EXTREME EVENTS

ERT 474/574

Open-Source Hydro Data Analytics

Sep 22nd 2025

University at Buffalo The State University of New York

Logistics

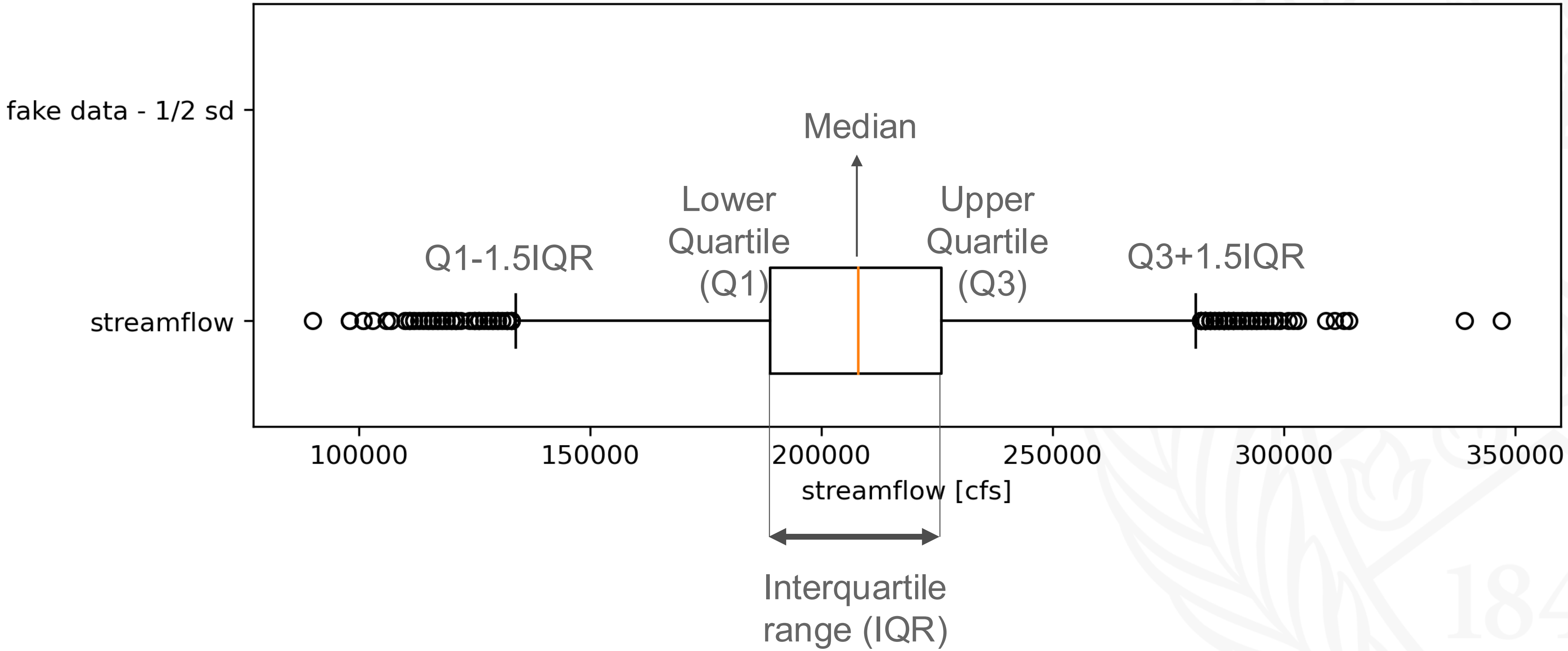
- We will have HW3 due next Wednesday (Oct 1st)
- HW1 grades were published.

Statistical methods are widely used in hydrologic modeling

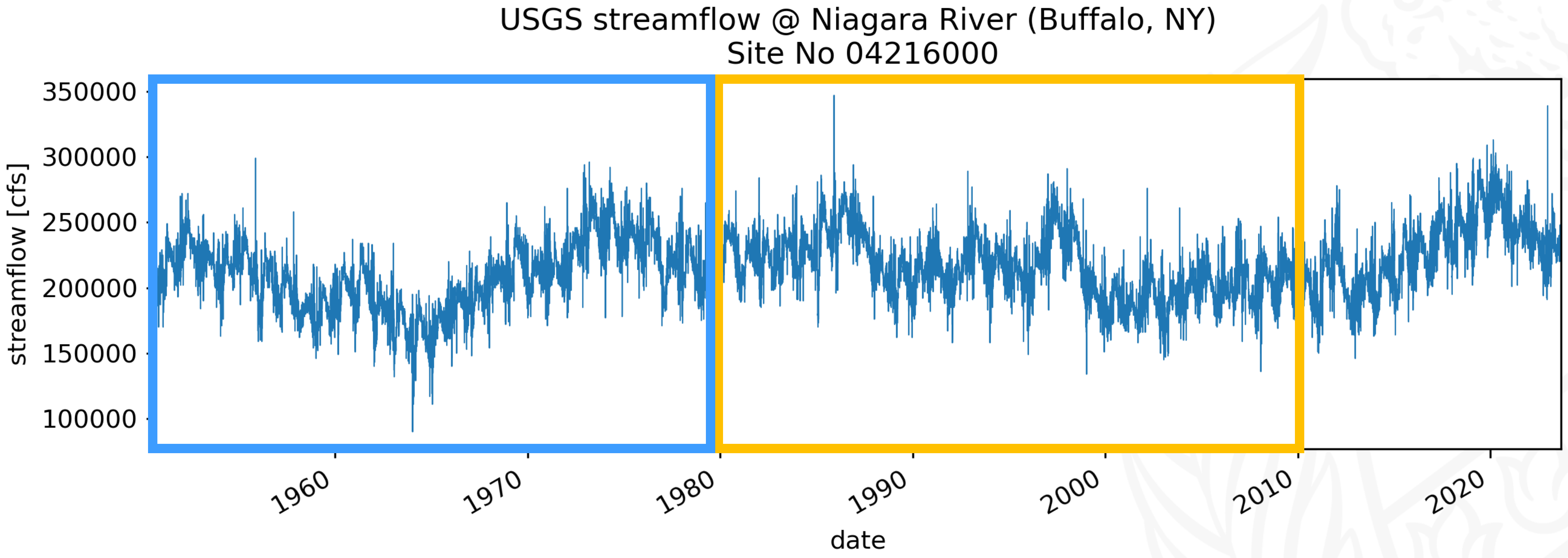
How do we do descriptive analysis when we get a data?

- Mean, variance, standard deviation (Box plot)
- PDF(Histogram), CDF (Quantile mapping), median (inter-quantile range)
- Extreme detection
  - z-score
  - 7Q10

Box-plot


fake data - 1/2 sd
streamflow
Median
Lower
Quartile
(Q1)
Upper
Quartile
(Q3)
Q1-1.5IQR
Q3+1.5IQR
100000
150000
200000
250000
300000
350000
streamflow [cfs]
Interquartile
range (IQR)

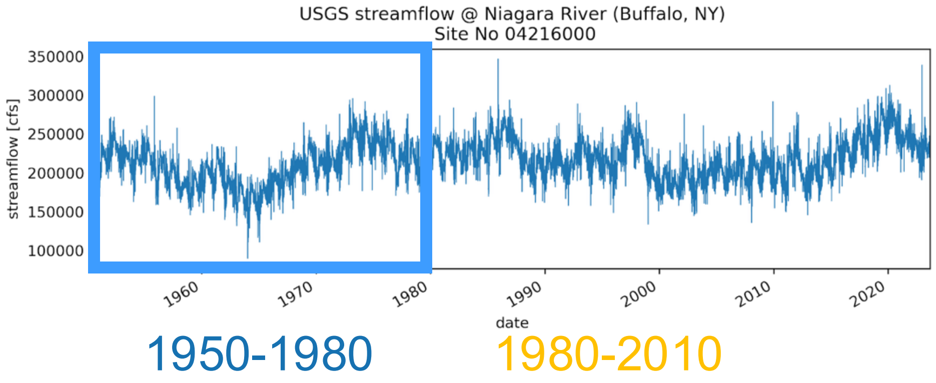
Use CDF to visualize the shift in hydrologic regimes

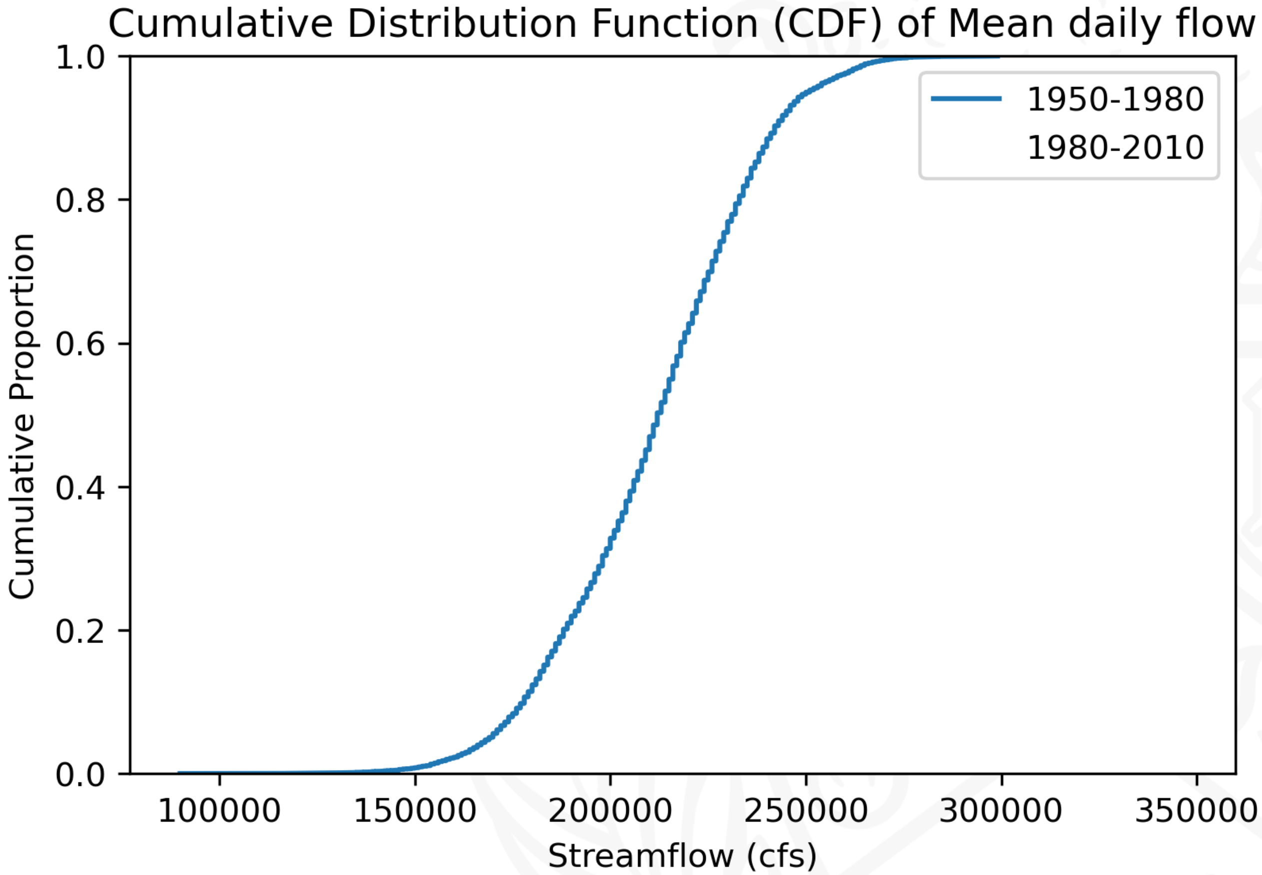


1950-1980

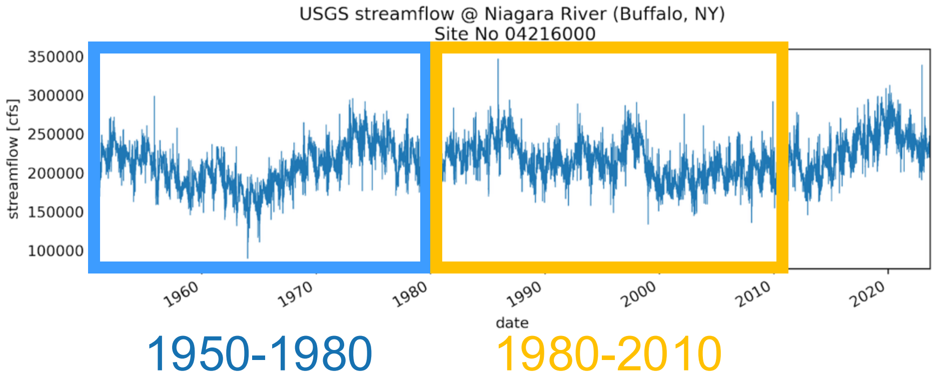
1980-2010

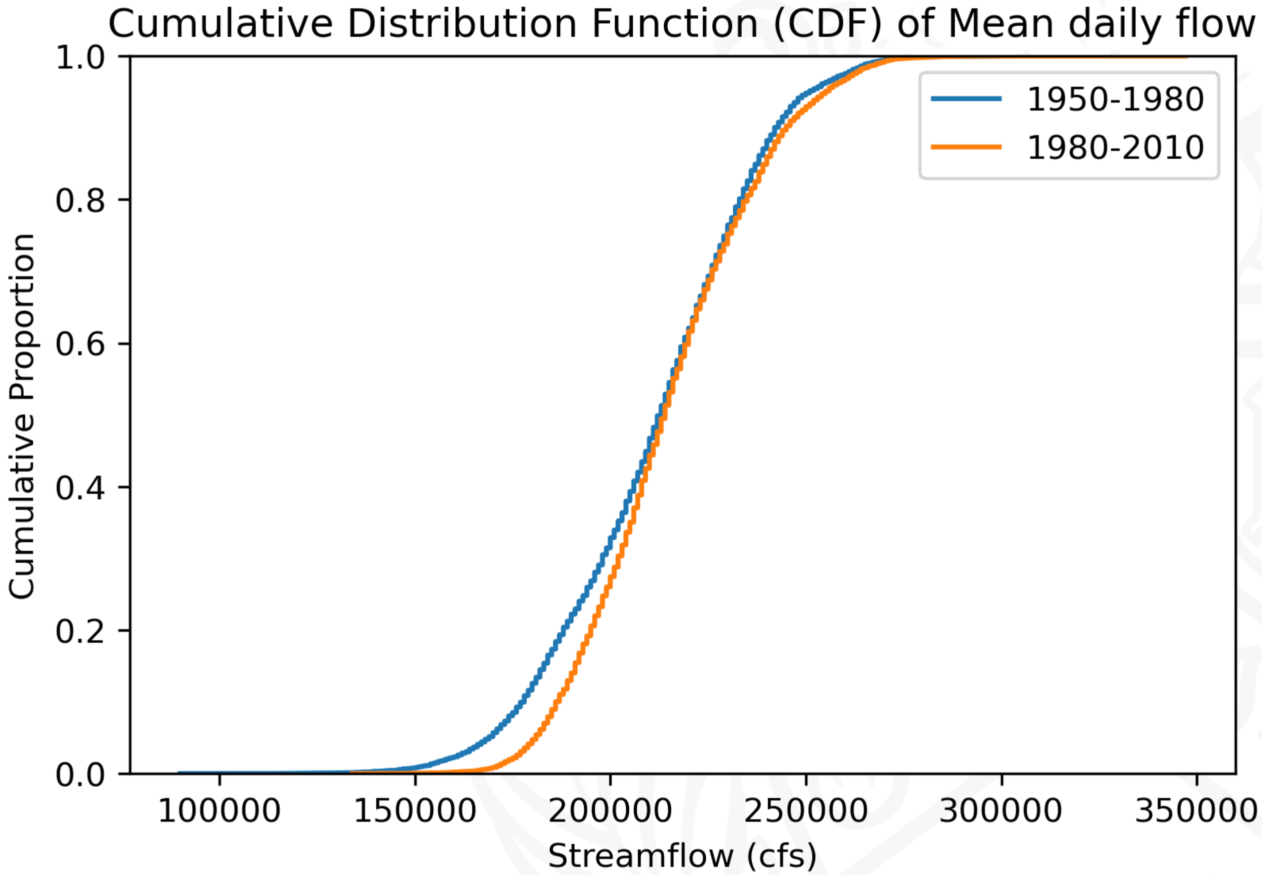
Use CDF to visualize the shift in hydrologic regimes





Use CDF to visualize the shift in hydrologic regimes





What information can we read from the plot in the right?

CDF for identifying floods?

New Data Reveals Hidden Flood Risk Across America

By Christopher Flavelle, Denise Lu, Veronica Penney, Nadja Popovich and John Schwartz June 29, 2020

Nearly twice as many properties may be susceptible to flood damage than previously thought, according to a new effort to map the danger.

Across much of the United States, the flood risk is far greater than government estimates show, new calculations suggest, exposing millions of people to a hidden threat — and one that will only grow as climate change worsens.

That new calculation, which takes into account sea-level rise, rainfall and flooding along smaller creeks not mapped federally, estimates that 14.6 million properties are at risk from what experts call a 100-year flood, far more than the 8.7 million properties shown on federal government flood maps. A 100-year flood is one with a 1 percent chance of striking in any given year.

What is a "100-year flood"?

Instead of the term "100-year flood" a hydrologist would rather describe *this extreme hydrologic event as a flood having a 100-year recurrence interval.*

Recurrence intervals

Recurrence intervals and probabilities of occurrences

Recurrence interval, years	Annual exceedance probability, percent
2	50
5	20
10	10
25	4
50	2
100	1
200	0.5
500	0.2

The term "100-year flood" is used to define a flow events that statistically has this same 1-percent chance of occurring.

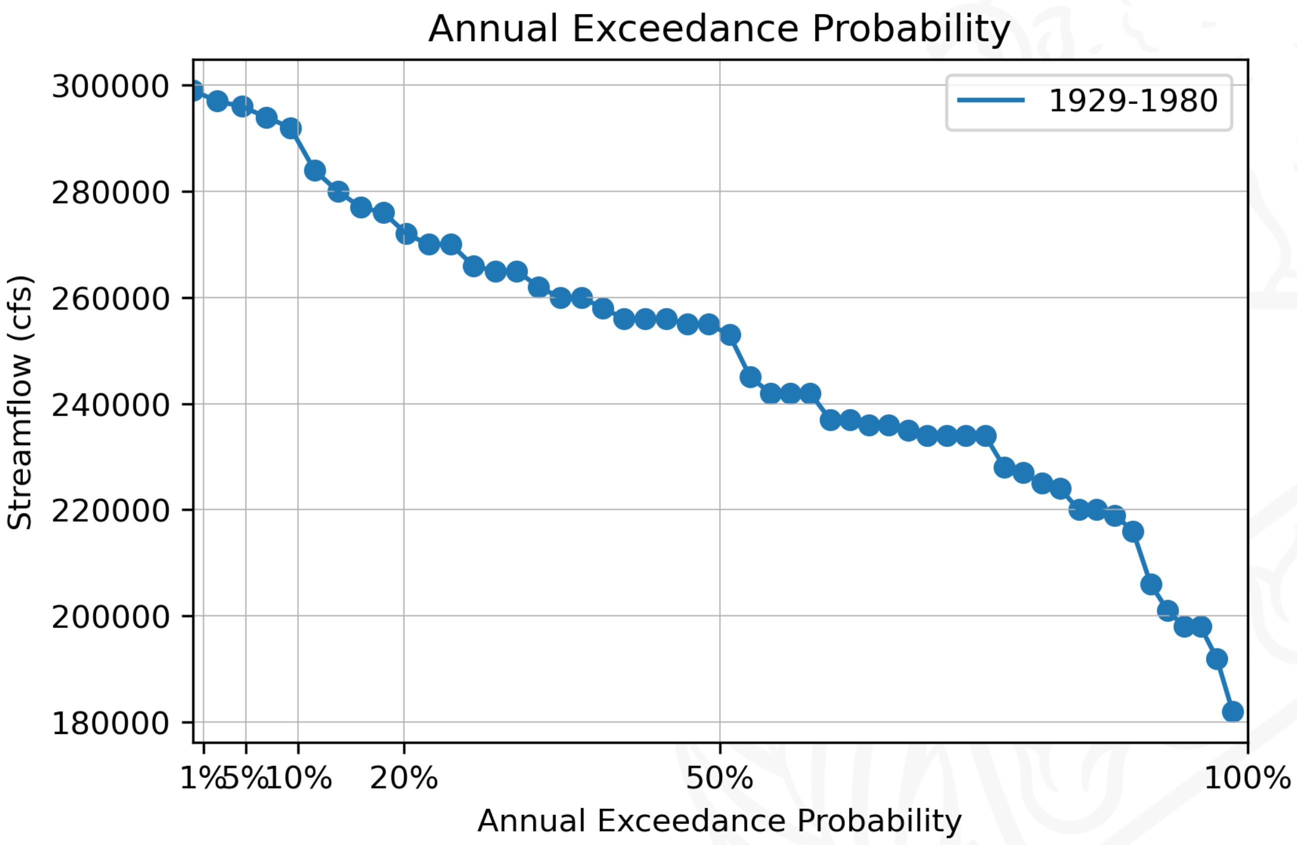
In other words, over the course of 1 million years, these events would be expected to occur 10,000 times.

But, just because it rained 10 inches in one day last year doesn't mean it can't rain 10 inches in one day again this year.

Annual exceedance probability (AEP)

- Step 1: Identify the peak flow for each year
- Step 2: Calculate the CDF for annual peak flow
- Step 3: AEP = 1 - CDF

If you were a hydrologist back in the 1980s, the construction of a bridge requires the information of 100-year flood

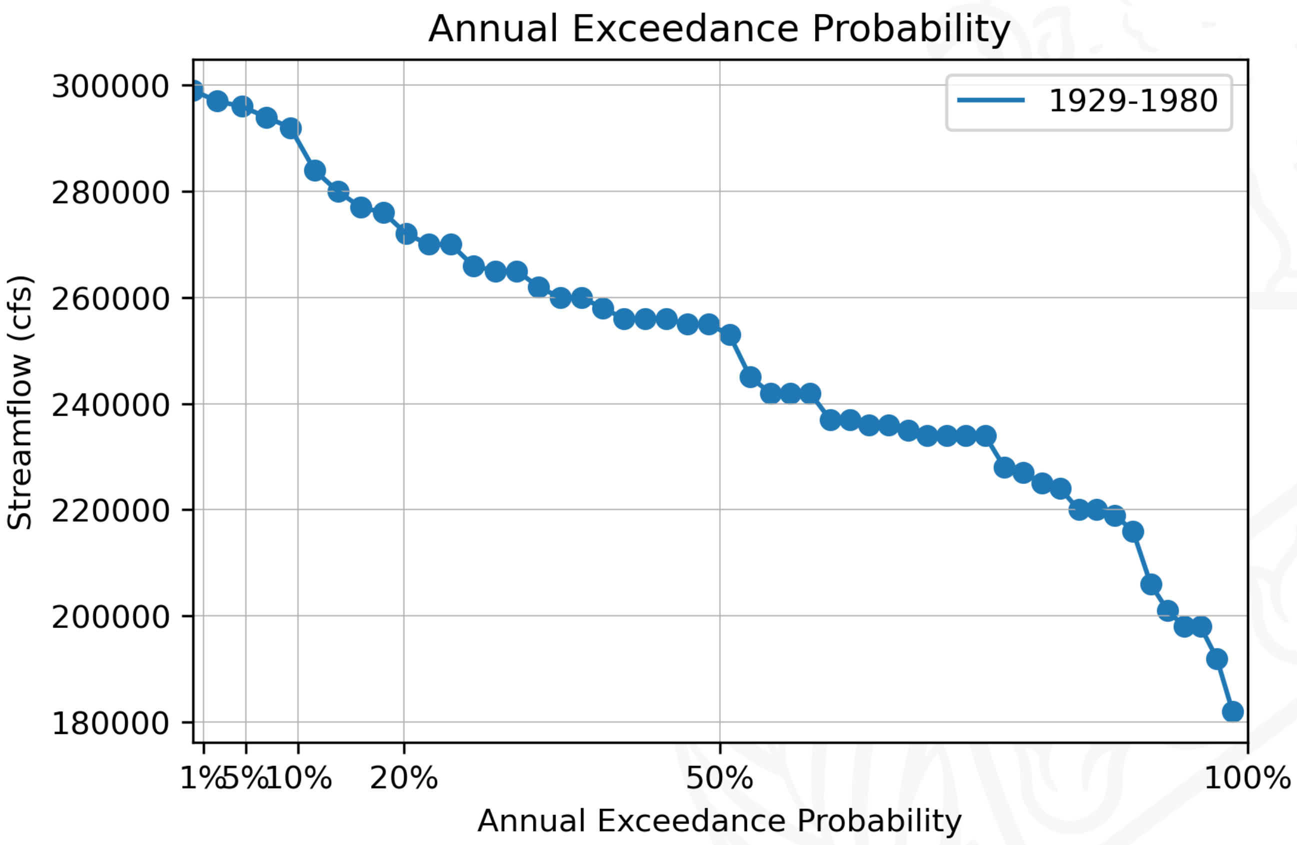


We can use the AEP plot on the right to retrieve the flood information, which corresponds to the value with **1% AEP!**

With more observations, it is important to revisit the flood designs

Now we are back to the 21st century, so we have observations for four more decades.

If we update the plots using more data, what have we observed?

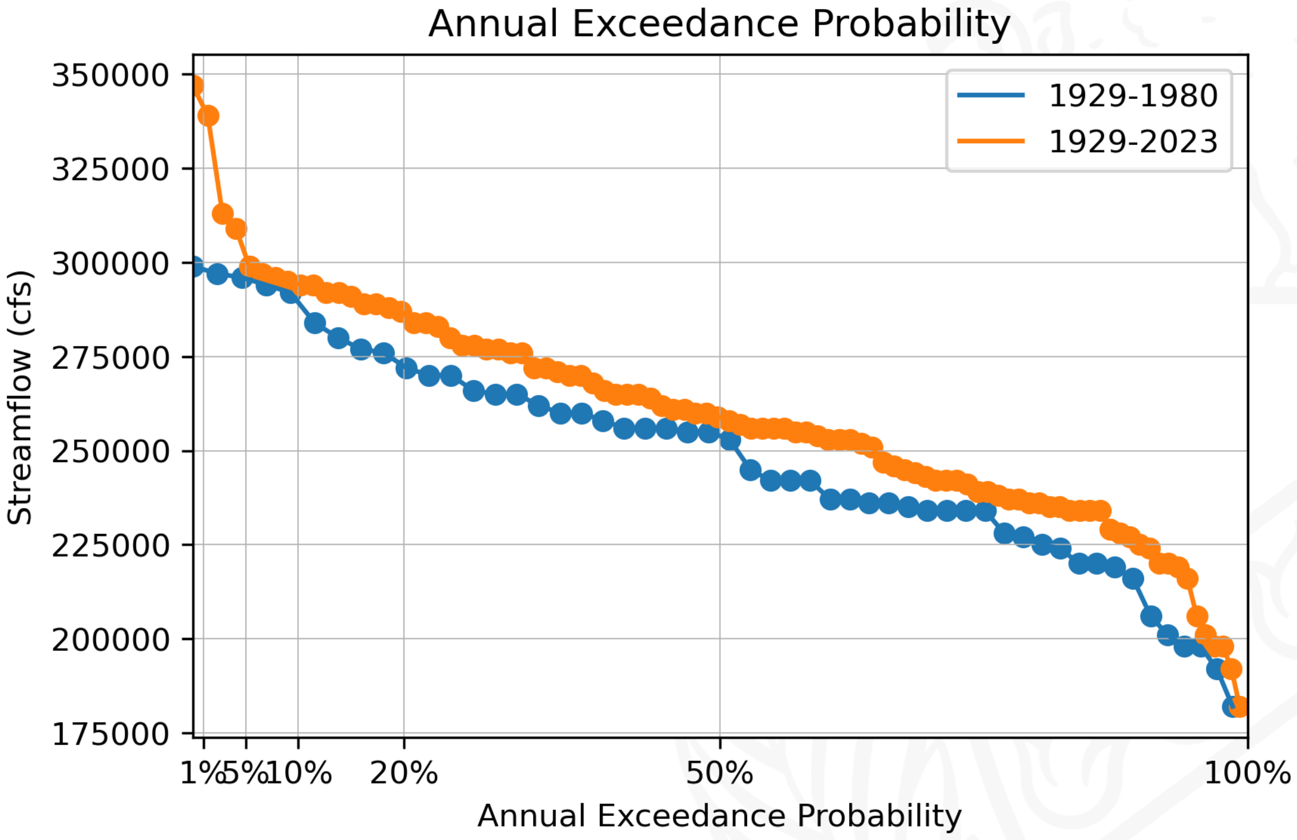


With more observations, it is important to revisit the flood designs

Now we are back to the 21st century, so we have observations for four more decades.

If we update the plots using more data, what have we observed?

The water volume of 100-year flood increased more than 10%!

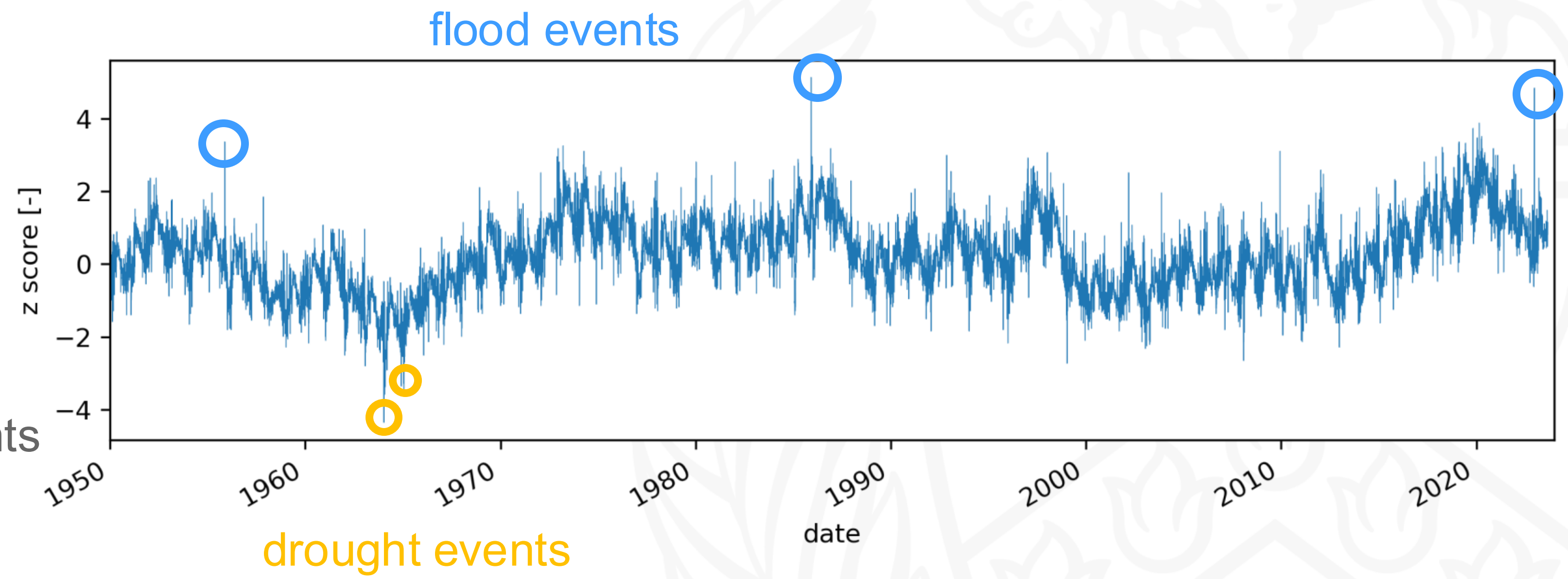


How can we more directly identify individual extreme events?

- **Z score**

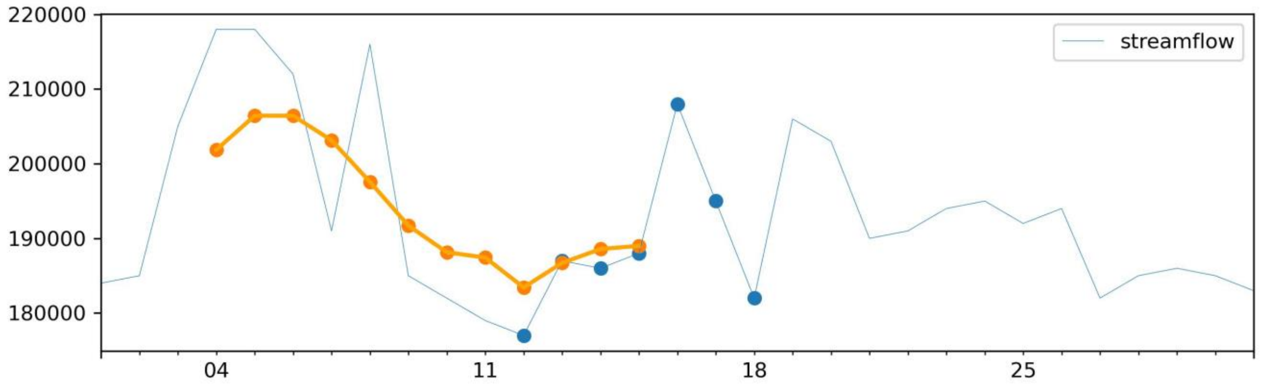
$$z = \frac{X_i - \bar{X}}{\sigma}$$

- Z scores will be assigned to every data points
- It is useful to identify extreme events across basins
- Not only useful for identify flood events, but also for low flow events



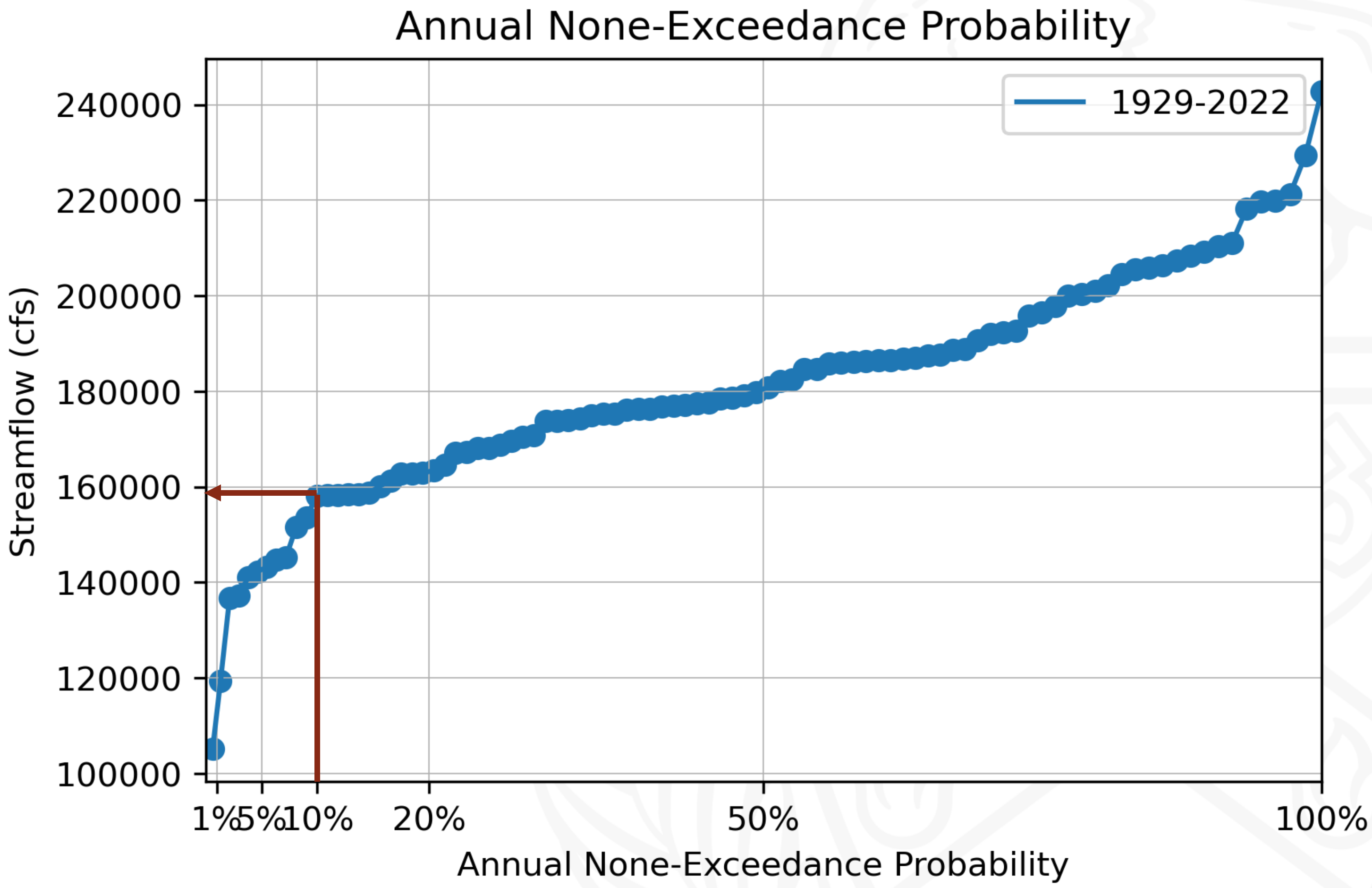
Statistical metrics for drought (low-flow events)?

- **7Q10**
  - The 7Q10 is the lowest 7-day average flow with a recurrence interval of 10 years.
- How do we calculate 7Q10?
  - Step 1: calculate the **7-day rolling average**



Statistical metrics for drought (low-flow events)?

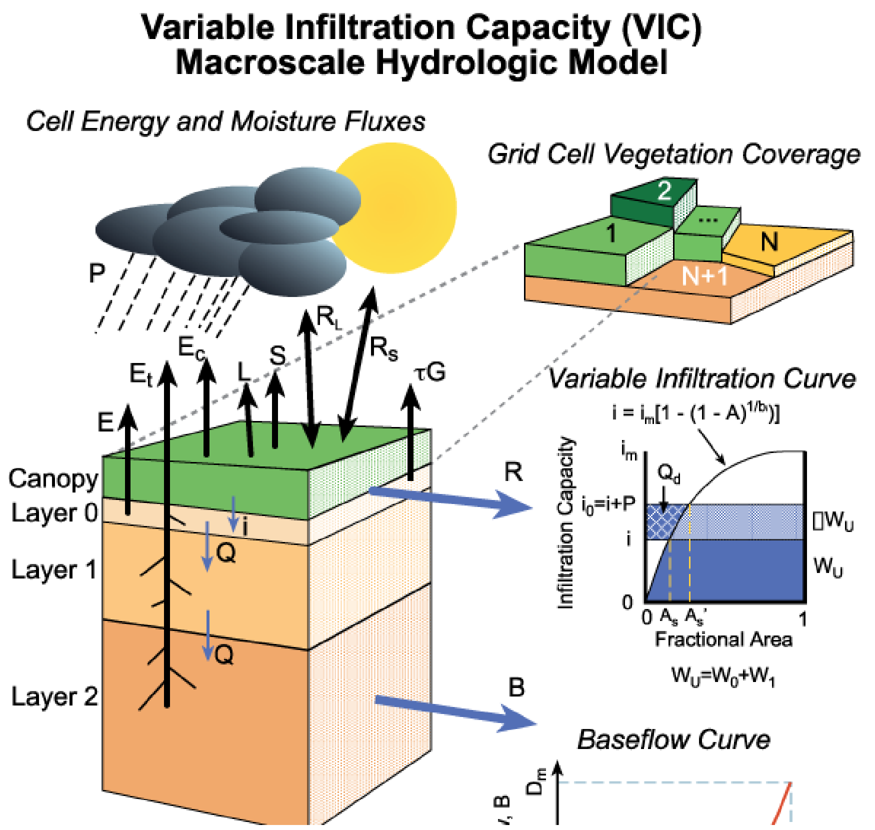
- **7Q10**
  - The 7Q10 is the lowest 7-day average flow with a recurrence interval of 10 years.
- How do we calculate 7Q10?
  - Step 1: calculate the **7-day rolling average**
  - Step 2: Find annual minimum 7-day average flow
  - Step 3: Annual Non-exceedance probability



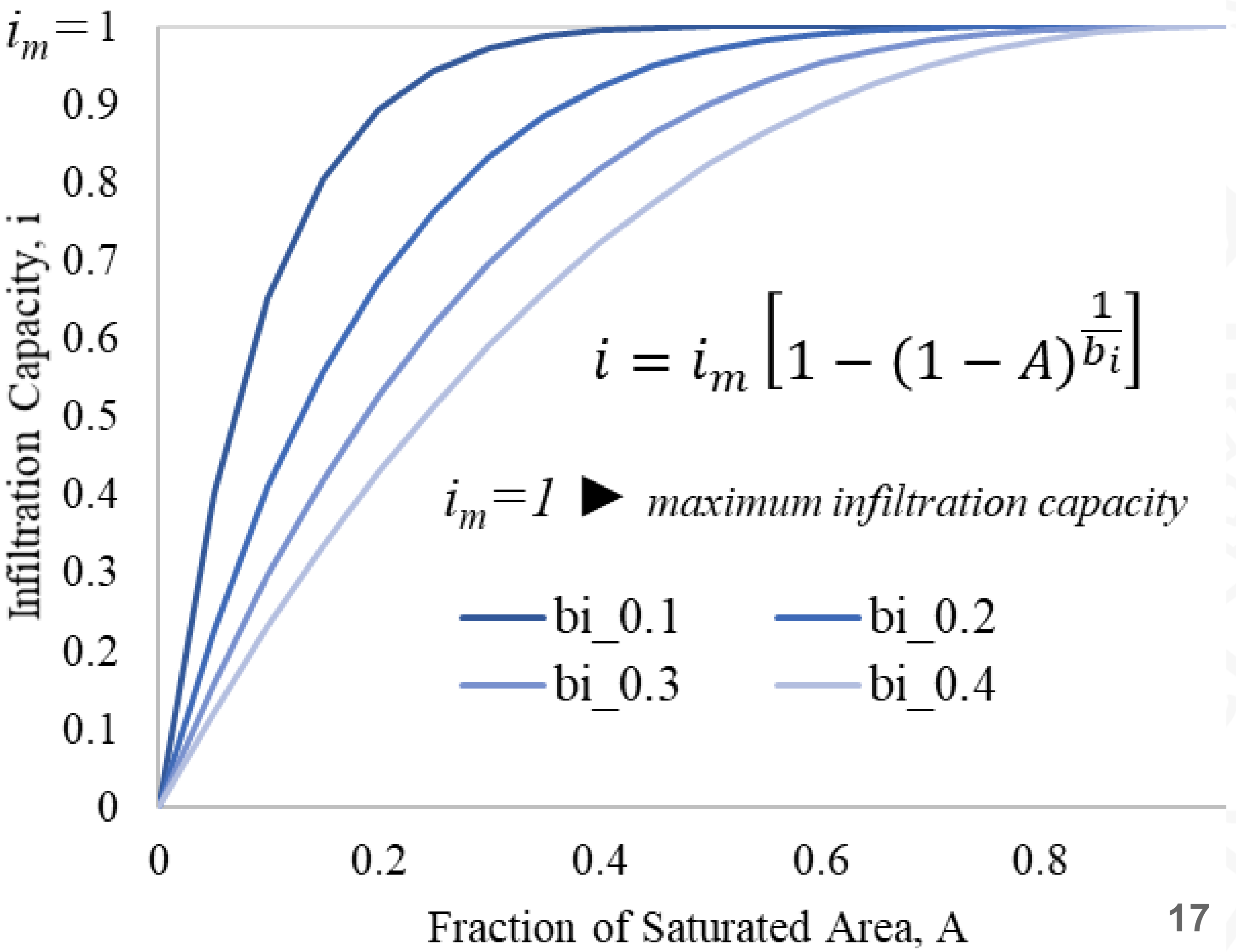
Comprehensive applications

Parameter Calibration of Hydrologic Models

What is a parameter?



The parameter b_i affects how fast the water can be infiltrated through the soil column



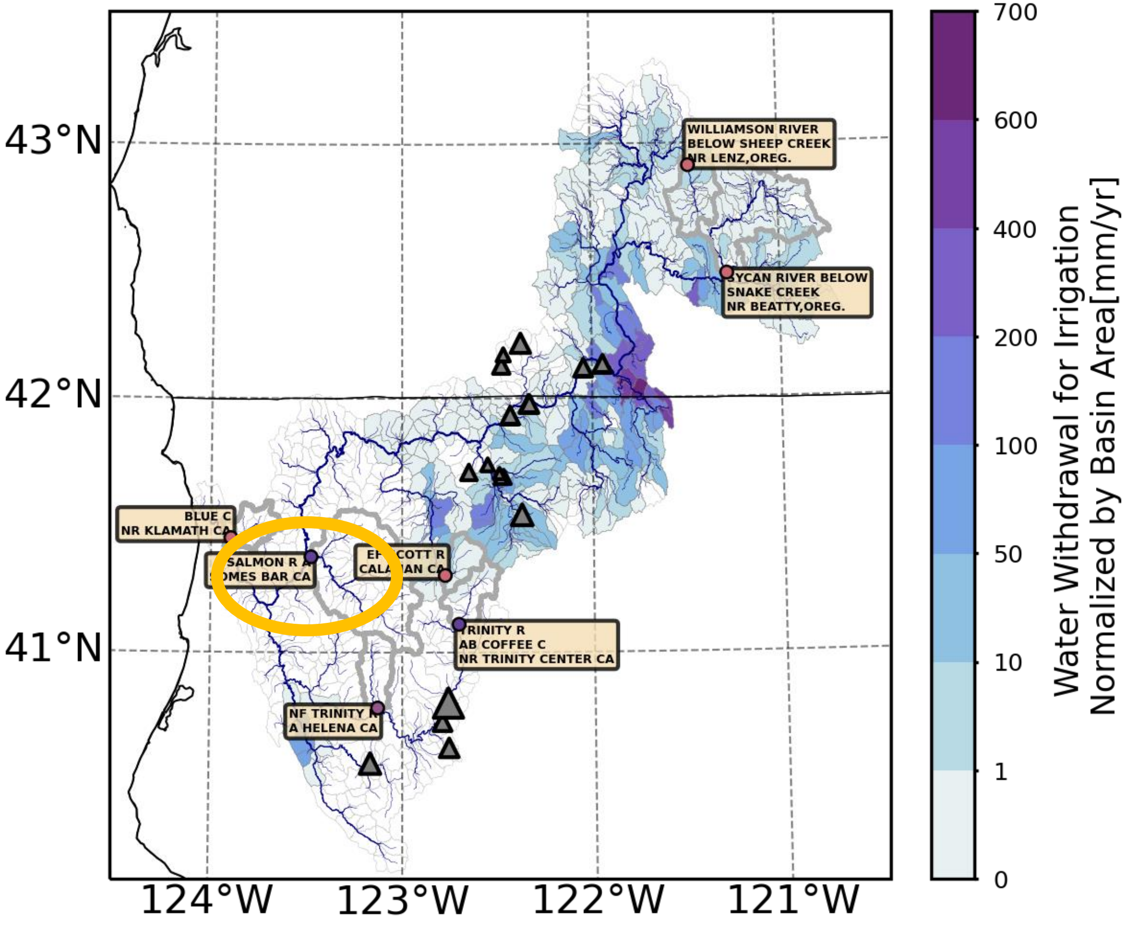
Steps for parameter calibration

- Step 1: Select observation sites and data
- Step 2: Model Setup
- Step 3: Selection of the objective function
- Step 4: Calibration
- Step 5: Evaluation of calibrated parameters

Steps for parameter calibration

- **Step 1: Select observation sites and data**
- Step 2: Model Setup
- Step 3: Selection of the objective function
- Step 4: Calibration
- Step 5: Evaluation of calibrated parameters

Model Evaluation using Statistics



This is the Klamath River Basin.

- The shaded colors denote the manual water withdrawal.
- The triangles denote the reservoirs.
- The dots with yellow text box denote USGS observation sites.

Hydrologic models usually cannot represent human interference in the river systems, and you will need to select one USGS site for tuning model parameters. Which sites will you choose?

Steps for parameter calibration

- Step 1: Select observation sites and data
- Step 2: Model Setup
- Step 3: Selection of the objective function
- Step 4: Calibration
- Step 5: Evaluation of calibrated parameters

Steps for parameter calibration

- Step 1: Select observation sites and data
- Step 2: Model Setup
- **Step 3: Selection of the objective function**
- Step 4: Calibration
- Step 5: Evaluation of calibrated parameters

Nash-Sutcliffe Efficiency (NSE)

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T} \left(Q_o^t - Q_m^t\right)^2}{\sum_{t=1}^{T} \left(Q_o^t - \overline{Q}_o\right)^2}$$

Purpose: Measures how well the model predicts observed values compared to the mean of observations.

Range:

- 1 = perfect match
- 0 = model is as good as using the mean
- < 0 = model is worse than using the mean

Sensitive to: High flows (outliers can dominate)

Interpretation: Good for assessing overall fit, but may mask poor performance in low flows.

Kling Gupta Efficiency (KGE)

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$$

Purpose: Decomposes model performance into three components:

- **r**: linear correlation (shape)
- **α**: variability ratio (spread)
- **β**: bias ratio (magnitude)

Range:

- 1 = perfect match
- Lower values indicate deviation in any component

Sensitive to: All aspects of the hydrograph (not just peaks)

Interpretation: More balanced; helps diagnose *why* a model is underperforming.

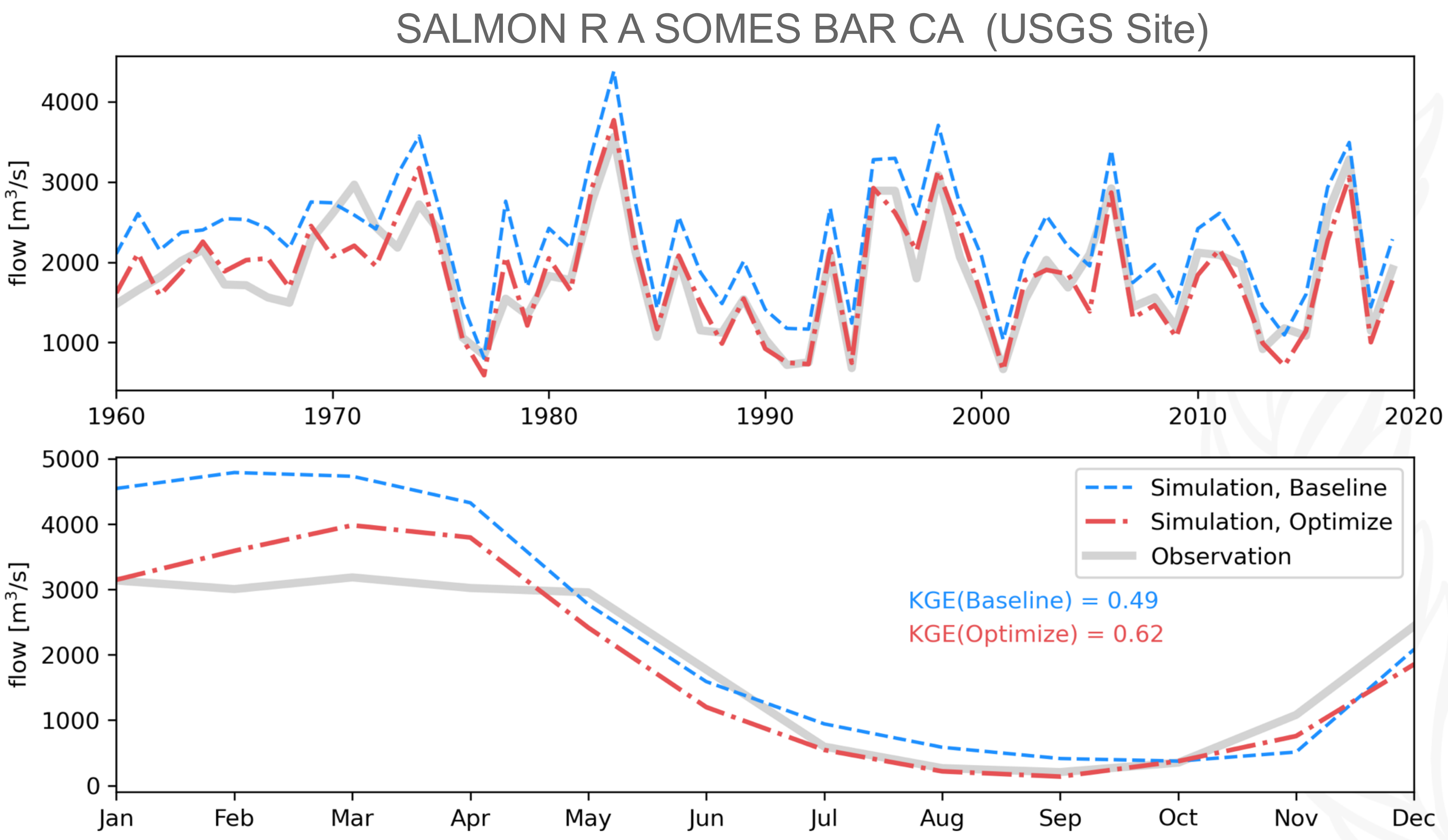
NSE versus KGE

Feature	NSE	KGE
Focus	Overall fit vs. mean	Fit decomposed into correlation, bias, variability
Sensitivity	High flows dominate	Balanced across flow regimes
Interpretability	Single score	Diagnostic (shows *why* model fails)
Range	$-\infty$ to 1	$-\infty$ to 1
Ideal Value	1	1
Common Pitfall	Can be misleading if low flows are poorly simulated	Requires more interpretation of components
Use Case	Quick performance check	Detailed model evaluation

Steps for parameter calibration

- Step 1: Select observation sites and data
- Step 2: Model Setup
- Step 3: Selection of the objective function
- Step 4: Calibration
- **Step 5: Evaluation of calibrated parameters**

Model evaluation



Except for the overall better KGE, did the calibration improve or worsen model performance?

Recap

- Basic statistics
  - Mean, variance, standard deviation
  - Boxplot (Interquartile range, IQR)
  - PDF, CDF, AEP (recurrence probability)
  - Identify extreme events (z score, 7Q10)
- Parameter calibration