Denison University
Data Analytics Program

Early Detection of Storm Driven Water Level Changes Using Sequential Analysis

DA 380 - Sequential Analysis and Applications
Instructor: Dr. Zhe Wang

Author: Minh Le

December 14, 2025

1 Introduction

1.1 Background and Objective

Coastal water levels are driven by predictable tidal cycles as well as atmospheric conditions such as wind, air pressure, and storms. Under typical conditions, tidal predictions describe water level variation accurately, but during extreme weather events, observed levels can deviate from predicted values due to storm surge and wind driven forcing. Between February 3 and February 5 2024, the San Francisco Bay Area experienced a strong atmospheric river event associated with an intensifying low pressure system, damaging winds, and heavy rainfall (National Weather Service, 2024). These conditions make this period well suited for examining storm related deviations in coastal water levels.

The objective of this study is to evaluate whether water level behavior in San Francisco exhibits statistically detectable anomalies or structural changes during this storm period. Instead of analyzing raw water levels, I focus on water level anomalies defined as the difference between observed measurements and NOAA predicted tide values. This transformation removes the dominant tidal signal and isolates behavior more directly influenced by meteorological forcing. Sequential analysis methods, including control charts and change point detection, are used to determine whether the observed anomalies represent short term fluctuations or sustained changes in the underlying process.

1.2 Data Source

All data used in this study were obtained from the National Oceanic and Atmospheric Administration Tides and Currents program through its public application programming interface (API). Water level observations and predicted tide values were collected from station 9414290 located in San Francisco, California for the period from January 29 to February 10 2024 at a 6-minute sampling interval. This window includes conditions before the atmospheric river event, the peak impacts observed between February 3 and February 5 2024, and several days of recovery afterward. Predicted tide values provided by NOAA are generated using harmonic analysis based on long term historical observations, which establishes a reference for normal tidal behavior against which storm related anomalies can be evaluated (NOAA Tides and Currents, 2024).

2 Data Exploration

2.1 Data Cleaning and Processing

After retrieval, the data were aligned by timestamp and merged into a single time series suitable for sequential analysis. Missing values within the series were handled using interpolation to preserve continuity while maintaining the temporal structure of the data. A water level anomaly variable was then constructed as the difference between observed water levels and predicted tides, removing the dominant tidal signal and isolating deviations associated with atmospheric forcing. To reduce short term noise, the anomaly series was smoothed using a centered rolling mean with a window size of 10 observations, corresponding to approximately 1 hour of data, and this smoothed anomaly serves as the primary signal for subsequent analysis.

2.2 Exploratory Data Analysis

After cleaning and processing, the final dataset contains 3,114 observations at a 6-minute sampling interval for the selected station and time window (Appendix A). This resolution is sufficient to capture storm driven changes in water level while preserving the sequential nature of the data.

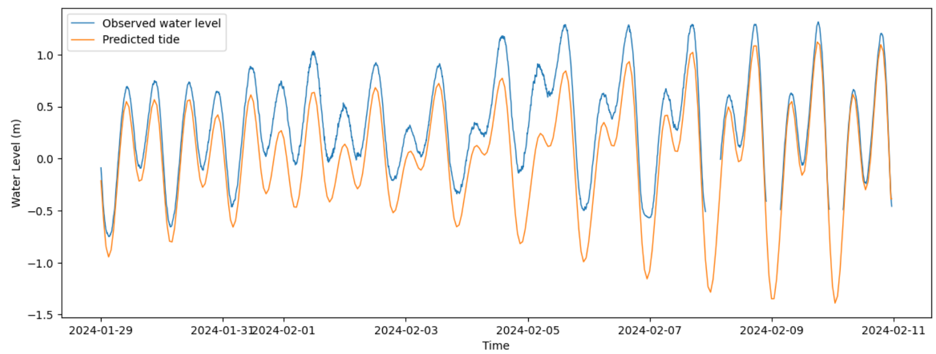

Figure 1: Observed vs. Predicted Water Levels - San Francisco

Over most of the study period, observed and predicted water levels track each other closely, indicating that tidal predictions capture the dominant astronomical signal under typical conditions (Figure 1). As the storm approaches, observed levels begin to exceed predicted tides, creating a widening gap that reflects additional forcing not explained by tidal dynamics. This behavior motivates the use of water level anomaly as the primary variable of interest.

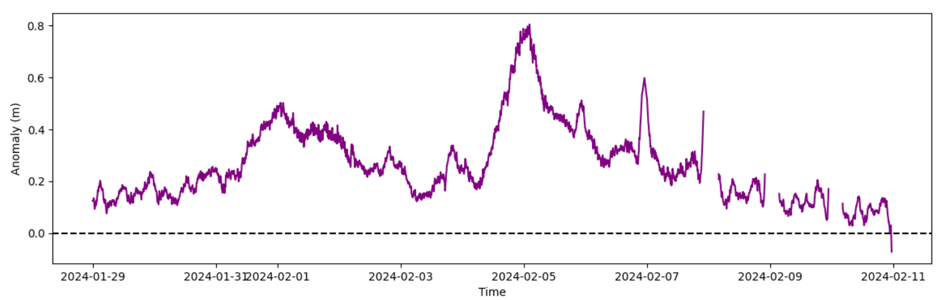

Figure 2: Water Level Anomaly (Observed – Predicted)

The anomaly series highlights this effect more clearly (Figure 2). Rather than oscillating around zero, anomaly values rise and remain elevated during the core storm period. The persistence of this elevation points to a temporary shift in the underlying process rather than isolated extremes, making the series well suited for sequential monitoring and change point detection.

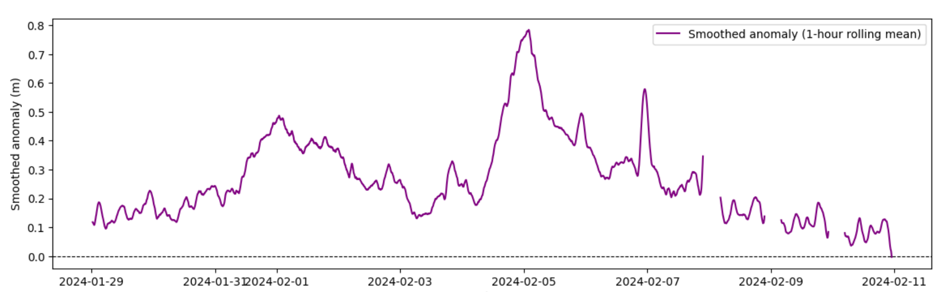

Figure 3: Smoothed Water-Level Anomaly (Observed - Predicted)

Figure 3 shows the anomaly after applying a centered rolling mean, which reduces short term variability while preserving the timing and structure of the storm signal. The main rise and decline become easier to distinguish, and baseline behavior before and after the event is more clearly defined. This supports the use of the smoothed anomaly as the primary input for the analysis in the following sections.

3 Methods and Results

3.1 Control Charts

Control charts provide a direct way to evaluate whether observed anomalies during the storm exceed what would be expected under typical conditions. In this study, I use them to identify unusually large deviations and to differentiate isolated extremes from longer-lasting abnormalities in water level behavior.

3.1.1 Shewhart Control Chart

The Shewhart control chart evaluates each observation independently by comparing it to fixed control limits derived from an in control distribution. Let X_t denote the smoothed water level anomaly at time t. The center line is defined as the historical mean μ, and the upper and lower control limits are given by:

$$UCL = \mu + 3\sigma, \quad LCL = \mu - 3\sigma$$

where σ is the historical standard deviation. Observations that fall outside these limits are flagged as out of control, indicating the possible presence of special cause variation rather than common background variability.

To estimate the in control parameters, I constructed a historical baseline using anomaly data from January 29 to February 2 across multiple years from 2020 to 2024. This avoids relying only on early 2024 data, which may already reflect storm related effects, and provides a more stable estimate of typical seasonal behavior.

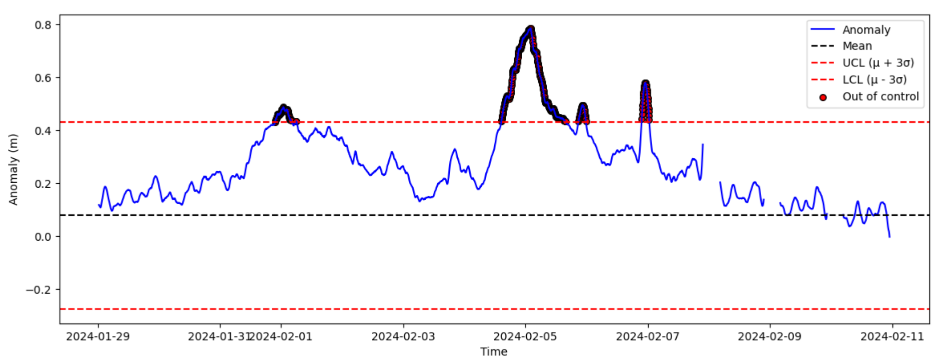

Figure 4: Shewhart Control Chart for Smoothed Water Level Anomaly

Figure 4 shows that most observations remain within the control limits prior to the storm, indicating stable baseline behavior. The chart also flags an out of control signal around February 1, before the peak impacts, which suggests that elevated water level anomalies began to emerge ahead of the main storm period. As the storm intensifies, several additional points exceed the upper control limit, corresponding to the largest anomalies observed during the event. Overall, the Shewhart chart effectively highlights extreme deviations but remains most sensitive to large and abrupt changes rather than gradual shifts.

3.1.2 Cumulative Sum (CUSUM) Control Chart

While the Shewhart chart focuses on individual observations, the CUSUM control chart is designed to detect small but persistent shifts by accumulating deviations over time. Let μ denote the target mean of the process and let k be a reference value that determines the size of shift of interest. The positive and negative CUSUM statistics are defined recursively as:

$$C_t^+ = \max\left(0, C_{t-1}^+ + (x_t - \mu - k)\right)$$

$$C_t^- = \max\left(0, C_{t-1}^- - (x_t - \mu + k)\right)$$

where k is a reference value controlling sensitivity. A signal occurs when either statistic exceeds a decision threshold h.

In this analysis, the reference value was set to $k = 0.25\sigma$, reflecting a focus on detecting sustained increases rather than isolated spikes. The decision threshold h was chosen using a short baseline period from February 1 to February 2, with h set to 1.25 times the maximum observed CUSUM magnitude during this baseline. This choice reduces false alarms while remaining sensitive to persistent deviations.

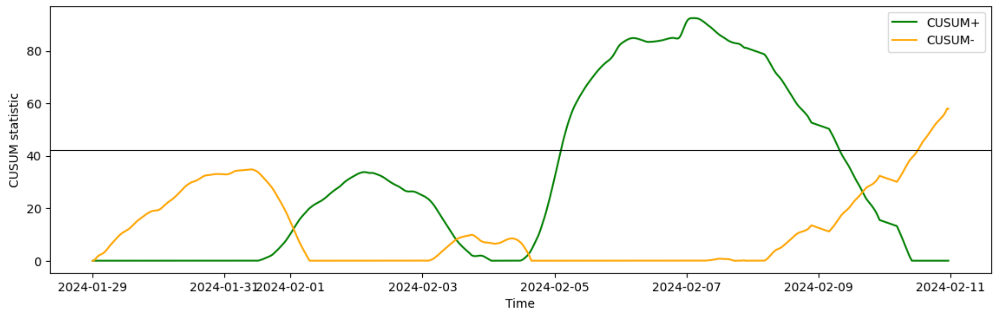

Figure 5: CUSUM Control Chart for Smoothed Water Level Anomaly

From Figure 5, the CUSUM statistics remain near zero early in the period but increase sharply as the storm intensifies, with the positive CUSUM exceeding the threshold and signaling a sustained upward shift in the anomaly. Unlike the Shewhart chart, the CUSUM continues to reflect accumulated evidence even after individual observations decrease, highlighting its effectiveness in detecting persistent storm driven changes.

3.2 Change Point Detection

I now apply change point detection methods to examine whether the atmospheric river event corresponds to a structural shift in the water level anomaly series rather than a collection of independent deviations. These methods allow me to identify changes in the statistical behavior of the series and to characterize how water level dynamics evolve before, during, and after the storm.

3.2.1 Frequentist Change Point Detection

For the frequentist change point analysis, I focus on detecting changes in the mean of the smoothed water level anomaly series. The underlying assumption is that the signal can be approximated by a sequence of segments, where each segment has a constant mean but the mean may shift at unknown time points. Formally, the observations are modeled as:

$$y_t \sim \mathcal{N}(\mu_i, \sigma^2),$$

where μ_i denotes the mean within segment i, and change points occur when the mean shifts from one segment to another.

I apply the PELT algorithm, which uses dynamic programming and pruning rules to find the exact global optimum effectively. The Bayesian Information Criterion (BIC) penalty is used to discourage over segmentation and favor a conservative solution.

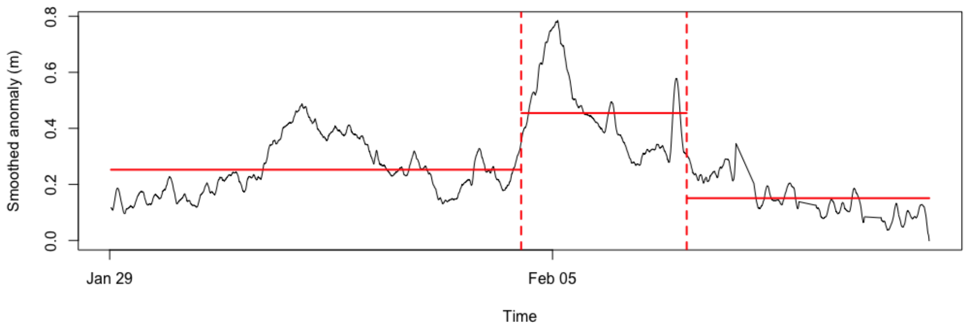

Figure 6: Mean Change Detection

Figure 6 shows the detected change points and segment means. The results indicate a clear increase in the mean anomaly during the storm period, followed by a decrease as conditions recover. The timing of these changes aligns with the onset and decline of the atmospheric river event, supporting the interpretation that the storm introduced a temporary shift in water level behavior.

3.2.2 Bayesian Change Point Detection

Bayesian online change point detection models change points probabilistically and updates beliefs sequentially as new data arrive. The key quantity is the run length r_t, defined as the number of observations since the most recent change point. At each time step, the method updates the posterior distribution:

$$P(r_t \mid x_{1:t})$$

which reflects how long the current regime is believed to have persisted given the observed data. The smoothed anomaly series is standardized prior to applying the model. A constant hazard function

$$H(r) = \frac{1}{\lambda}$$

is used to specify the prior probability that a change point occurs at any time step. This choice reflects an assumption that structural changes occur infrequently and are roughly independent of how long the current regime has lasted. I set $\lambda \approx 600$ to favor regime lengths spanning multiple days rather than short term fluctuations. A change point probability threshold of 0.85 is used, along with minimum separation and run length constraints, to limit sensitivity to noise.

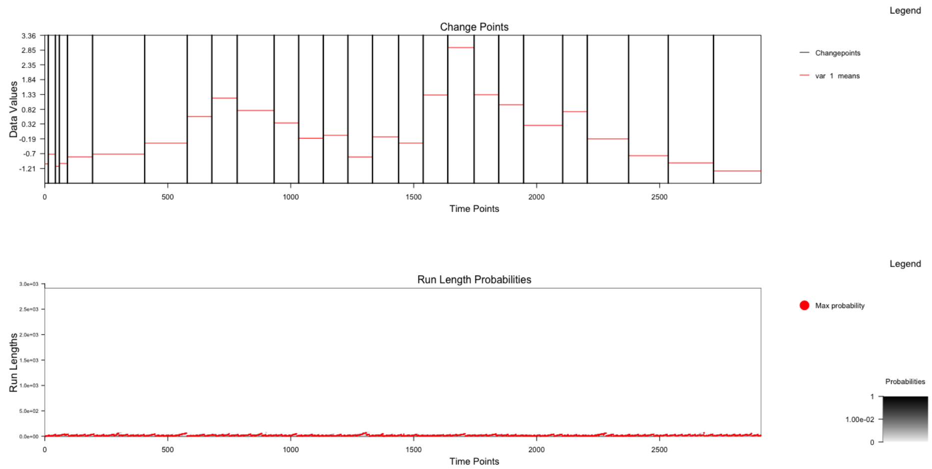

Figure 7: Bayesian Online Change Point Detection

Figure 7 shows that the Bayesian method produces many short segments with consistently low run length values. Rather than identifying a small number of persistent regimes, the model frequently resets, indicating high sensitivity to local variation in the anomaly series. As a result, the Bayesian approach does not clearly capture the storm as a single sustained change, likely due to the gradual evolution of the anomaly and the constant hazard assumption.

4 Discussion

This project shows how different sequential analysis tools highlight different aspects of storm driven water level behavior. Control charts and frequentist change point detection both provide clear and interpretable signals during the atmospheric river event, identifying periods when water level anomalies depart from typical behavior and when the underlying mean shifts. In contrast, the Bayesian online change point detection struggles to represent the storm as a single sustained regime, instead reacting to local fluctuations in the anomaly series. Comparing these results helped me better understand the strengths and limitations of each method. Threshold based tools such as Shewhart and CUSUM are effective for monitoring departures from a known baseline, while frequentist change point detection offers a clear retrospective segmentation of the event. The Bayesian approach, although appealing in principle, requires careful modeling choices to perform well in settings where changes evolve gradually rather than abruptly.

5 Limitations and Next Steps

One limitation of this analysis is that several modeling assumptions are only approximately satisfied. The frequentist change point method assumes constant variance within segments, while the anomaly series exhibits moderate variability changes during the storm period. Although smoothing helps stabilize the signal, remaining variance changes may influence the detected change points. In addition, control chart thresholds and CUSUM parameters were calibrated using historical baseline periods, which introduces some subjectivity and may limit generalizability to other locations or storm events.

Future work could address these limitations by applying models that allow both the mean and variance to vary over time or by using adaptive hazard functions within the Bayesian framework to better reflect evolving storm conditions. Extending the analysis to multiple coastal stations would also help assess the spatial consistency of storm driven water level responses and strengthen the robustness of the findings.

6 References

1. National Weather Service. 2024. Atmospheric river event impacts across the San Francisco Bay Area, February 3 to February 5 2024. National Oceanic and Atmospheric Administration. https://www.weather.gov/mtr/AtmosphericRiver-February_3-5_2024

2. National Oceanic and Atmospheric Administration Tides and Currents. 2025. Tides and Currents data and station information. https://tidesandcurrents.noaa.gov/stations.html

3. National Oceanic and Atmospheric Administration Tides and Currents. 2025. Harmonic constituents and tide prediction methods. https://tidesandcurrents.noaa.gov/about_harmonic_constituents.html

4. Wang, Z. 2025. Applied Sequential Analysis. Course notes, Denison University

7 Appendix

7.1 Appendix A: Table Summary

Table 1: Summary of variables used in the analysis

Variable	Description
datetime	Timestamp of each observation recorded in UTC time
water_level_m	Measured coastal water level at the San Francisco station in meters
predicted_tide_m	Tide level predicted by NOAA harmonic analysis in meters
anomaly_m	Difference between observed water level and predicted tide in meters
anomaly_smooth	Smoothed version of the anomaly using a rolling mean

7.2 Appendix B: Data and Code Repository

All data used in this study and the corresponding code implementations are available at the following GitHub repository: `https://github.com/lqminhhh/storm-water-level-sequential-analysis`