With this identification, these understandings could be used to further increase the industry's ability to regionally predict streamflow.

Bibliography

1. Barlow, M., et al. (2001). ENSO, Pacific decadal variability, U.S. summertime precipitation, drought, and stream flow. J. Clim., 14, 2105–2128.

   Summary The goal of this work is to examine the sea surface temperatures of the ENSO, PDO, and North Pacific mode teleconnection patterns and relate them back to precipitation and drought. This group used a covariance-based RPCA, and the NCEP-NCAR reanalyses to determine the strong links that these datasets have to U.S. drought events.

   Relevance The succes of this work inspired the idea of connecting streamflow, including streamflow drought, to these larger scales climate processes. If they can be connected to meteorological drought, surely they can show correlation to streamflow drought. The method used is interesting, but a bit outdated and overshadowed by more recent tools.

2. Canchala, T., Loaiza Cerón, W., Francés, F., Carvajal-Escobar, Y., Andreoli, R.V., Kayano, M.T., Alfonso-Morales, W., Caicedo-Bravo, E., Ferreira de Souza, R.A. (2020). Streamflow Variability in Colombian Pacific Basins and Their Teleconnections with Climate Indices. Water, 12, 526.

   Summary The goal of this paper is to relate teleconnection patterns to the flow of different river basins. They used wavelet analysis as well as PCA in order to find proof of multiscale connections between the two.

   Relevance This paper is extremely relevant to my project because they are connecting large-scale streamflow regions to teleconnection patterns by using one of the same tools I plan to use. I will most certainly be looking into using PCA for my project.

3. Chen, C., Lee, T. (2017). An Investigation into the Relationship between Teleconnections and Taiwan's Streamflow. Hydrology and Earth System Sciences Discussions, 1-34.

   Summary This paper's goal is to explore the relationship between climate-oscillations and streamflow in Taiwan with the goal of applying that understanding to a prediction model. They successfully accomplish this after finding that the West-Pacific and Pacific-Japan patterns had correlation values of 0.5 to streamflow, and then integrating these into a forecasting model.

   Relevance This paper relates to mine because they explore teleconnection pattern data enough to find correlation to large-scale streamflow data. They picked out the most important teleconnection patterns by using a lagged correlation approach. I plan on trying this in my project.

4. Chiew, F. H. S., et al. (2002). Global ENSO streamflow teleconnection, streamflow forecasting and interannual variability. Hydrological Sciences Journal, 47:3, 505-522.

   Summary This paper's goal is to identify the connections that ENSO has to streamflow all over the world. The authors accomplish this goal by fitting a first harmonic of the streamflow data to ENSO measurement gauges.

Relevance This paper explores the idea of directly relating streamflow data to teleconnection pattern data, and using them to pull out broad relationships that can be extrapolated to broader regions. This is quite close to the goal I am trying to accomplish, so it is definitely relevant.

5. Chikamoto, Y., et al. (2020). Colorado River water supply is predictable on multi-year timescales owing to long-term ocean memory. Communications Earth Environment, 1(1), 1-11.

   Summary The goal of this paper is to predict the Colorado River water supply years in advance. They accomplish this by using a complete climate model as well as years of river data.

   Relevance This paper may be the closest to the goal of my own project. Their success makes me hopeful that these kind of large scale connections will be found between Colorado River Basin streamflow data and teleconnection data. Their methods are not useful to me though as I do not have access to a full climate model.

6. Coulibaly, P., Burn, D.H. (2004). Wavelet analysis of variability in annual Canadian streamflows. Water Resources Research, 40.

   Summary The goal of this paper is to understand the variability in Canadian streamflows by comparing it to sea surface temperature oscillations. They used wavelet analysis in order to do this and discovered some strong connections to different teleconnection patterns during different years.

   Relevance This is a very relevant paper to my project. The authors are using the same wavelet analysis tools in order to discover more about the relationship between river flow and teleconnection patterns, and their approach has inspired my own.

7. Decastro, M., Lorenzo, N., Taboada, J., Sarmiento, M.A., Alvarez, I., Gómez-Gesteira, M. (2006). Influence of teleconnection patterns on precipitation variability and on river flow regimes in the Miño River basin (NW Iberian Peninsula). Climate Research, 32, 63-73.

   Summary The aim of this paper is to investigate how the Arctic Oscillation (AO) impact precipitation and streamflow in the Iberian Peninsula. They found that precipitation and river flow variablity are correlated, however, did not find a correlation between streamflow and the AO.

   Relevance This paper is relevant to my project because it provides background as to just how differently each sea surface temperature oscillation develops, behaves, and impacts the world. Some oscillations just do not have an impact on certain parts of the world, at least not obviously or directly. This reaffirms my idea that exploring multiple teleconnection pattern relationships is the best plan of action. None of the tools used in this paper will be applied to mine.

8. Dettinger, M. D., (2000). Multiscale streamflow variability associated with El Niño/Southern Oscillation. El Niño and The Southern Oscillation: Multiscale Variability and Global Regional Impacts Edited by: Diaz, H. F. and Markgraf, V. 113–147. Cambridge, UK: Cambridge University Press.

   Summary The aim of this paper is to explore the association of ENSO with streamflow values from the North and South Americas. They use principal component analysis on streamflow data to determine how significant the components are that are related ENSO values.

   Relevance This work is relevant to my project because it ties together streamflow and teleconnection patterns in a similar way that I plan on doing. It also looks at the data at a regional scale like my plan is. The use of PCA could be applied to my project, although there are currently no plans to.

9. Fleming, S.W., Dan Moore, R., Clarke, G.K. (2006). Glacier-mediated streamflow teleconnections to the Arctic Oscillation. International Journal of Climatology, 26.

   Summary The goal of this paper is to explore whether or not glacier and snowmelt fed rivers will behave differently to the Arctic Oscillation using streamflow data from the Yukon and British Columbia. They found that glacially-fed rivers respond positively to AO, whereas snowmelt fed rivers do not. They did so by using PCA and composite analyses.

   Relevance This paper is relevant to my project because they used statistical tools to discover a relationship between streamflow and a sea surface temperature oscillation. I will be exploring the idea of using PCA for my project.

10. Fu, C., et al. (2012). Analyzing the combined influence of solar activity and El Niño on streamflow across southern Canada. Water Resources Research, 48:5.

Summary This paper looks into how streamflow is impacted by both solar activity and ENSO. They use fourier spectrum analysis (FSA), continuous wavelet transform (CWT), and cross wavelet transform coherence analysis (WTC) to accomplish this goal.

Relevance This is definitely the article that is the closest to the goal of my paper. It looks into account solar activity and SSTs, and has been a spectacular guide for my project so far. The methods they used will be directly applied to achieve my project goal.

11. Graumlich, L.J., Pisaric, M.F., Waggoner, L.A., Littell, J.S., King, J.C. (2003). Upper Yellowstone River Flow and Teleconnections with Pacific Basin Climate Variability during the Past Three Centuries. Climatic Change, 59, 245-262.

Summary This paper investigates the connection between river flow in the Upper Yellowstone river and Pacific sea surface temperature oscillations by using a climate model. They could not find decision linkages between the two, and stated that it was likely due to the small streamflow dataset they had.

Relevance This paper relates to my project through the exploration of a relationship between a river in the western United States to a teleconnection pattern in the Pacific. However, they use a climate model to do so, and that is not a tool I have access to.

12. Haslinger, K., et al. (2014). Exploring the link between meteorological drought and streamflow: Effects of climate-catchment interaction. Water Resour. Res., 50, 2468–2487.

Summary This paper's goal is to determine the connection between streamflow and drought, but they do so by using some SST data. The tool they used was rank correlation analysis and were able to find some correlations as high as 0.8.

Relevance This paper relates to mine in the sense that it uses large scale data to discover something about streamflow. Rank correlation analysis is a tool that I may use later on in my project.

13. Ionita, M., Lohmann, G., Rimbu, N., Chelcea, S. (2011). Large-scale teleconnection patterns associated to Rhine river streamflow variability in spring and autumn. Geophysical Research Abstracts, 13.

Summary The goal of this paper is to associate streamflow variability in the Rhine river to sea surface temperature oscillations. They find that the two are correlated during shoulder, or transition, seasons (spring and fall). They used wavelet and cross spectra techniques to find this.

Relevance This paper is relevant to my project because their goal is very similar, but just on a more granular scale. This papers success gives me hope that the large scale streamflow data I have will turn up some results. I will be investigated the cross spectra method used in this paper.

14. Kennedy, A.M., Garen, D.C., Koch, R.W. (2009). The association between climate teleconnection indices and Upper Klamath seasonal streamflow: Trans-Niño Index. Hydrological Processes, 23.

Summary This paper aims to improve streamflow forecasting in the Upper Klamath river by integrateing teleconnection data into the model. They were able to successfully improve the model's prediction power via the 0.7 correlation to the Trans-Nino Index.

Relevance This paper relates to mine because it the authors searched for a strong enough connection between streamflow and a sea surface temperature oscillation that it would be considered significant, however, they applied that relationship to a forecasting model. Their analysis showed that this connection is present throughout the northwestern United States, which is promising for my research.

15. Kim, J., Jain, S., Moon, Y.I. (2012). Atmospheric teleconnection-based conditional streamflow distributions for the Han River and its sub-watersheds in Korea. International Journal of Climatology, 32.

Summary The aim of this paper is to predict streamflow distributions for a river basin in Korea by using supporting information from teleconnection pattern data. The used a decision tree to find the most contributing teleconnection patterns, and plugged those into a classifier to determine 50-year

highs and lows.

Relevance This paper relates to my project in the sense that it finds a connection between streamflow using 40-years of data, the same as mine, and teleconnection patterns, however, my project is not a prediction problem. The fact that their model was successful is encouraging for me when it comes to finding a defined relationship between the two in my project.

16. Kingston, D.G., McGregor, G.R., Hannah, D.M., Lawler, D. (2006). River flow teleconnections across the northern North Atlantic region. Geophysical Research Letters, 33.

    Summary The goal of this paper is to investigate how streamflow in North America is teleconnected to streamflow within Europe. They used cluster analysis in order to accomplish this, showing evidence of inverse river flow relationships between the two regions during autumn.

    Relevance This paper is only mildly relevant to my paper, but the idea behind it provides even more evidence that there could be regional relationships between streamflow and teleconnection patterns. Cluster analysis will be considered for my project.

17. Kwon, H., et al. (2009). Seasonal and annual maximum streamflow forecasting using climate information: application to the Three Gorges Dam in the Yangtze River basin, China. Hydrological Sciences Journal, 54, 582-595.

    Summary The goal of this paper is to predict maximum streamflow at different time scales for a single river. They used a Bayesian based prediction model to accomplish this.

    Relevance This is relevant to my project because they input SSTs and other climate information into the prediction model for streamflow, so seeing their feature selection process has been a helpful guide for me. I will not be doing any prediction for this project, but their data approaches are useful.

18. Lee, J. H., et al. (2019). Large-scale climate teleconnections with South Korean streamflow variability. Hydrological Sciences Journal.

    Summary The goal of this paper is to explore the connections between teleconnection patterns and South Korean streamflow variability. They use correlation and regression analyses in order to determine certain trends that connect the large scale processes to streamflow values.

    Relevance This paper is super relevant to my project. It explores very similar datasets with the same goal of inference rather than forecasting. The same wavelet analysis for correlation will be used in my project.

19. Maity, R., Kashid, S.S., (2010). Short-term basin-scale streamflow forecasting using large-scale coupled atmospheric oceanic circulation and local outgoing longwave radiation. Journal of Hydrometeorology, 11 (2), 370–387.

    Summary This paper looks into the relationship between streamflow and SST teleconnection patterns with the goal of forecasting streamflow values with that correlation. They used genetic programming to accomplish this and were able to improve weekly basin-scale streamflow predictions with it.

    Relevance This is useful to my project because even though their method and goal are quite a bit different from mine, the way that they select which variables of streamflow data to relate to ENSO has been important to my own decision-making about my project.

20. Mauas, P. J. D., et al. (2008). Solar Forcing of the Streamflow of a Continental Scale South American River. Physical Review Letters, 101 (16).

    Summary The goal of this research is to explore the relationship between streamflow values and sun spot numbers. They do so by using a Fourier filter, and find that the correlation coefficient between the two is r=0.78.

    Relevance This relates well to my project because it is one of the few papers I have found that uses modern statistical analysis techniques to explore streamflow and sun spot data together. The Fourier filter they used is in my current plan for next steps of my project.

21. Mishra, A.K., Singh, V.P., Özger, M. (2011). Seasonal streamflow extremes in Texas river basins: Uncertainty, trends, and teleconnections. Journal of Geophysical Research, 116.

    Summary The goal of this paper is to explore how season streamflow extremes are related over a large time scale, as well as compare that relationship to teleconnection patterns like ENSO and PDO. They used wavelet analysis, cross-correlation analysis and cross-wavelet transform to observe these relationships, and discovered powerful seasonal connections.

    Relevance This paper is super relevant to my project because it not only looks at the relationship between streamflow and teleconnection patterns, but also does so using the exact tools I plan on using.

22. Mosley, M. P. (2000). Regional differences in the effects of El Nino and La Nina on low flow and floods. Hydrol. Sci. J., 45(2), 249–268.

    Summary The goal of this paper is to explore how SST teleconnection patterns impact flooding in different regions of New Zealand. They do so by using regression analysis, and find specific regions of high risk.

    Relevance This research relates to mine because it is extrapolating value from large scale processes in order to predict surface water values. Their methods will not be used in my project however.

23. Ouachani, R., Bargaoui, Z., Ouarda, T.B. (2013). Power of teleconnection patterns on precipitation and streamflow variability of upper Medjerda Basin. International Journal of Climatology, 33.

    Summary The goal of this paper is to relate ENSO, PDO, and NAO to precipitation and streamflow values within the upper Medjerda Basin in the Mediterranean region. They found strong correlations at multiple time scales between precipication and streamflow to teleconnection patterns by using cross wavelet analysis.

    Relevance This paper is relevant to my project because they are determining correlation between streamflow and sea surface temperature oscillations by using the same tools I plan to use. This paper has been helpful to me in determining my wavelet analysis approach.

24. Pekárová, P., Pekař, J. (2007). Teleconnections of inter-annual streamflow fluctuation in Slovakia with Arctic Oscillation, North Atlantic Oscillation, Southern Oscillation, and Quasi-Biennial Oscillation phenomena. Advances in Atmospheric Sciences, 24, 655-663.

    Summary The aim of this paper is to relate long-term streamflow data in Slovokia to different teleconnection patterns. They used spectral analysis to find that there are strong correlations between the two at select time patterns.

    Relevance This paper heavily relates to my project through the fact that they searched for connections between river basin flow and teleconnection patterns using a statistical tool, however, I have not heard of spectral analysis until now. I will do research into whether or not it could help me with my project goal.

25. Rashid, M., et al. (2015). Assessment of Trends in Point Rainfall Using Continuous Wavelet Transforms. Adv. Water Resour., 82, 1–15.

    Summary This paper's goal is to determine rainfall trends as a way of looking into climate variability. They do so by use the continuous wavelet transform (CWT), and develop a technique that can be used for trend detection in other hydrological variables.

    Relevance This paper is relevant to my project because it directly applies a method I plan on using to a similar dataset.

26. Rust, W., et al. (2021). Exploring the role of hydrological pathways in modulating multi-annual climate teleconnection periodicities from UK rainfall to streamflow. Hydrol. Earth Syst. Sci., 25, 2223–2237.

    Summary The goal of this paper is to understand how teleconnection patterns interact with rainfall and streamflow. They use multi-resolution analysis to discover statisticall significant evidence of correlation for certain regions in the UK.

    Relevance This paper is relevant to my project because of the overlap it has with the datasets used and the goal of inference rather than foreacasting. The multi-resolution analysis used is interesting and could potentially be applied to my project in the future.

27. Sen, A.K. (2012). Streamflow variability in the Southern Appalachians and atmospheric teleconnections. River Research and Applications, 28.

    Summary The aim of this paper is to explain streamflow variability in the Southern Appalachians in the United States by using teleconnection pattern data. They use continuous wavelet transform (CWT) to find an interannual fluctuation that is hypothesized to be related to the Pacific North American teleconnection pattern.

    Relevance This paper relates to my project because it uses the same USGS streamflow dataset, just on a different region, to compare to teleconnection data using CWT. They did not find solid evidence of this connection though, which varifies my idea of using more than one analysis technique.

28. Steirou, E., et al. (2017). Links between large-scale circulation patterns and streamflow in Central Europe: A review. Journal of Hydrology, 549, 484-500.

    Summary The goal of this project is to determine what the link is between river discharge in Central Europe and large-scale climate indices. They found significant correlations for NAO and discharge through the use of wavelet transformation.

    Relevance This project is relevant to mine because of the similar datasets and the region-wise analysis using wavelet transformations. This paper has been an inspiration as to how to apply this tool to my project.

29. Switanek, M. B., Troch, P. A. (2011). Decadal prediction of Colorado River streamflow anomalies using ocean-atmosphere teleconnections. Geophys. Res. Lett. 38, L23404.

    Summary The goal of this paper is to explore how streamflow anomalies can be predicted by utilizing AMO and PDO data. They accomplish this by using regression and correlation analysis to determine a relationship, and then apply that relationship to a forecast model.

    Relevance This research relates to my project because it uses very similar datasets and the way they clean and prepare their data has been a useful guide for me so far. Their methods used are not helpful to my goal of inference, though.

30. Xu, K., Brown, C.M., Kwon, H., Lall, U., Zhang, J., Hayashi, S., Chen, Z. (2007). Climate teleconnections to Yangtze river seasonal streamflow at the Three Gorges Dam, China. International Journal of Climatology, 27.

    Summary The aim of this paper is to predict streamflow in the Yangtze river in China by integrating correlations to different teleconnection patterns. They successfully created a prediction model to help guide water management at the Three Gorges Dam.

    Relevance This paper relates to mine because they used statistical analysis to determine the regions that were the most highly correlated to the teleconnection pattern data they had. They used a linear correlation map to do this, which is the same correlation approach I tried although my correlations were not significantly high.