Leveraging USGS Stream Gauge Records to Quantify Channel Change in Response to Extreme Precipitation Events in California

Yuliya Dominguez, Kai Hidaka, Emmett Jaecklein

With Guidance from Dr. Allison Pfeiffer and Dr. Brian Hutchinson

Dr. Allison Pfeiffer

Western Washington University - Geology

Focus: Geomorphology, the study of the physical features of the surface of the earth and their relation to its geological structures.

Research project supervisor for various WWU geomorphology groups.

Personal work done on stream flow and river incision and aggregation in Washington State.

USGS and River Channel Change

The United States Geological Survey (USGS) is a scientific agency within the United States federal government.

The USGS is one of the oldest scientific agencies in the nation.

The USGS keeps extensive waterdata records, including streamflow data recorded by stream gauge sites.

Stream flow is used to directly calculate sediment flow, and thus channel change.

Goals

- Collect data on river bed elevation change in California.
- Create an automated method for the selection of valid candidate sites.
- Create clean visuals for the residual values from the rating curves.
- Gain insight into the river bed elevation variability in California.
- Deliver code to scrape and clean field measurements, rating curve and water-year summary data from any state for future research.

Data Scraping and Processing

Important information for the model selection was located on the USGS gage site page.

Using the BeautifulSoup Python library, we can scrape this information.

Scrapped information contains gage location, site specifications, and gage remarks.

Data Scraped:

- Water-Year Summaries.
- Site Rating Curves.
- Field Measurements.
- USGS Remarks.

Water-Year Summary for Site 11109000

Available data for this site: Water-Year Summary GO

Click to hide Water-Year Summary instructions and information

- Water-Year Summary reports summarize a year of hydrologic data in a printer-friendly format.
- For water years 2014 onward, choose a water year and parameter and an on-demand Water-Year Summary report will be generated.
- For water years 2006 through 2013, choose a water year and an Annual Water Data Report will be provided.
- For water years 2005 and earlier, Annual Water Data Reports may be available in digital format.
- Not all sites have reports available for every water year.

Available Parameters

- 00060 Discharge(Mean)
- 00065 Gage height(Max.,Min.)

Select a water year 2006

Water year 2006: 2005-10-01 to 2006-09-30

GO

11109000 Santa Clara River near Piru, CA

LOCATION - Lat 34°24'13", long 118°44'18" referenced to North American Datum of 1927, Ventura County, CA, Hydrologic Unit 18070102, in San Francisco Grant, on right downstream bank, on private property owned by Newhall Farms, 0.1 mi south of Highway 126, 3 mi east of Piru, and 8 mi west of intersection of Highway 126 and Interstate Highway 5.

DRAINAGE AREA - 645 mi².

REVISIONS HISTORY - WDR-US 2009: 1997 (maximum discharge).

SURFACE-WATER RECORDS

PERIOD OF RECORD - October 1927 to September 1932, October 1996 to current year.

GAGE - Water-stage recorder and crest-stage gage. Elevation of gage is 725 ft above NGVD of 1929, from topographic map.

REMARKS - Base flow affected by pumping from wells along stream for irrigation. Flow partly regulated since January 1972 by Castaic Lake (station 11108133), capacity, 323,700 acre-ft. Imported water from California Water Project stored and released at Castaic Dam. See schematic diagram of Santa Clara River Basin available from the California Water Science Center.

EXTREMES FOR PERIOD OF RECORD - Maximum discharge, 32,000 ft³/s, Jan. 10, 2005, estimated based on hydrographic comparison and Castaic Lake release records, maximum gage height, 14.54 ft, Mar. 20, 2011; no flow for many days during the summers of 1929-32.

WATER-QUALITY RECORDS

PERIOD OF RECORD - Water year 2005 to 2015. WATER TEMPERATURE: Water year 2005-07 (discontinued). SEDIMENT DATA: Water year 2005 to 2015 (October 2007 to April 2015 (seasonal) (discontinued).

PERIOD OF DAILY RECORD - SUSPENDED-SEDIMENT DISCHARGE: October 2004 to September 2007 (discontinued).

EXTREMES FOR PERIOD OF DAILY RECORD - SEDIMENT CONCENTRATION: Maximum daily mean, 65,100 mg/L, Jan. 10, 2005; minimum daily mean, 9 mg/L, Aug. 21, 2006. SUSPENDED-SEDIMENT LOAD: Maximum daily, 3,490,000 tons, estimated, Jan. 10, 2005; minimum daily, 0.56 ton, Aug. 21, 2006.

OpenAI and ChatGPT

Pros:

- Interpetable prompt engineering.
- Easy manipulation of criteria.
- Simple setup with little to no parameter tuning.

Cons:

- OpenAI charges requests to their API.
- Prompt engineering is inconsistent and unreliable.
- Interpretation of gage selection is not as clear.

NLP Model and LightGBM

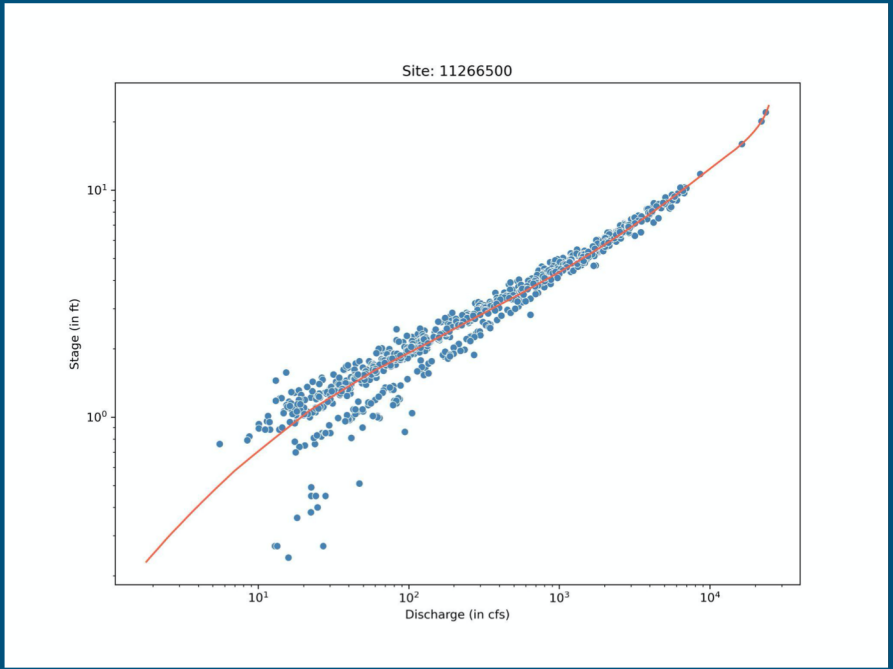
To label our data on water-year summaries we used LightGBM (light gradient-boosting machine) - a highly efficient gradient-boosting framework that is based on decision tree algorithms. LightGBM achieves high performance with sparse data, which makes the algorithm ideal for word feature vectors and matrices.

For our model, LightGBM outperformed similar frameworks such as XGBoost and had a superior runtime against SVM classifiers. Ensemble classifiers and the Multinomial Naive Bayes classifier yielded similar accuracy, however, both of these methods had a tendency to score low recall on possibly eligible gauges.

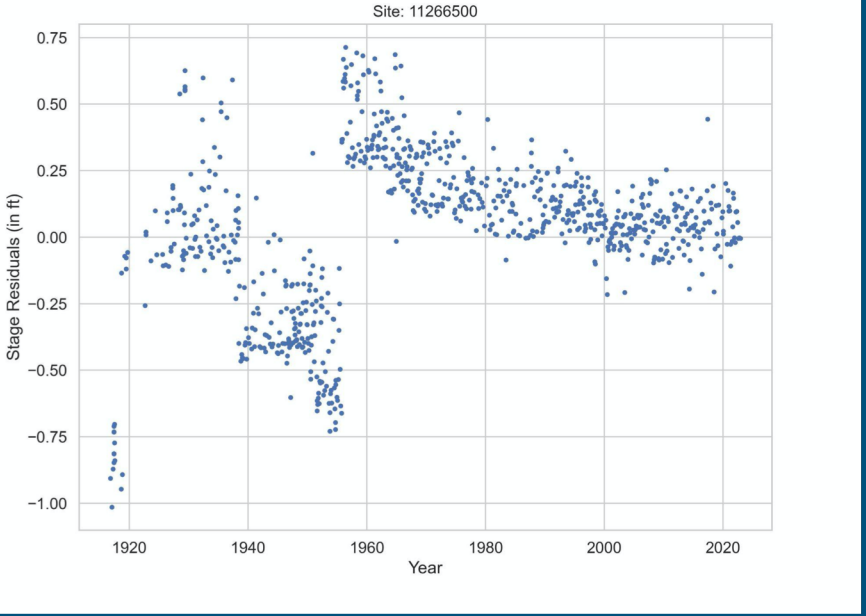
Stage-Discharge and Rating Curves

The relationship between stage height and discharge is taken from field measurements data that we scraped, and log-scaled before plotting.

Rating Curves are calculated based on the stage-discharge data using the Loess Curve method to make the best fit and then posted on the USGS website. The curve changes over time when more data is recorded.



Stage Residuals Time Series

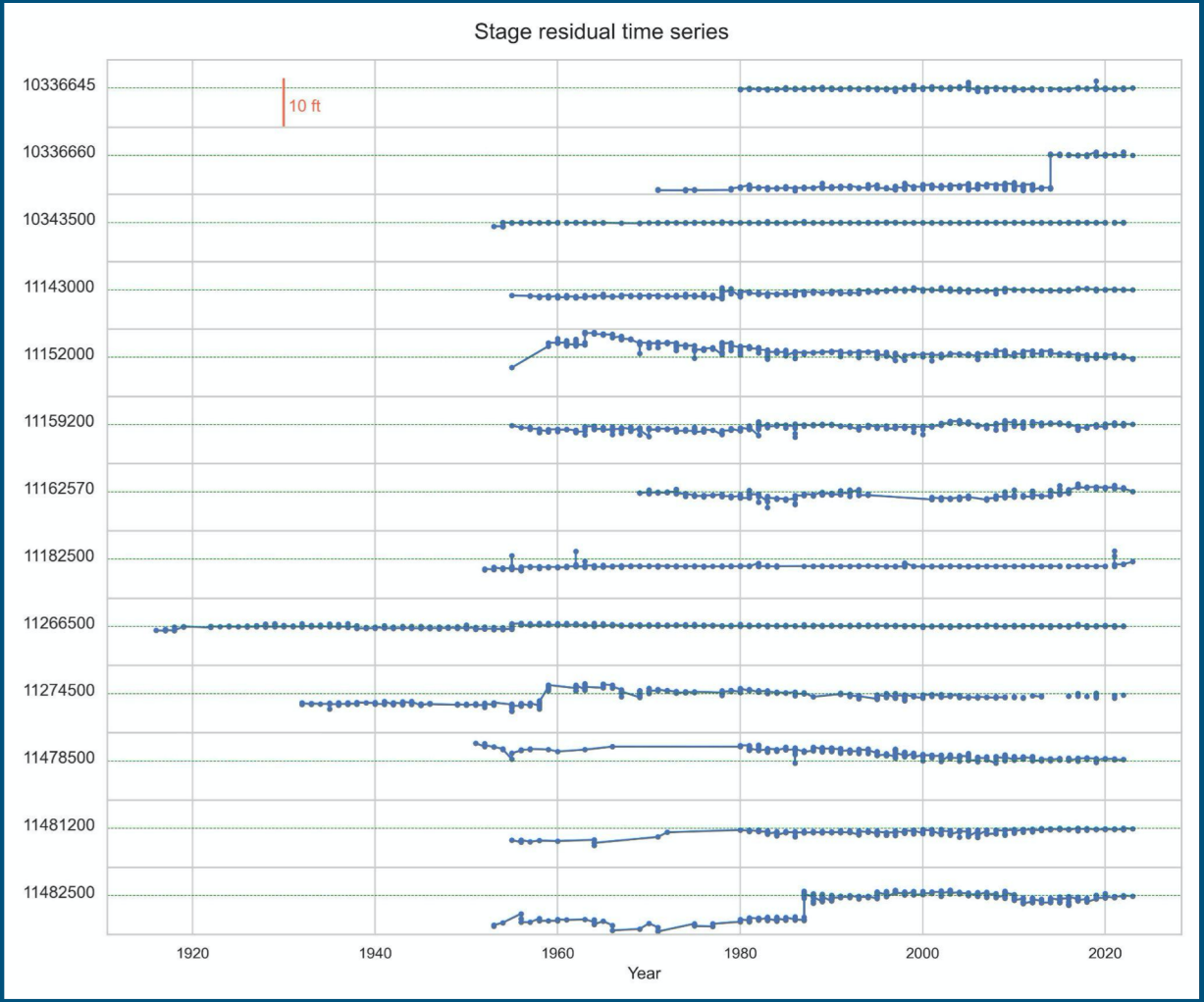


Finding the distance between the field measurement data point to the rating curve involved interpolation to fill in the missing values for height and selecting monotonically increasing discharge.

Then, we calculated stage residuals for each gauge by finding the difference between original stage height and an interpolated height.

Stage Residuals Time Series

For a better analysis and to be able to see how each gauge's residuals compare to other gauges, we plotted a time series with all gauges' residuals next to each other. The y-range has been scaled according to the min and max of all residuals.

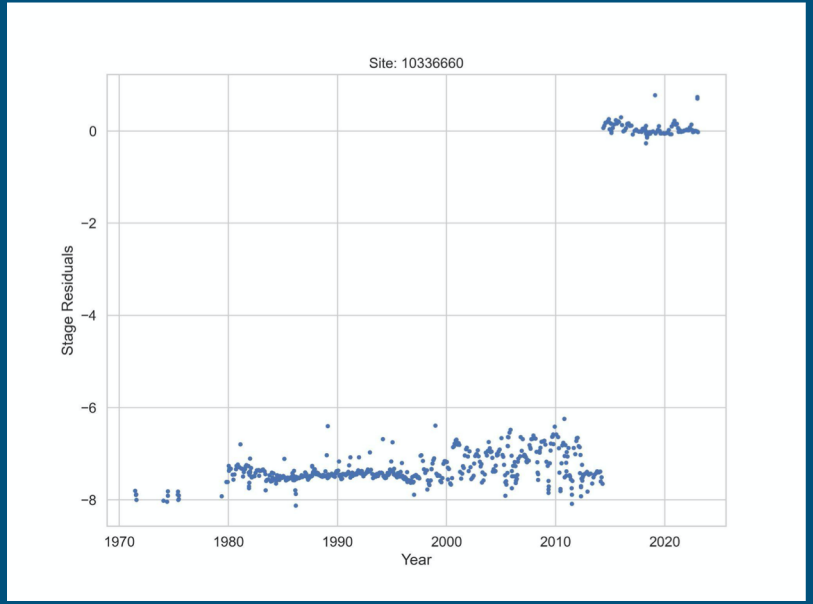


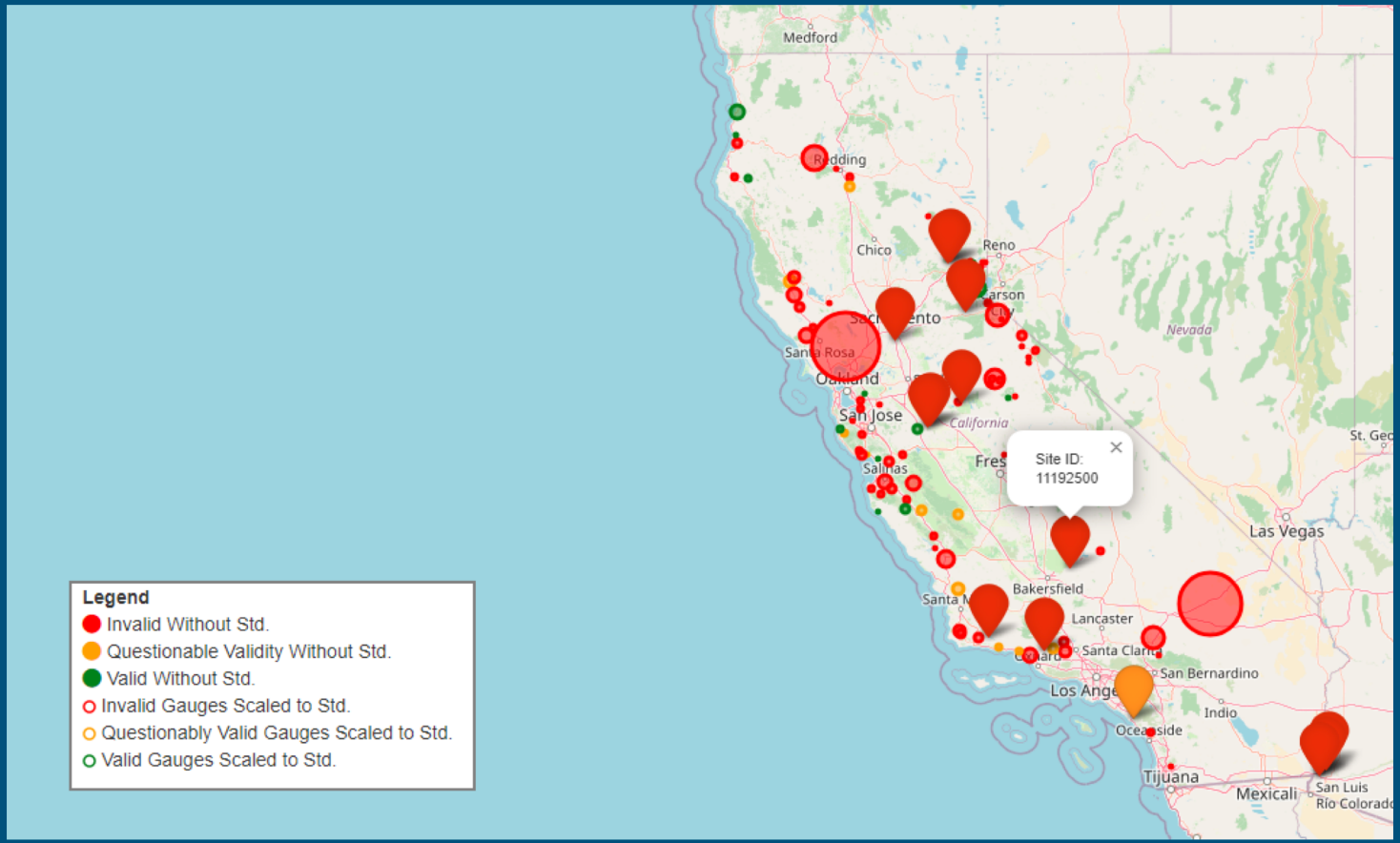
Stage Residual Standard Deviation

Stage-residual standard deviation is calculated by taking the standard deviation of all the stage-residual values.

Given a USGS rating curve, each gage has a standard deviation value representative of the variability in the stage-residuals of that gage.

Large standard deviations are indicative of a datum shift or invalid sites.





Interactive Folium Map

Conclusions

Deliverables:

Our work provides the client with a flexible framework for processing USGS data. It allows easy modifications to accommodate changes in the project scope without affecting the code and processes. Our trainable model, graphics generation, and data processing are applicable to all input USGS stream gages.

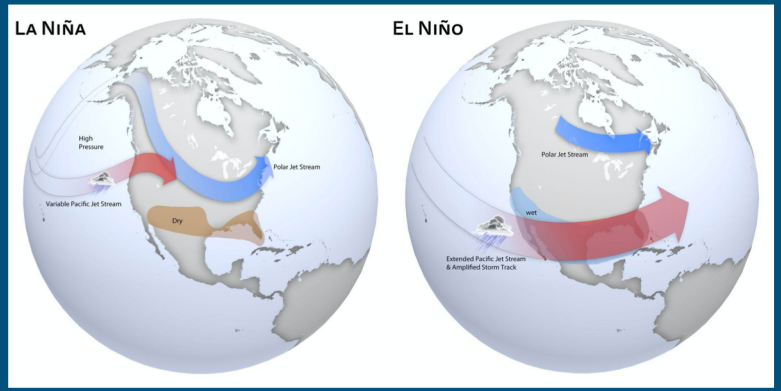
Interesting Findings:

We saw that the gages that our model labeled as valid also were gages with very low standard deviation, likewise, gages with higher standard deviation were usually labeled as invalid.

Regulated rivers are more variable than natural ones: gauges with dams, reservoirs, and other regulations are changing more than the channels that respond to natural events.

Future Work

This work can be easily replicated for all states such that a thorough study of stream gage sites and stream flow can be done in the United States. La Niña and El Niño cycles, extreme weather events, and forest fires can be referenced in channel response to see if there is correlation. Data can be compared with other studies on channel response to streamflow. Extensive collections of images and maps can be created for United States and published such that anyone can have access.



Thank you for
your time, feel free
to ask questions.