# Historical Climate Trends for Seattle City Light

Ross Dam, Skagit County, Washington (Seattle City Light)

*Prepared by*


Guillaume Mauger, UW Climate Impacts Group
Karin Bumbaco, Office of the WA State Climatologist
Robert Norheim, UW Climate Impacts Group
Jason Won, UW Climate Impacts Group
Nick Bond, Office of the WA State Climatologist


December 30th, 2016

# Contents

*Recommended citation format:* Mauger, G.S., Bumbaco, K.A., Norheim, R.A., Won, J.S., Bond, N.A. Observed Climate and Hydrologic Trends for Seattle City Light. Report prepared for Seattle City Light. Climate Impacts Group, University of Washington, Seattle. http://doi.org/10.7915/CIG3ST7FF

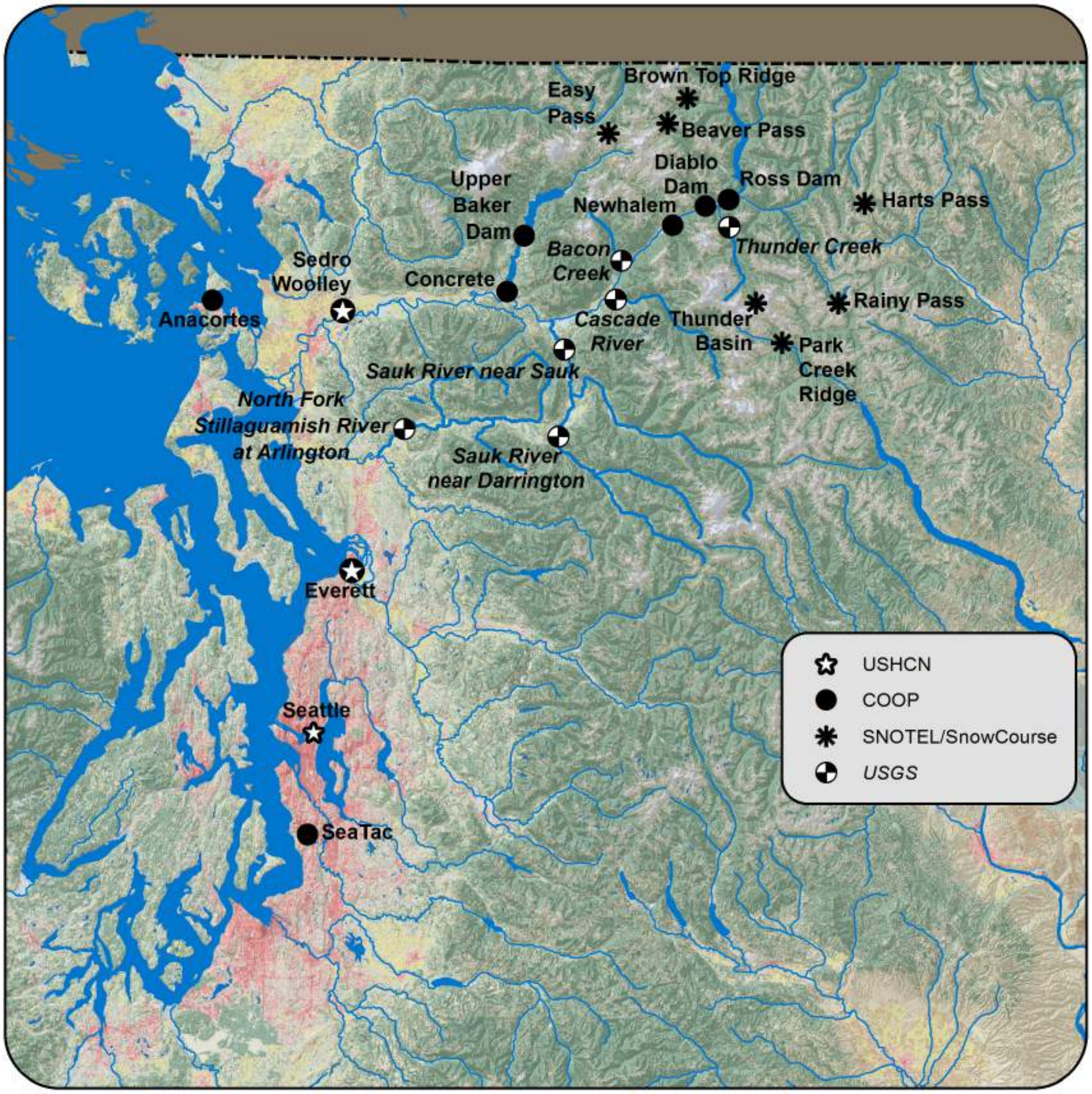
*Funding:* This project was primarily funded by Seattle City Light (SCL) via Crystal Raymond, SCL's climate change strategic advisor. Additional funds were provided by the State of Washington.

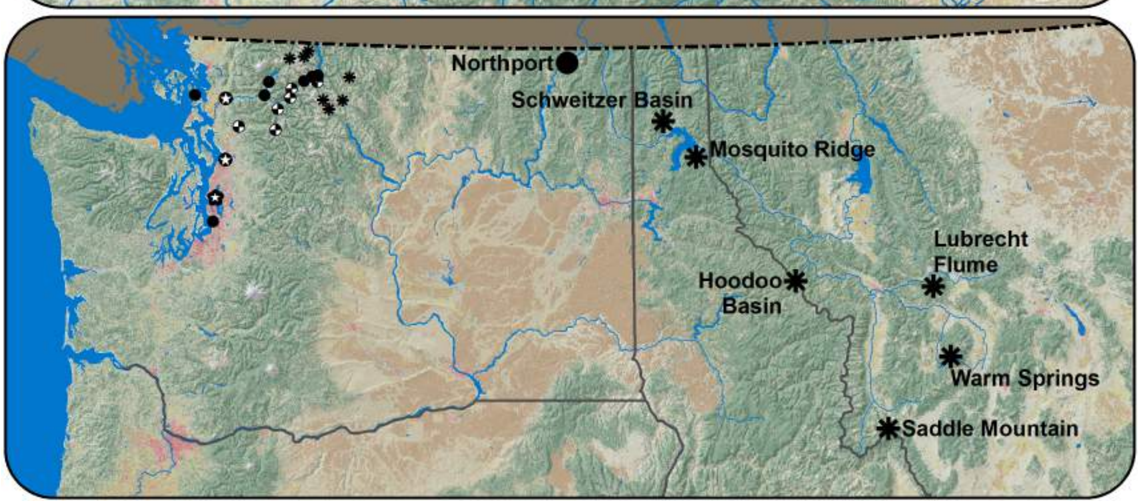
*Cover photo source:* https://www.seattle.gov/cityarchives/exhibits-and-education/online-exhibits/a-city-at-work/seattle-city-light, photo courtesy of Seattle City Light.

# 1 Purpose of this project

This work was designed to provide a centralized assessment of historical trends across a range of locations and metrics that are of interest to Seattle City Light. The intent is that this resource can help ensure that historical trends are interpreted consistently across the agency. In addition, the dataset is intended to support engagement efforts by providing greater information on the observed changes in climate to date.

**Figure 1** *(following page)*. Maps showing the locations of the observing stations used in this analysis. These include weather stations (temperature, precipitation, cloudiness), snow monitoring, and streamflow gauges. All were selected based on the interests of Seattle City Light. The top panel shows the stations that are located in western Washington, while the bottom panel shows stations that extend beyond this domain, including sites in northeastern Washington and the Rocky Mountains in Idaho and Montana.


Brown Top Ridge
Easy Pass
Beaver Pass
Diablo Dam
Ross Dam
Harts Pass
Upper Baker Dam
Newhalem
Thunder Creek
Bacon Creek
Sedro Woolley
Concrete
Rainy Pass
Anacortes
Cascade River
Thunder Basin
Park Creek Ridge
Sauk River near Sauk
North Fork Stillaguamish River at Arlington
Sauk River near Darrington
Everett
USHCN
COOP
SNOTEL/SnowCourse
USGS
Seattle
SeaTac


Northport
Schweitzer Basin
Mosquito Ridge
Lubrecht Flume
Hoodoo Basin
Warm Springs
Saddle Mountain

## 2 Data

We evaluated historical observations of temperature, precipitation, cloud cover, snowpack, and streamflow for a selection of observing locations throughout the Pacific Northwest (Figure 1). For each site, we calculated trends for a series of metrics of interest to Seattle City Light (see Tables in the Appendix). Data were obtained from the sources listed in the following subsections. Each section also describes the quality control efforts that were applied to the data.

### 2.1 Cooperative Observer Program (COOP) Network[1]

The Cooperative Observer Program (COOP) Network provides daily measurements of precipitation and minimum, maximum, and average daily temperature. For the Seattle station, we also considered observations of cloud cover. Maintained by volunteers, records at some locations extend back to the late 19th century. We obtained daily COOP data for each of the sites listed in Table 1. For each of these, we evaluated trends for all metrics listed in the Appendix (Table A1).

**Table 1.** NOAA Cooperative Observer Program (COOP) stations used in this project.

| Name | State | ID | Elev. (ft) | Lat. | Long. | Years |
|---|---|---|---|---|---|---|
| SeaTac | WA | 457473 | 370 | 47.44 | -122.31 | 1948-2014 |
| Everett | WA | 452675 | 60 | 47.98 | -122.20 | 1919-2012 |
| Anacortes | WA | 450176 | 20 | 48.51 | -122.63 | 1912-2006 |
| Sedro Woolley | WA | 457507 | 60 | 48.50 | -122.24 | 1898-2010 |
| Concrete | WA | 451679 | 195 | 48.54 | -121.74 | 1910-2014 |
| Upper Baker Dam | WA | 458715 | 690 | 48.65 | -121.69 | 1966-2014 |
| Diablo Dam | WA | 452157 | 891 | 48.71 | -121.14 | 1916-2014 |
| Newhalem | WA | 457690 | 525 | 48.68 | -121.24 | 1911-2014 |
| Ross Dam | WA | 457185 | 1236 | 48.73 | -121.07 | 1961-2014 |
| Northport | WA | 455946 | 1477 | 48.87 | -117.87 | 1900-2014 |

#### *2.1.1 Quality Control*

NOAA's National Center for Environmental Information (NCEI) flags COOP data that fails one of several quality tests, including: duplicate values, measurements that are outside of the

[1] https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/cooperative-observer-network-coop

expected range, lack of consistency with other observations, etc. As a conservative approach, we removed all data that had been flagged.

The temperature data were also filtered to remove anomalous monthly values. Specifically, data are considered invalid if the monthly average, after removing the long-term trend, is more than 3 standard deviations outside of the mean. Although some valid data could be flagged in this process, this is likely to be rare: the longest COOP record is 113 years in length, whereas normal temperature variations would only result in variations beyond the 3rd standard deviation about once in every 300 years. This process only resulted in the removal of 142 data points out of a total of 45,360 for all COOP stations and temperature variables.

Finally, it should be noted that quality control for COOP stations can be challenging. Many sites have been moved from time to time, observers have recorded observations at different times, and the instruments used to make measurements have changed substantially over the course of each record. These data should be considered the most likely to contain biases, especially prior to about 1950.

To synthesize the data into specific metrics, we applied the standard rules that are used by NCEI. Specifically, we only computed monthly averages if less than five (5) days of data were missing for that month, and only computed seasonal and annual averages if valid data were available for all of the months concerned. Some of the metrics are calculated using daily data that spans more than one month; these were treated similarly. For example, the number of days below freezing were only computed if all 12 calendar months had valid data, whereas the days above 86°F metrics were only computed if the months of April through October all had valid data. This approach could still result in some biases, since these are accumulated metrics and some of the days within each time period might be missing, even if less than five are missing per month. Since this represents at most 15% of the days, it is unlikely to have a large effect on the data. Nonetheless, we conducted sensitivity tests in which we compared the number-of-days and degree-day results both with and without adjustments for missing days. We adjusted the data for missing values by converting the number-of-days metrics (Warm Days, Freeze Days, Ice Days) to percent-of-days, and by replacing missing daily temperature data with the monthly average for the degree-day metrics (Heating Degree Days, Cooling Degree Days). Our tests showed that the adjusted data correlated highly ($r>0.95$) with the unadjusted metrics in all cases, that there were no systematic biases in the trend estimates, and that very few of the statistically significant trend estimates were biased by more thant 10%. The latter trend differences were confined to a small number of the heating degree day metrics; spreadsheets including comparisons for all sites and metrics are included in the online resources (Section 5). Since the degree-day corrections require us to estimate missing data – which is both subjective and requires creating synthetic data – and since so few trend estimates are affected, we have not applied any adjustments due to missing data. Note that our sensitivity tests did not evaluate the consecutive days metrics, since they concern much shorter time periods which are less likely to be affected by missing data and would also require more complex approaches to correct for missing values.

## 2.2 U.S. Historical Climatology Network (USHCN)[2]

The U.S. Historical Climatology Network (USHCN) provides monthly averages of daily minimum, maximum, and average temperature. This dataset is constantly being updated; in this project we analyzed trends using USHCN version 2.5.5.20160309. Data were obtained for the three stations listed in Table 2. For each of these, we assessed changes in the USHCN metrics listed in Table A1..

**Table 2.** NOAA Historical Climatology Network (USHCN) stations used in this project.

| Name | State | ID | Elev. (ft) | Lat. | Long. | Years |
|---|---|---|---|---|---|---|
| Seattle | WA | 457458 | 19 | 47.65 | -122.30 | 1896-2015 |
| Everett | WA | 452675 | 60 | 47.98 | -122.20 | 1897-2015 |
| Sedro Woolley | WA | 457507 | 60 | 48.50 | -122.24 | 1895-2015 |

### *2.2.1 Quality Control*

USHCN sites are a designated subset of long-term COOP weather stations that have been selected due to their long records, spatial coverage, and data completeness. The USHCN sites have been subjected to additional quality control on the monthly time scale to homogenize the record. Specifically, an objective comparison technique is used to identify breakpoints – or inhomogeneities – in the temperature record by comparing an individual station record to the surrounding stations (Menne et al. 2009). There are multiple quality control procedures that ultimately result in an adjusted dataset that has been corrected for biases resulting from urbanization, station moves, instrument upgrades, and time of observation changes (Menne et al. 2009).

We used the USHCN data as-is, without doing any additional quality control.

[2] https://www.ncdc.noaa.gov/oa/climate/research/ushcn/

## 2.3 NRCS Snow Courses[3]

Snow Course observations are made manually, with records extending back to the early 20th century. Snow Course sites are maintained by the USDA Natural Resources Conservation Service (NRCS). We obtained monthly Snow Course data for each of the sites listed in Table 3; the same sites used for the SNOTEL data (Section 2.4). For each of these, we evaluated trends in snowpack on April 1st and and May 1st (Table A2).

**Table 3.** NRCS Snow Course stations used in this project. The same sites were analyzed using SNOTEL data, described below.

| Name | State | ID | Elev. (ft) | Lat. | Long. | Years |
|---|---|---|---|---|---|---|
| Park Creek Ridge | WA | 20A12 | 4600 | 48.44 | -120.91 | 1929-2014 |
| Thunder Basin | WA | 20A07 | 2400 | 48.52 | -120.99 | 1948-2015 |
| Rainy Pass | WA | 20A09 | 4780 | 48.52 | -120.74 | 1930-2015 |
| Harts Pass | WA | 20A05 | 6200 | 48.72 | -120.66 | 1941-2015 |
| Easy Pass AM | WA | 21A07 | 5390 | 48.86 | -121.44 | 1959-2015 |
| Beaver Pass | WA | 21A01 | 3621 | 48.88 | -121.26 | 1944-2015 |
| Brown Top Ridge | WA | 21A28 | 6000 | 48.93 | -121.20 | 1971-2014 |
| Schweitzer Basin | ID | 16A10 | 6090 | 48.37 | -116.64 | 1980-1991 |
| Mosquito Ridge | ID | 16A04 | 5200 | 48.06 | -116.23 | 1937-1995 |
| Saddle Mountain | MT | 13D22 | 7940 | 45.69 | -113.97 | 1961-2002 |
| Warm Springs | MT | 13C43 | 7800 | 46.27 | -113.16 | 1961-2008 |
| Lubrecht Flume | MT | 13C38 | 4680 | 46.88 | -113.32 | 1961-2010 |
| Hoodoo Basin | MT | 15C10 | 6050 | 46.98 | -115.03 | 1961-2006 |

### *2.3.1 Quality Control*

Data are typically collected by measuring both the depth and the weight of the snowpack. The weight of the snowpack gives the snow water equivalent (SWE): the depth of liquid water that would be obtained by melting the snowpack. Measurements are usually repeated at 5 to 10 locations for each Snow Course, and these repeat observations are averaged to obtain the final estimate for each site.

We did not perform any additional quality control on the Snow Course data.

3 http://www.wcc.nrcs.usda.gov/about/mon_manual.html

## 2.4 NRCS Snowpack Telemetry (SNOTEL) Network[4]

The Snowpack Telemetry (SNOTEL) is a network of automated snow monitoring stations. These sites contain more measurements and at a higher resolution than the Snow Course sites, but have been in existence for a much shorter time period. SNOTEL sites are also maintained by the USDA Natural Resources Conservation Service (NRCS). We obtained daily SNOTEL data for each of the sites listed in Table 4; the same sites used for the Snow Course data. For each site, we evaluated trends for the metrics listed in Table A2.

**Table 4.** NRCS Snowpack Telemetry (SNOTEL) stations used in this project. The same sites were analyzed using Snow Course data.

| Name | State | ID | Elev. (ft) | Lat. | Long. | Years |
|---|---|---|---|---|---|---|
| Park Creek Ridge | WA | 681 | 4600 | 48.44 | -120.92 | 1979-2015 |
| Thunder Basin | WA | 817 | 4320 | 48.52 | -120.99 | 1989-2015 |
| Rainy Pass | WA | 711 | 4890 | 48.52 | -120.74 | 1983-2015 |
| Harts Pass | WA | 515 | 6490 | 48.72 | -120.66 | 1983-2015 |
| Easy Pass | WA | 998 | 5270 | 48.86 | -121.44 | 2010-2015 |
| Beaver Pass | WA | 990 | 3630 | 48.88 | -121.26 | 2002-2015 |
| Brown Top | WA | 1080 | 5830 | 48.93 | -121.20 | 2010-2015 |
| Schweitzer Basin | ID | 738 | 6090 | 48.37 | -116.64 | 1982-2015 |
| Mosquito Ridge | ID | 645 | 5200 | 48.06 | -116.23 | 1982-2015 |
| Saddle Mtn. | MT | 727 | 7940 | 45.69 | -113.97 | 1968-2015 |
| Warm Springs | MT | 850 | 7800 | 46.27 | -113.16 | 1978-2015 |
| Lubrecht Flume | MT | 604 | 4680 | 46.88 | -113.32 | 1971-2015 |
| Hoodoo Basin | MT | 530 | 6050 | 46.98 | -115.03 | 1967-2015 |

### 2.4.1 *Quality Control*

SNOTEL sites transmit data in real time via telemetry. Once a day, these observations are evaluated to ensure that they are within a pre-determined range of suitable values for each site based on the site's climatology. Values that are outside of this range are flagged as suspect and then manually inspected on a weekly basis, which includes tests for consistency among precipitation, temperature, and SWE measurements, as well as comparisons with nearby sites. Data may be replaced by redundant measurements if they are taken at the site in question, by using established precipitation to SWE or snow depth to SWE relationships, or by using adjusted data from nearby sites. An additional review of all daily values is performed at the end of each

4 http://www.wcc.nrcs.usda.gov/about/mon_automate.html

water year. Additional information on the quality control procedures can be found in NRCS (2011).

Following Serreze et al. (1999; see Appendix), we performed the following additional data checks:

1. Years with missing data: since SWE is accumulated, this could lead to errors later in the season.
2. Negative SWE values.
3. Days with greater than 10 in. of accumulation.
4. Days with back-to-back accumulations and losses greater than 2.5 in.

Of these, only the last criteria flagged any data, and only for seven days out of all of the SNOTEL records evaluated. Furthermore, of these seven instances only 2 occurred during the accumulation season. For those cases, SWE data were replaced with missing values.

As an additional safeguard against missing or biased data, several authors have suggested different metrics of spring snowpack and melt dates (personal communication: Andrew Slater, National Snow & Ice Data Center; and Scott Pattee, NRCS WA). For the April 1st and May 1st SWE metrics, we compared the values for these specific days to both the median and average of the 11 days surrounding each date. For example, we compared the following three estimates of April 1st SWE: (1) SWE value on April 1st, (2) Average SWE for March 25th through April 8th (11 day average), and (3) Median SWE for March 25th through April 8th (11 day median). The latter two metrics are intended to guard against anomalous and missing values. Comparisons across all of the SNOTEL stations included in this study showed that all three metrics gave nearly identical results: correlations between each were greater than 0.99.

Multiple definitions exist for the date of snowmelt. NRCS uses the first day with zero SWE (also used by Pederson et al. 2011), whereas others commonly use the first day with 2 in. of SWE remaining, and Hamlet et al. (2007) use the date when only 10% of peak SWE is remaining. Only dates falling after the water year peak SWE are considered. Since these may be sensitive to factors other than measurement error (rates of melt change throughout the season and as a function of the thickness of the snowpack), we have included results for each of these three metrics in the final dataset.

## 2.5 USGS Streamflow[5]

The USGS National Water Information System (NWIS) maintains a database of water resource monitoring sites across the country. These include daily streamflow observations for designated gauges. We evaluated daily streamflow observations from the six sites listed in Table 5. For each gauge, we evaluated trends for the metrics listed in Table A3.

In this project we do not consider streamflow gauges that are downstream of an existing dam. This is because of the large influence that dam operations can have on the seasonal patterns of streamflow. Since these flow modifications would confound our efforts to estimate long-term trends, our analysis only includes river sites that are not influenced by an upstream reservoir.

**Table 5.** USGS streamflow stations used in this project.

| Name | State | ID | Elev. (ft) | Lat. | Long. | Years |
|---|---|---|---|---|---|---|
| NF Stillaguamish at Arlington | WA | 12167000 | 89 | 48.26 | -122.05 | 1929-2015 |
| Sauk Nr Darrington | WA | 12187500 | 525 | 48.25 | -121.58 | 1915-2015 |
| Sauk Nr Sauk | WA | 12189500 | 266 | 48.42 | -121.57 | 1929-2015 |
| Cascade River | WA | 12182500 | 330 | 48.53 | -121.41 | 1929-2015 |
| Bacon Creek | WA | 12179900 | 410 | 48.60 | -121.40 | 1944-2015 |
| Thunder Creek | WA | 12175500 | 1220 | 48.67 | -121.07 | 1931-2015 |

### *2.5.1 Quality Control*

Although there are many different approaches to estimating streamflow, most gauges measure river stage – the elevation of the water surface – and use an established relationship ("stage-discharge relationship") to relate river stage to streamflow. USGS field agents typically visit a gauge once every six to eight weeks to make in situ flow measurements and check that the existing stage-discharge relationship still applies. Variations in this relationship could stem from seasonal effects of vegetation growth, which can limit the rate of flow, as well as long term changes, for example as sediments accumulate or are scoured from the river channel. These adjustments are applied over time as changes are detected by the field agents.

Provisional data undergo a series of checks before being considered final. This includes the adjustments resulting from field measurements as well as efforts to fill in missing data, which could be caused by ice effects or other measurement issues. When missing data are present,

5 http://waterdata.usgs.gov/nwis

flows are estimated from nearby unregulated gauges (i.e., where flows are not affected by a reservoir), based on scalings obtained from times when both gauges were operational. These corrections are likely to be good during normal flow conditions, but may be biased during periods of high flows. All daily streamflow data are accompanied by a subjective assessment of accuracy, which ranges from excellent (95% of daily discharges are correct within 5%), good (within 10%), fair (within 15 %), to poor which is less than fair accuracy.

We did not perform any additional quality control on the USGS streamflow data.

In estimating the flow timing metrics, we varied the start date relative to which flow timing is assessed. When peak or low flows coincide with the beginning or end of the water year, this can result in an artificially noisy dataset that simply reflects the arbitrary start and end dates of a 12-month calendar. In order to minimize this effect, we simply shifted the calendar in such a way as to minimize this "aliasing" (e.g., by using June $1^{st}$ instead of October $1^{st}$). Low flows are generally assessed relative to June $1^{st}$ while peak flows are assessed relative to August $1^{st}$. The one exception is Thunder Creek, for which April $1^{st}$ is used to assess the timing of peak flows.

# 3 Evaluating Trends

We assessed trends by considering both the linear regression as well as changes in the long-term (30-year) average. In this project we use the term “trend” to refer to changes over time, even though not all records show statistically significant changes over time.

## 3.1 Linear Trends

We computed linear trends using a standard ordinary least squares regression. This is the simplest approach to estimating a line from noisy data, and works by simply minimizing the difference between the data and a fitted line. The difference is calculated as the sum of the squared difference between the line and the data for each point. When the data are normally distributed (as is often the case for aggregate metrics like annual average temperature), ordinary least squares provides the maximum likelihood trend estimate.

## 3.2 Statistical Significance

We used two tests for statistical significance: the Student’s t-test (Student 1908) and the Mann-Kendall test (Kendall 1938). The Student’s t-test is based on the t-distribution, which is the expected distribution of random samples drawn from a normal distribution. The distribution changes with sample size and approaches the normal distribution at large sample sizes. Whereas the Student’s t-test assumes data are drawn from a normal distribution, the Mann-Kendall test is non-parametric: it does not assume that data are drawn from any particular distribution. This is useful for evaluating the significance of trends derive from metrics that are not normally distributed. The test works by counting the number of increases over time and comparing it to the number of decreases. In the case of random data, this statistic is expected to have a normal distribution; the p-value is derived by assessing the extent to which these departures (increases v. decreases) are consistent with a normal distribution.

In both cases, p-values indicate the likelihood that a particular trend could come about as a result of random chance. In this project we chose to report the 95% confidence level (a p-value of 0.05) as statistically significant, and further required that this condition be met by both tests to label a trend as significant. Since trends could be either positive or negative we used a two-tailed significance test. This is a more stringent test than the one-tailed alternative. The p-values for each significance test are included in the final products, allowing users to choose some other confidence limit for determining significance.

## 3.3 Autocorrelation

Significance tests require an estimate of the sample size used to calculate a trend. The *effective* sample size (sometimes referred to as the degrees of freedom) may be less than the actual number of points used to compute a trend if there is some relationship between the measurements in one year and the year after: in other words, if the data are autocorrelated. We tested for autocorrelation in all metrics by correlating each time series with itself for lags ranging from one to 10 years (e.g., Figure 2; results for all other metrics are included in the final products). Although autocorrelation is often an issue in time series analysis, annual data is frequently uncorrelated from one year to the next. The vast majority of metrics showed negligible correlation at one year lag. Some metrics did indicate some degree of persistence over time. These were mostly confined to the minimum temperature metrics, but also included some average and maximum temperature metrics as well as annual heating degree-days, and the persistence appears to be more important for summer temperature. However, these results were generally not consistent across observing sites or even when compared across similar metrics. As a result, all calculations used in this study assume that the effective sample size is the same as the actual sample size for this dataset. This does imply that the uncertainty estimates from the analysis are artificially low, but not to a substantial extent. Even for the records such as minimum temperatures that include non-negligible autocorrelation, the number of effective degrees of freedom is typically in the range of 0.5

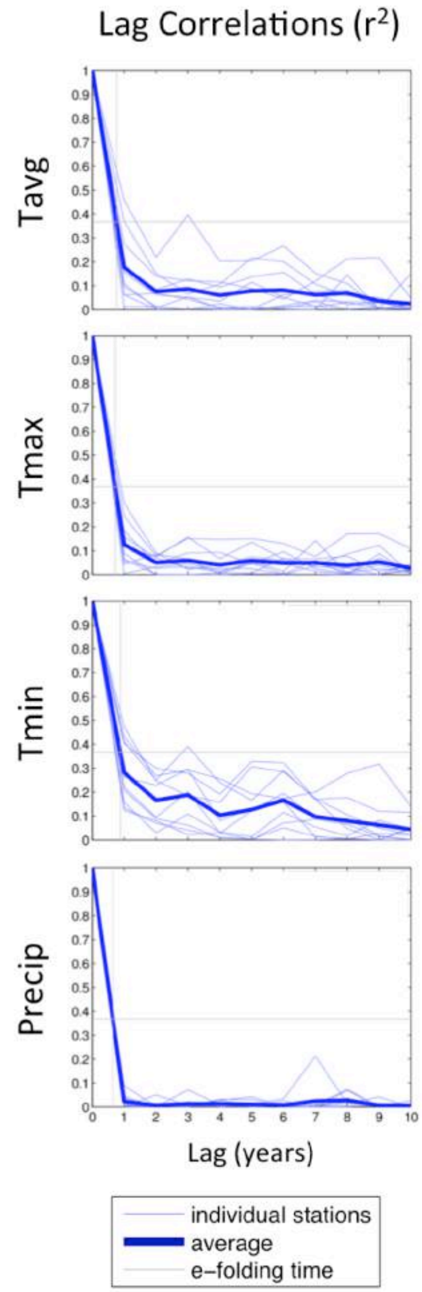



**Figure 2.** Year-to-year variations are essentially uncorrelated for the metrics considered in this study. The plot shows the autocorrelation (squared lag correlations; $r^2$) for annual average, minimum, and maximum temperature, and for precipitation. Each plot shows the results for all COOP stations. The e-folding time is often used to determine the decorrelation time scale and the implications for the number of degrees of freedom.

to 0.7 times the number of samples, resulting in minor changes in confidence levels.

## 3.4 Long-term Averages

Linear trends can be influenced by natural variability and may be unduly influenced by the start and end dates chosen for the analysis. As a complement to the linear regressions, we also computed the statistics of select 30-year periods for each metric and observing station. Although the choice of 30-years is arbitrary, it is the standard used in climate analyses and was chosen to minimize the impact of long-term variability on the averages. In addition, the 30-year averages are consistent with the climate normals that form the basis for many NOAA climate products.

In order to facilitate comparisons, we include the following statistics for each 30-year "epoch": average, standard deviation, and the $5^{th}$, $25^{th}$, $50^{th}$, $75^{th}$, and $95^{th}$ percentile values. As with the trends, statistics are calculated at 5-year intervals (e.g., 1976-2005, 1981-2010, 1986-2015) in order to facilitate evaluations of changing conditions over time.

Changes from one 30-year average to another may not be indicative of a long-term trend. Users can evaluate whether a changes is robust by looking for consistent changes across multiple 30-year averages, assessing the magnitude of change compared to variability, and checking if similar results are found from the linear regression analysis.

## 3.5 Criteria for Assessing Trends

To ensure that the calculations were robust – not overly influenced by missing data or anomalous values – we developed the following four criteria.

### *3.5.1 <u>Missing Data</u>*

Missing data could lead to biases in trends, since the remaining valid data may not be evenly sampled across warm and cold years or across time. For example, there could be more warm years missing than cold years, or vice versa. Similarly, an entire decade of data could be missing. In order to minimize this effect, we only

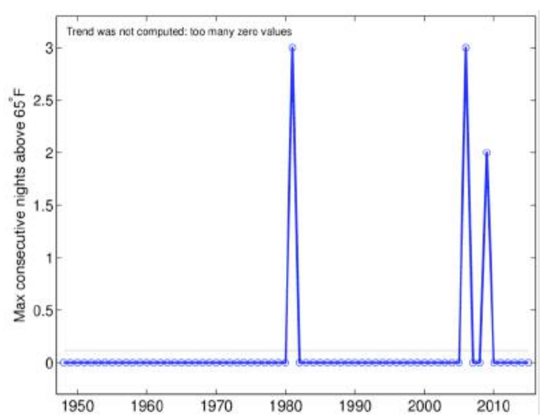



**Figure 3.** Example time series that is dominated by zero values: computing a trend from this time series would be misleading, and is inconsistent with the premise of the student's t-test for statistical significance. The example shows the number of Nighttime Warm Spells (defined in Table A1) for the SeaTac COOP station.

computed trends if valid data was present for at least 80% of all years.

### 3.5.2 *Repeat Values*

Several of the metrics remain pegged at zero for a substantial proportion of the record (e.g., Figure 3). This can lead to a false sense of certainty in trend estimates, since small departures may appear statistically significant, whereas in reality there exists variability that is not accounted for in the metric (i.e., the metric does not include variability for values that do not register above zero). In addition, the student's t-test is not robust when the probability distribution departs significantly from a gaussian. In order to ensure robust results, we only estimated trends if at least 50% of the data is non-zero.

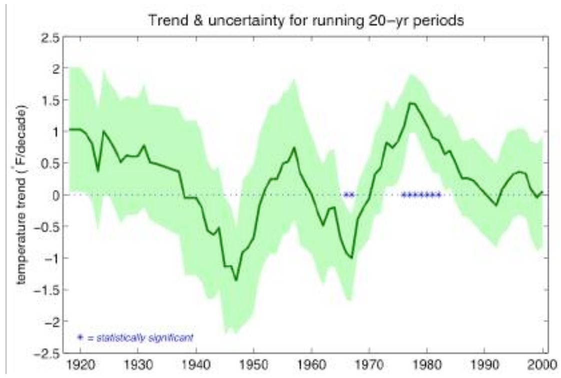



**Figure 4.** Trends based on short records can vary widely depending on the specific start and end years that are selected. This plot shows the estimated trend and uncertainty for running 20-year periods spanning from the start to the end of the record, based on the time series of average annual temperature from the Sedro Woolley COOP station. Estimates range from large positive to large negative values, and very few are statistically significant.

### 3.5.3 *Minimum Record Length*

Trends computed from short records can be overly biased by natural variability (Figure 4). Specifically, short-term variability can result in trends that are either much higher or much lower than the long-term trend (for example, by picking an El Niño year as the starting point and a La Niña year as the end point). This problem is minimized with longer records because the larger number of points reduces the impact of anomalous values. To avoid this pitfall, we only computed trends for periods of at least 30 years in length.

### 3.5.4 *Start Year*

Even for longer records, trend estimates can change depending on which start and end years are chosen. This is particularly true given the increased potential for measurement biases prior to about 1950.

The sensitivity to start and end dates is very clearly illustrated by the time series of annual precipitation at the Diablo Dam COOP site (Figure 5; see also Figure 4). Due to several years of particularly low precipitation in the 1930s and 1940s, trends that include these years tend to show a statistically significant increase in precipitation. In contrast, trends starting after about

1945 show no change over time.[6] Regardless of the cause of these low precipitation values, they appear to be more associated with short-term variability than indicative of a long-term trend.

To provide additional context for evaluating trends, we estimated the long-term trend for different start years, beginning with the start of the record and continuing at five year intervals until fewer than 30 years are available for a trend estimate. All trend calculations continue until the latest available observations (i.e., the start year is varied, but not the end year of each trend). These results are archived with the final dataset, as described below.

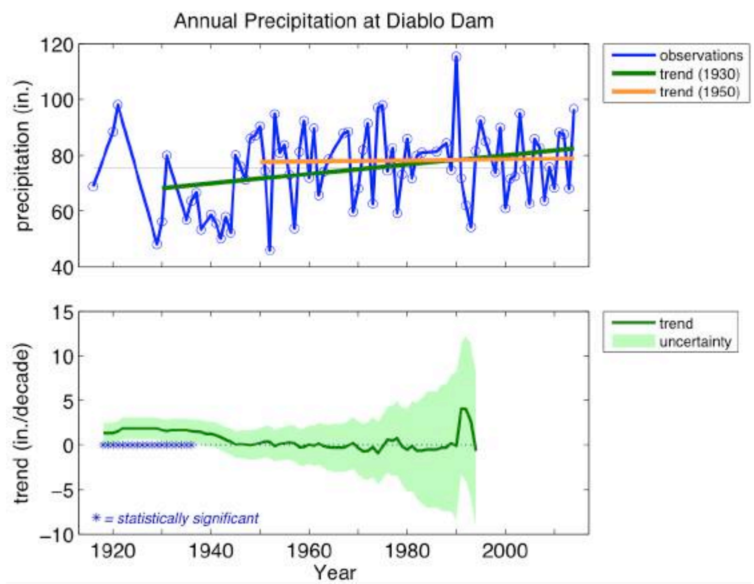



**Figure 5.** Trend estimates differ depending on which years are included in the analysis. Both plots show the effect of evaluating trends starting at different times in the past. The top panel shows the observed annual precipitation at the Diablo Dam COOP site (blue) along with two trend lines: One calculated using data from 1930-2015 (green; this trend shows a statistically significant increase), and another using data from 1950-2015 (orange; this trend is not statistically significant). The bottom panel shows the trend, and 95% uncertainty limits, calculated for observations starting in each year from 1918-1994 (trends are calculated after 1994 become highly uncertain due to the increasingly short record). Years with statistically significant trends are highlighted with an asterisk. For example, for the year 1920, the plot shows that a trend calculated from the years 1920-2015 shows a statistically significant increase, where as the trend for 1960-2015, shown for the year 1960, is not statistically significant.

[6] Although the timing of the low precipitation values suggests that measurement error may be an issue, this period roughly coincides with the timing of the Dust Bowl, a time that was associated with below average precipitation across much of the U.S., including many of the other COOP sites included in this analysis. As of this time it is not known how much the precipitation deficits in the 1930s and early 1940s in the datasets used here reflect actual weather conditions versus measurement errors.

## 4 Connection to Large Scale Variability

The presence of a long-term trend does not necessarily imply that it is due to climate change, nor is the converse necessarily true in the absence of a statistically significant trend. One useful diagnostic for evaluating the possible role of natural variability is to examine the relationship between a time series and a measure of the large-scale atmospheric circulation. Although changes in circulation can be driven by global warming, variations in the large-scale circulation are thought to be largely driven by internal variability and therefore not related to human-driven changes in climate.

In order to provide a straightforward first estimate of possible large-scale influence, we calculate the correlation between 500 hPa heights (hereafter referred to as Z500) and each metric, and compare that to the long-term trend in Z500. These were calculated using the NOAA Earth System Research Laboratory's online tool,[7] which can be used to calculate correlations with a number of climate and weather metrics, including Z500. Correlations are calculated for the years 1948-2015, when data are available.

The 500 hPa level is situated about halfway up the troposphere, and can be considered the "steering level" for atmospheric wind and pressure patterns. Where Z500 is high it is associated with relatively warm temperatures and often a high pressure "ridge". Similarly, when Z500 is low this represents generally cooler temperatures in the lower atmosphere, often associated with a low pressure "trough".

Figure 6 includes maps of the correlations with five key metrics, selected to represent the bulk of those analyzed in this study:

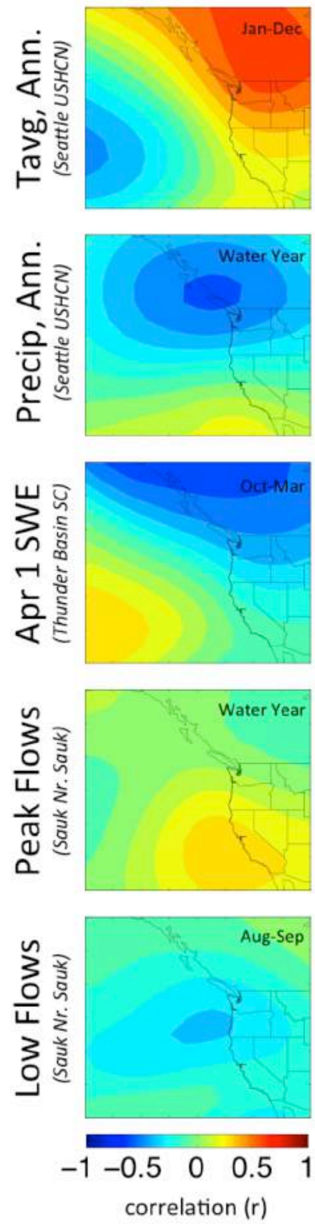



**Figure 6.** Large-scale variability explains some but not all of the changes over time. The maps show the correlation between select metrics and the height of the 500 hPa surface, a key measure of the large-scale circulation.

7 https://esrl.noaa.gov/psd/data/correlation/

- Average annual Temperature, Seattle USHCN site (correlated with Jan-Dec Z500)
- Annual Precipitation, SeaTac COOP site (correlated with Oct-Sep Z500)
- April 1$^{st}$ SWE, Thunder Basin Snow Course site (correlated with Oct-Mar Z500)
- Peak Daily Streamflow, Sauk River near Sauk (correlated with Oct-Sep Z500)
- Low 7-day Streamflow, Sauk River near Sauk (correlated with Aug-Sep Z500)

Overall, each of these shows the anticipated relationship to large-scale conditions. Warm temperatures are associated with anomalous ridging to the north of the Pacific Northwest, which promotes sinking motion and the advection of warm air from the interior to the coasts. Similarly, both precipitation and snowpack are associated with negative anomalies over the Pacific Northwest and anomalous westerly flow: these suggest cooler conditions and a stronger storm track. The Z500 correlation for SWE suggests that higher snow accumulations are associated with an increased influence of cold air coming from the Gulf of Alaska. Both peak and low flows are not as well correlated with Z500. Low flows are related to warm conditions and ridging in summer. Peak flows are more difficult to interpret but may be associated with warmer conditions, resulting in lower snow accumulation and a greater proportion of precipitation falling as rain.

Correlations are the greatest for annual average temperature, which is just over a correlation of 0.7 at its maximum point over British Columbia, while none of the others exceed 0.5 anywhere in the domain. This implies that variations in Z500 explain about 50% of the long-term changes that are observed for temperature ($0.7^2 = 0.49$; this is consistent with other estimates in the literature, e.g. Wallace et al. 1996), while for the other variables it explains about 25% of the variability. For all of the metrics, the actual influence of natural variability is greater, since Z500 is not the only natural driver of variability in each metric.

# 5 Project Outputs

The following subsections describe the project outputs. These can all be accessed at the project website: https://cig.uw.edu/seattle-city-light-trends/

## 5.1 Data and Trends Archive

An online archive contains all of the observational data used in this project as well as the calculated trends for each metric. All files are stored in a comma-delimited format (.csv) with a 5-line header that describes the file's contents, the date on which it was created, and contact information for any questions that arise. All data are stored in a separate folder denoting the network, ID, and name of each observing site. For example, all metrics and trends computed for the SeaTac COOP site are stored in the folder entitled *COOP_457473_SeaTac*.

Data for each variable (temperature, precipitation, etc.) are stored in monthly and daily files as appropriate (COOP, SNOTEL, and USGS data are all daily in time resolution, whereas the USHCN and Snow Course data only exist at a monthly time step). Metrics are stored in the same fashion, but the data are yearly in time resolution. Files are named similarly to the directories, with an additional suffix to denote the variable or metric stored in each file, following the "Short Name" convention listed in Tables A1, A2, and A3. For example, year-to-year observations of the number of warm days at SeaTac are stored under *COOP_457473_SeaTac_WarmDays.csv*

As described in Section 3, the data for each metric is used to estimate trends and the statistics for 30-year epochs. These results are stored in two separate data files along with two associated figures, each with a suffix denoting *"lin_trend"* or *"30-yr_avg"*. For example:

- *COOP_457473_SeaTac_WarmDays.lin_trend.csv*
- *COOP_457473_SeaTac_WarmDays.lin_trend.png*
- *COOP_457473_SeaTac_WarmDays.30-yr_avg.csv*
- *COOP_457473_SeaTac_WarmDays.30-yr_avg.png*

The trends files contain trend estimates for each start year beginning with the first year of observations and ending when the criteria for estimating a trend is no longer met (often the last valid trend estimate is for 1985-2015). The following information is included for each trend estimate:

- Start and end years used for the trend calculation
- Number of years of valid data used to estimate the trend
- Linear trend
- 95% confidence limits (lower and upper bound) on the trend estimate

- p-value (Student's t-test)
- p-value (Mann-Kendall test)
- minimum detectable trend
- trend-estimated values for the start and end years (used in plotting).

The minimum detectable trend is derived from the Student's t-distribution, and can be compared to the observed trend to note how close or far the estimate is from the minimum threshold for detectability.

The 30-yr average files are similarly organized, but with different statistics and 5-year increments that are slightly offset from the trends analysis. For consistency with the climate normals that are commonly used by NOAA and others, the 30-year averages are computed starting on the $1^{st}$ and $6^{th}$ year of each decade. For example, whereas linear trends might be computed for 1980-2015 and 1985-2015, the 30-year averages are computed for 1981-2010 and for 1986-2015. This allows for a straightforward comparison to standard NOAA products.

The following information is included for each 30-year epoch:

- Start and end years used for to calculate statistics
- Number of years of valid data
- Average
- Standard deviation
- Uncertainty in the mean (standard deviation divided by the square root of the sample size)
- $5^{th}$, $25^{th}$, $50^{th}$, $75^{th}$, and $95^{th}$ percentile values, respectively.

In addition to the data and trends information described above, the online directory also includes reference figures showing the results of the autocorrelation analysis for all metrics, the comparisons with large-scale conditions for a selection of key metrics, and spreadsheets comparing results both with and without adjustment for missing data. These are stored in separate directories entitled "Autocorrelation", "LargeScale_Comparisons", and "MissingData_Tests", respectively.

As in Figure 2, each autocorrelation figure combines the results for all measurement sites in one figure. This is done to facilitate side-by-side comparisons, since autocorrelation on such coarse time scales can be influenced by measurement biases and is best interpreted in aggregate.

The large scale comparisons are constructed by separately comparing the time evolution of a metric with the 500 hPa heights (Z500) at each point. This produces a map that highlights the patterns of association between Z500 and each selected metric. Since correlations alone can sometimes be misleading – due to the fact that a strong correlation does not necessarily imply a large effect – regression maps are included for comparison with the correlations. These maps were obtained directly from ESRL in their automated figure format.

The missing data comparisons directory includes two spreadsheets that summarize the comparisons between the number-of-days and degree-day metrics, calculated both with and without adjustment for missing data. For each measurement site and metric, the spreadsheets show the correlation between the two time series, the linear trend and p-values for each, as well as the differences between the two trend estimates. The final column flags statistically significant trends for which the adjusted time series differs from the unadjusted data by more than 10%.

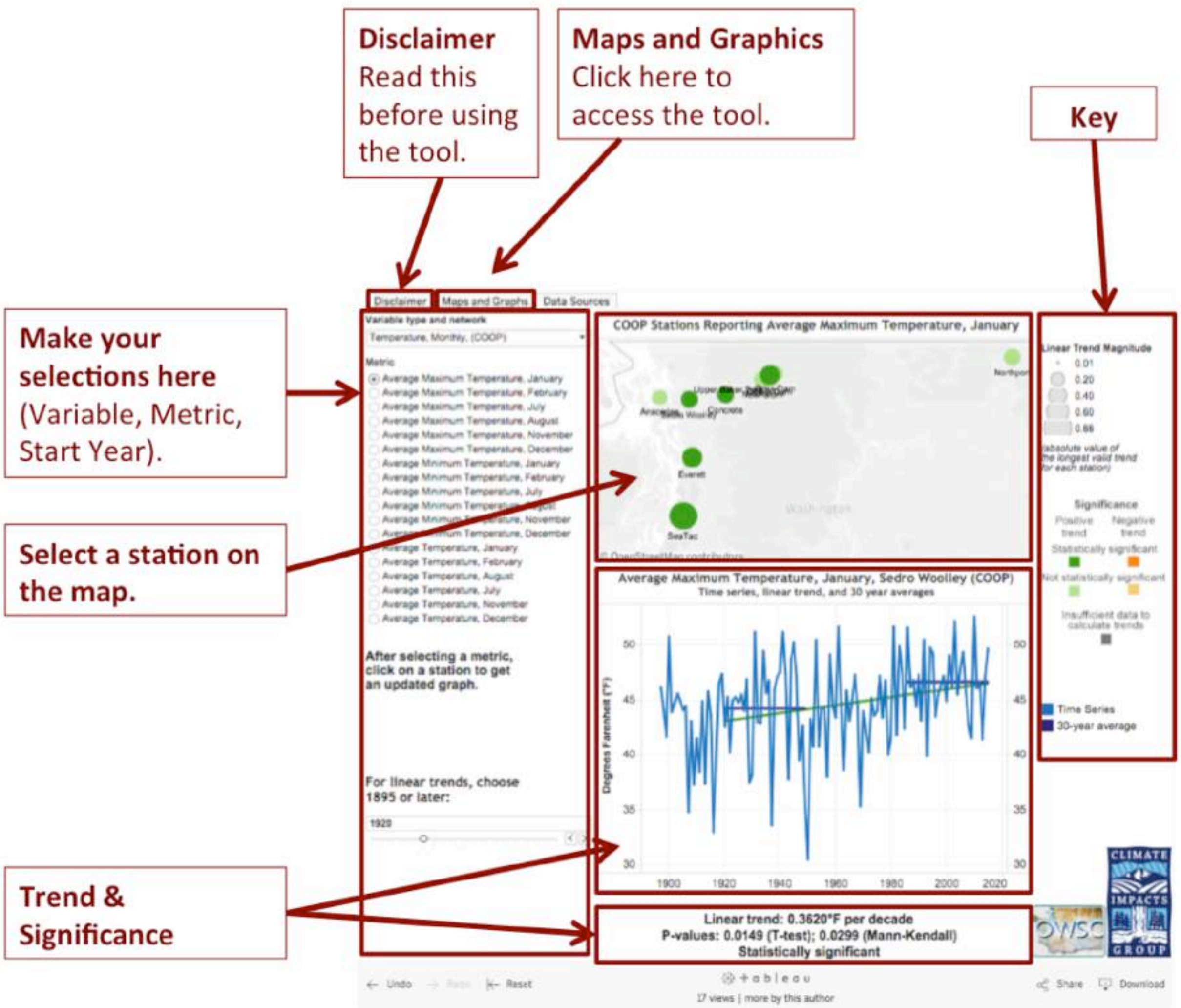



**Figure 7.** Screenshot, with descriptions, of the web tool.

## 5.2 Tableau Tool

As a complement to the reference data products, we have also produced a tool that is intended to allow users to easily visualize and query observed changes for each metric (Figure 7).

By selecting a particular metric, users can view the long-term trend for all stations that include observations for that metric. The map shows the trend for the full record of available data at each station, and includes information on both the sign (increasing or decreasing) and magnitude of the trend along with the statistical significance of each. Consistency across many stations is an indication of a robust trend.

After selecting a particular station on the map, the time series figure will update, showing both trends and 30-year averages. The text below the plot lists the trend estimate and the results of the significance tests. When possible, 30-year averages are included for both the beginning and end of the record, allowing a comparison of changes over time. The slider bar on the left can be used to adjust the start year for the trend analysis and the first of the two 30-year averages that are displayed. Both trends and 30-year averages are displayed as long as the necessary criteria are met (as described in Section 3). In order to maintain readability, the first 30-year average is not displayed if there are 15 years or more of overlap between it and the final 30 year average.

# 6 Interpreting the Results

This section describes some of the factors that should be considered in interpreting the results of this analysis.

## 6.1 Not all metrics show a trend

Although the title of this project includes the word “trend”, this does not imply that all observations were assumed to have a long-term trend. In fact, a major motivation for this project was to distinguish between observations that do show a statistically significant trend and those that do not. In order to do this, the analysis included a careful evaluation of data quality, two separate tests of statistical significance, and information on changes in trends over time.

## 6.2 Several factors influence trends

This dataset highlights the presence of long-term trends in some metrics. In some cases, greenhouse gas emissions have played a role in contributing to the trends. However, other factors can also influence trends. These include natural year-to-year fluctuations in climate, measurement biases, and changes in land use and development. In general, long-term trends should be interpreted as a combination of these factors, the relative contributions of which should be assessed on a case-by-case basis.

## 6.3 Trend estimates can be affected by errors in the data

Data may be subject to errors due to changes in instruments, measurement location, or other issues, and these errors can result in inaccurate trend estimates. Although we applied standard quality control approaches to the data, some of these errors are likely to persist in the data presented here.

## 6.4 Identifying robust trends

Robust trends tend to be consistent over time and space. If many stations show a similar trend, especially if many of these are statistically significant, this suggests that the trend is real and not an artifact. Similarly, if the trend remains statistically significant when assessed for different time periods – for example, by changing the start year for trend calculations – this too provides some evidence that the trend is robust.

## References


Hamlet, A. F., Mote, P. W., Clark, M. P., & Lettenmaier, D. P. (2007). Twentieth-century trends in runoff, evapotranspiration, and soil moisture in the Western United States. Journal of Climate, 20(8), 1468-1486. http://dx.doi.org/10.1175/JCLI4051.1

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81-93. http://dx.doi.org/10.1093/biomet/30.1-2.81

Kennedy, E. J. (1983). Computation of continuous records of streamflow (No. 03-A13; 53 pp.). US GPO. https://pubs.usgs.gov/twri/twri3-a13/

Menne, M. J., Williams Jr, C. N., & Vose, R. S. (2009). The US Historical Climatology Network monthly temperature data, version 2. *Bulletin of the American Meteorological Society*, *90*(7), 993. http://dx.doi.org/10.1175/2008BAMS2613.1

(NRCS) U.S. Department of Agriculture, Natural Resources Conservation Service. 2011. Snow survey and water supply forecasting. National Engineering Handbook, Part 622. Washington, DC. https://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=33164.wba

Pederson, G. T., Gray, S. T., Ault, T., Marsh, W., Fagre, D. B., Bunn, A. G., ... & Graumlich, L. J. (2011). Climatic controls on the snowmelt hydrology of the northern Rocky Mountains. Journal of Climate, 24(6), 1666-1687. http://dx.doi.org/10.1175/2010JCLI3729.1

Rantz, S. E. (1982). *Measurement and computation of streamflow: volume 2, computation of discharge* (No. 2175; pp. 285-631). US GPO. http://pubs.water.usgs.gov/wsp2175

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry(SNOTEL) data. Water Resources Research, 35(7), 2145-2160. http://dx.doi.org/10.1029/1999WR900090

Student. (1908). The probable error of a mean. *Biometrika*, 1-25. http://dx.doi.org/10.2307/2331554

Wallace, J. M., Zhang, Y., & Bajuk, L. (1996). Interpretation of interdecadal trends in Northern Hemisphere surface air temperature. *Journal of Climate*, *9*(2), 249-259. http://dx.doi.org/10.1175/1520-0442(1996)009%3C0249:IOITIN%3E2.0.CO;2

## Appendix A – Metrics

**Table A1.** Temperature, precipitation, and cloud cover metrics included in this study. Cloud cover was only assessed for SeaTac, whereas all other metrics were assessed for every COOP site. Since the USHCN data are monthly, only the monthly, seasonal, and annual temperature metrics were assessed for this dataset, as noted in the final column of the table.

| Metric | Short Name | Aggregation | Units | Notes | USHCN |
|---|---|---|---|---|---|
| Precipitation, Annual | Precip-ANN/WY | sum | in | Calendar Year, Water Year | |
| Precipitation, Seasonal | Precip-MMM-MMM | sum | in | Oct-Dec, Oct-Mar, Sep-Apr, Apr-Jun, Jun-Aug, Jul-Sep, Aug-Sep | |
| Precipitation, Monthly | Precip-MMM | sum | in | Nov, Dec, Jan, Feb, Jul, Aug | |
| Precipitation, Extremes | Precip-p95Avg | average | in | Daily Precipitation, Average of top 5% | |
| Precipitation, Extremes | Precip-MAX | n/a | in | Daily Precipitation, Annual Max | |
| Average Temperature, Annual | Tavg-ANN/WY | average | °F | Calendar Year, Water Year | ✔ |
| Minimum Temperature, Annual | Tmin-ANN/WY | average | °F | Calendar Year, Water Year | ✔ |
| Maximum Temperature, Annual | Tmax-ANN/WY | average | °F | Calendar Year, Water Year | ✔ |
| Average Temperature, Seasonal | Tavg-MMM-MMM | average | °F | Oct-Mar, Dec-Feb, Mar-May, Jul-Sep | ✔ |
| Minimum Temperature, Seasonal | Tmin-MMM-MMM | average | °F | Oct-Mar, Jul-Sep | ✔ |
| Maximum Temperature, Seasonal | Tmax-MMM-MMM | average | °F | Oct-Mar, Jul-Sep | ✔ |
| Average Temperature, Monthly | Tavg-MMM | average | °F | Nov, Dec, Jan, Feb, Jul, Aug | ✔ |
| Minimum Temperature, Monthly | Tmin-MMM | average | °F | Nov, Dec, Jan, Feb, Jul, Aug | ✔ |
| Maximum Temperature, Monthly | Tmax-MMM | average | °F | Nov, Dec, Jan, Feb, Jul, Aug | ✔ |
| Diurnal Temperature Range | DiurnalTempRange-Aug | average | °F | USHCN | |
| Warm Days | WarmDays | n/a | days | Number of days with Tmax > 86°F | |
| Warm Spell Duration | WarmSpellDuration-Days | n/a | days | Max Consecutive Days with Tmax > 86°F | |
| Warm Spell Duration | WarmSpellDuration-Nights | n/a | days | Max Consecutive Days with Tmin > 65°F | |
| Cooling Degree Days (base 65°F), Seasonal | CDD-65F-MMM-MMM | n/a | °F-days | Jun-Aug | |
| Cooling Degree Days (base 75°F), Seasonal | CDD-75F-MMM-MMM | n/a | °F-days | Jun-Aug | |
| Cooling Degree Days (base 65°F), Annual | CDD-65F-ANN | n/a | °F-days | Calendar Year | |
| Cooling Degree Days (base 75°F), Annual | CDD-75F-ANN | n/a | °F-days | Calendar Year | |
| Heating Degree Days (base 65°F), Seasonal | HDD-65F-MMM-MMM | n/a | °F-days | Oct-Dec, Jan-Mar | |
| Heating Degree Days (base 65°F), Annual | CDD-65F-ANN | n/a | °F-days | Calendar Year | |
| Number of Frost Days | FrostDays | n/a | days | Number of days with Tmin<32°F | |
| Number of Freeze Days | FrostDays | n/a | days | Number of days with Tmax<32°F | |
| Cloud Cover, Midnight-Midnight | Cloud-24hour-MMM-MMM | average | % | Dec-Feb, Jun-Aug. SeaTac Only. | |
| Cloudiness, Sunrise-Sunset | Cloud-24hour-MMM-MMM | average | % | Dec-Feb, Jun-Aug. SeaTac Only. | |

**Table A2.** Snow metrics included in this study. All metrics were assessed for every SNOTEL site. Since the Snow Course (SC) data are monthly, only the monthly metrics were assessed for this dataset, as noted in the final column of the table. All melt date criteria are assessed only for dates following the maximum SWE for a given water year.

| Metric | Short Name | Units | Notes | SC |
|---|---|---|---|---|
| April 1$^{st}$ Snow Water Equivalent (SWE) | Apr1-SWE | in | | ✔ |
| May 1$^{st}$ Snow Water Equivalent (SWE) | May1-SWE | in | | ✔ |
| Maximum Snow Water Equivalent | Max-SWE | in | Maximum Daily SWE for the Water Year | |
| Melt Date | MeltDate-0in | day of WY | First day with 0 in. of SWE | |
| Melt Date | MeltDate-2in | day of WY | First day with < 2 in. of SWE | |
| Melt Date | MeltDate-10pct | day of WY | First day with < 10% of Max-SWE remaining. | |

**Table A3.** Streamflow metrics included in this study.

| Metric | Short Name | Aggregation | Units | Notes |
|---|---|---|---|---|
| Late Summer Streamflow | Aug15-Sep15-Flows | average | cfs | Average Streamflow for Aug 15$^{th}$ - Sep 15$^{th}$ |
| Minimum 7-day Flow | LowFlows | n/a | cfs | Water Year minimum of 7-day average flows |
| Minimum 7-day Flow Timing | LowFlowDate | n/a | day* | Date of minimum 7-day streamflow |
| Peak Daily Flow | PeakFlows | n/a | cfs | Water Year maximum of daily flows |
| Peak Daily Flow Timing | PeakFlowDate | n/a | day* | Date of maximum daily streamflow |

* *These were assessed relative to other dates (Apr 1$^{st,}$ Jun 1$^{st}$, Aug 1$^{st}$) in order to minimize aliasing (Section 2.5.1).*