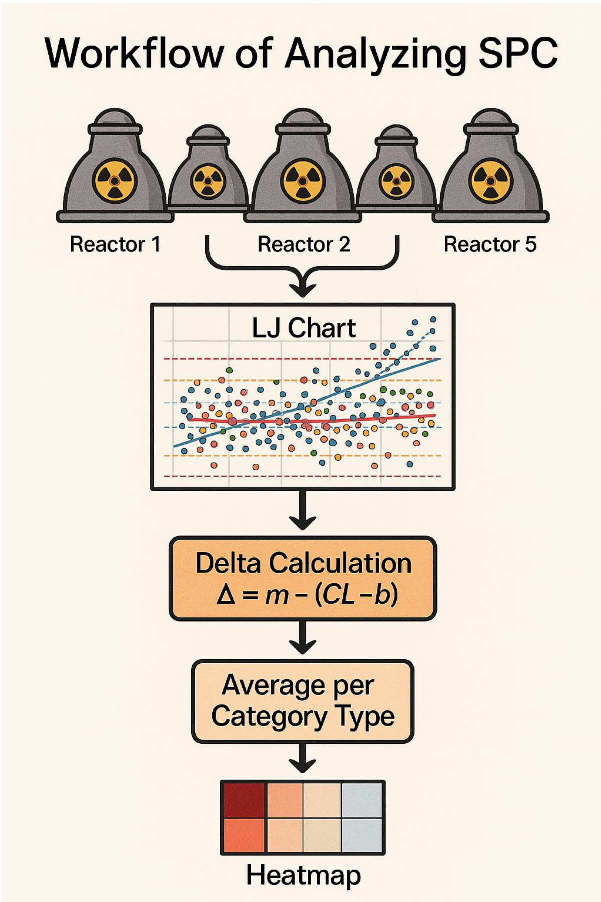
Correlative Advance Statistical Process Control (SPC)

Salvador Gutiérrez Portocarrero

August 10, 2025



Keywords

SPC, Levey–Jennings, regression analysis, high-volume manufacturing, automated manufacturing, process control, Python.

Origin of the Levey-Jennings Chart

In 1950, Stanley Levey and E.R. Jennings published a paper titled: *The Use of Control Charts in the Clinical Laboratory*, in the journal *American Journal of Clinical Pathology*. [1] Their goal was to adapt the principles of Statistical Process Control (SPC) — which were already widely used in manufacturing — to clinical chemistry, where ensuring the precision and reliability of laboratory measurements is critical.

The chart they introduced is essentially a Shewhart control chart using standard deviation-based control limits ($\pm 1\sigma, \pm 2\sigma, \pm 3\sigma$) centered around the mean. [2] However, instead of plotting means of samples, it plots individual test results.

Building on the original Levey-Jennings concept, we explore how this type of control chart can be extended through real-time correlation, and automated using Python. To illustrate this, we first generate simulated data sets that mimic laboratory measurement variability. These Python-based simulations help to demonstrate how process control behavior can be modeled computationally, providing a foundation for automated quality monitoring.

How to Obtain the Data Using Simulations

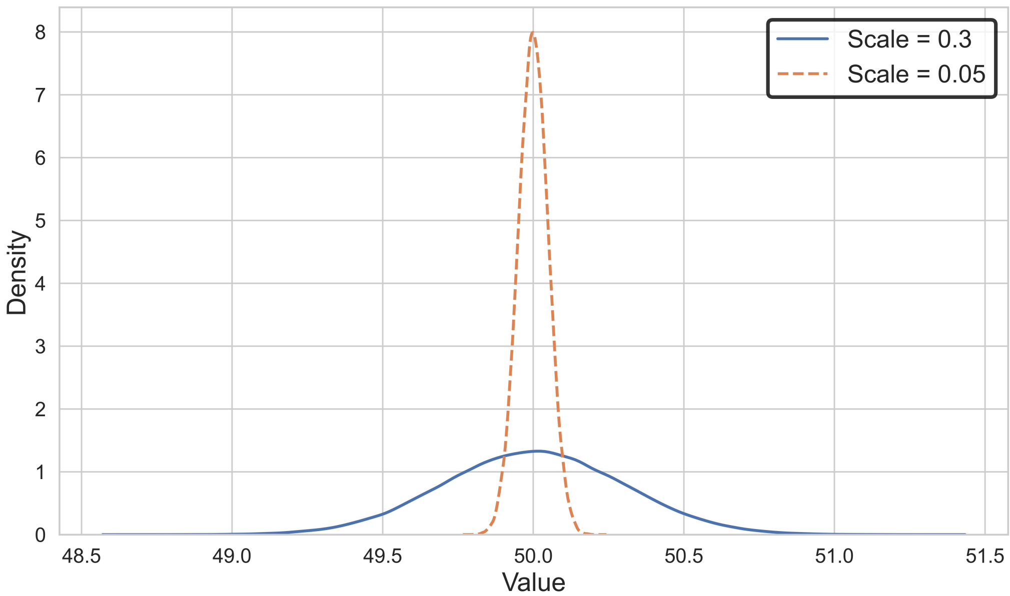
This section explores how varying the standard deviation affects the spread of simulated data. By generating multiple samples from normal distributions, we can simulate random processes and observe how the normal distribution changes based on different spread parameter, normal function scale or standard deviation.

In such simulations, we could pick a process mean (e.g, $\mu = 50$), then define the shape and spread of the random number distribution function. This parameter, as mentioned earlier, determine how the random values are propagated. Specifically, the **standard deviation** is often referred to as **the scale** of the normal distribution in random number generation, and we will use this term from now on to avoid any confusion.

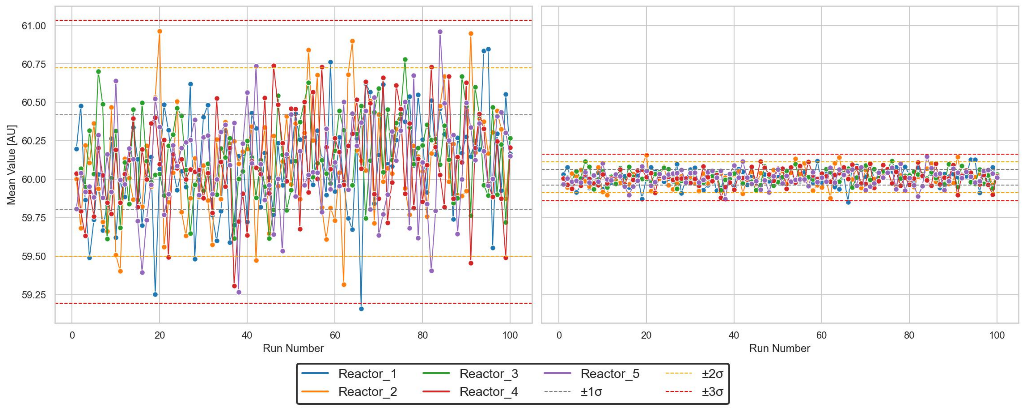
Figure 1a shows the resulting probability density functions. The distribution with the smaller standard deviation is sharply peaked, while the distribution with the larger standard deviation is broader.

- **Scale = 0.3:** Represents a distribution with a wider spread.
- **Scale = 0.05:** Represents a distribution with a narrower, sharper peak.

These simulated datasets demonstrate how the choice of scale (random number standard deviation) affects the variability of the generated random numbers. In **Figure 1b**, we propagate a mean value with arbitrary units, AU, with value of $\mu = 50$. In real terms, the variability of a process is crucial in quality control and SPC contexts, as greater spread can lead to more frequent deviations beyond control limits.



(a) Comparison of two normal distributions with different scales.



(b) Levey-Jennings plots for the same structure using scales values of 0.3 (left) and 0.05 (right) over five reactors.

Figure 1: Simulation Framework

Levey-Jennings Chart with Dynamic Regression Window

Jumping to the enhanced implementation of the Levey-Jennings plots. We included two types of linear regression analyses for all reactors:

- **Full Data Regression:** A linear fit considering all available data points, capturing the overall trend across the entire dataset.
- **Recent Data Regression:** A linear fit applied to the most recent n data points (where n can be dynamically selected), which highlights recent trends or shifts that may not be visible in the full dataset fit.

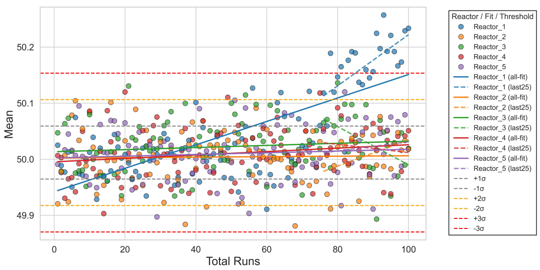
This approach provides more nuanced insights into both long-term stability and short-term variation within each reactor's measurements. Additionally, a linear drift (slope) was intentionally introduced into the simulated process data. Specifically, a constant increment of 0.0002 AU is added to the characteristic level (CL) with each event, meaning that over 10,000 events, the CL will shift by approximately 2 AU.

Control limits are calculated using the mean and standard deviation of the selected dataset, corresponding to $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ thresholds, which are clearly marked on the chart. The regression lines are plotted using distinct line styles and colors for easy differentiation, with a dynamic legend that reflects the current choice of n for the recent data regression.

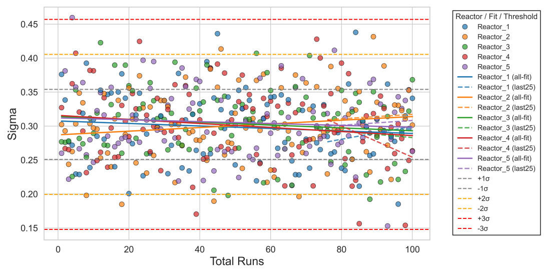
This visualization technique enhances monitoring sensitivity and supports more informed decision-making in process control. For example, a noticeable mean drift in **Reactor 1** of Process A is clearly visible at the final data point, as shown in **Figure 2a**. This significant deviation is apparent even to the untrained eye.

However, more subtle drifts in the process can remain difficult to detect, even with the aid of linear trend lines. **Figure 2c** illustrates such a case, where a drift is present in the same dataset—the same final points of **Reactor 1**—but is much less obvious. This is challenging for the even for a trained engineer, thus we need to use other resources to dredge out the information.

(a) Process A - Mean



(b) Process A - Sigma



(c) Process B - Mean

(d) Process B - Sigma

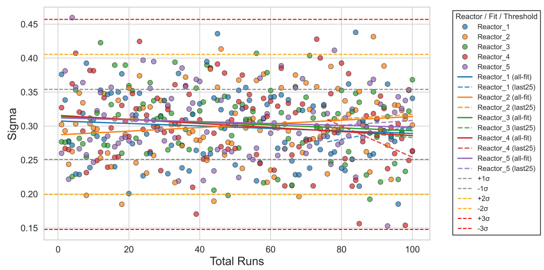


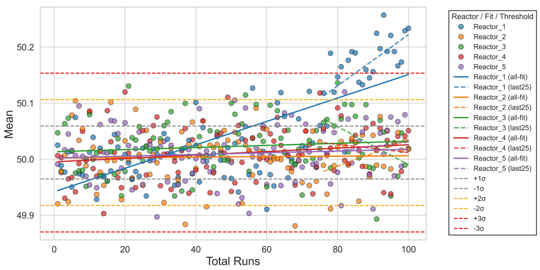
Figure 2: Levey-Jennings charts for five reactors showing full data regression (solid lines) and recent n-point regression (dashed lines) with control limits at $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$.

1 Control Limits and Regression Data Relations

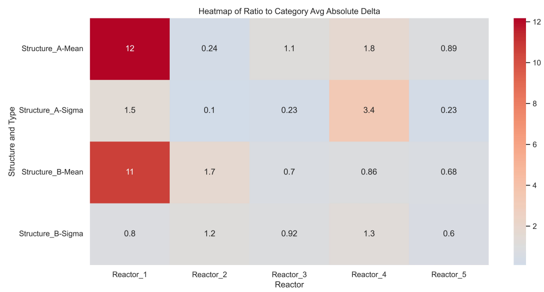
The use of Levey-Jennings control limits—specifically the calculated fleet center line (CL), weighted by the historical slope of the process—is sensitive enough to detect changes occurring in individual reactors within the fleet. Figure 3b confirms observations consistent with the latest point linear trend analysis. However, in the case of Process B, the reactor drift was initially obscured by the natural variability of the distribution. By incorporating fleet-level parameters into the linear trend weighting, the heatmap visualization makes it possible to clearly identify the underlying issue—Reactor 1—as shown in Figure 3d.

Note that we have simulated two means and two standard deviations for Processes A and B. This setup was chosen both to evaluate the effect of the selected random seed and to generate a richer color matrix, allowing us to demonstrate how this code and analytical workflow can scale to multiple variables. Because the same random seed was used for both processes, the random number sequences for A and B are identical—resulting in color matrices with the same underlying numerical values for sigma. You can notice the equality on sigma in Figures 2b and 2d, which are similar but they are different throws. Mean values differ due to the addition of different slope values. SPC charts for the mean and sigma of Structure B are not shown in order to avoid introducing unnecessary data noise.

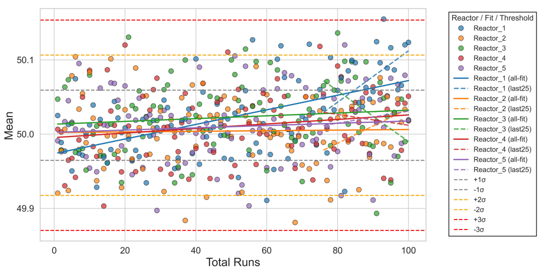
(a) Process A - Mean



(b) Process B - All Reactors Heatmap



(c) Process B - Mean



(d) Process B - All Reactors Heatmap

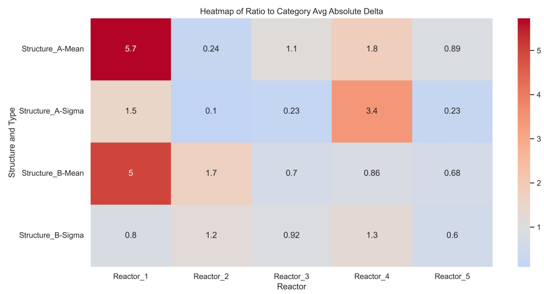


Figure 3: Levey-Jennings charts for five reactors showing full data regression (solid lines) and recent n-point regression (dashed lines) with control limits at $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$.

In addition to the heatmaps, bar plots offer a complementary and quantitative perspective by presenting the ratio of the average absolute delta regression for each reactor against the fleet-wide category-specific average. This normalization allows us to visually compare how far each reactor deviates from the expected behavior, considering the combined effect of slope and control limit statistics. Reactors with values significantly greater than 1 exhibit abnormal behavior relative to their peers and stand out clearly in the bar chart. For instance, **Reactor 1**—previously identified as a potential culprit of process drift in the heatmap—also registers a high deviation in the bar plot, reinforcing its role as an outlier in the process. These bar plots thus provide a clear and intuitive diagnostic tool for ranking reactor stability and detecting process shifts. Notice how the two visualizations—heatmaps and bar plots—are complementary: together, they allow us to both visualize and confirm which reactors or processes are drifting, enabling targeted investigation into potential root causes.

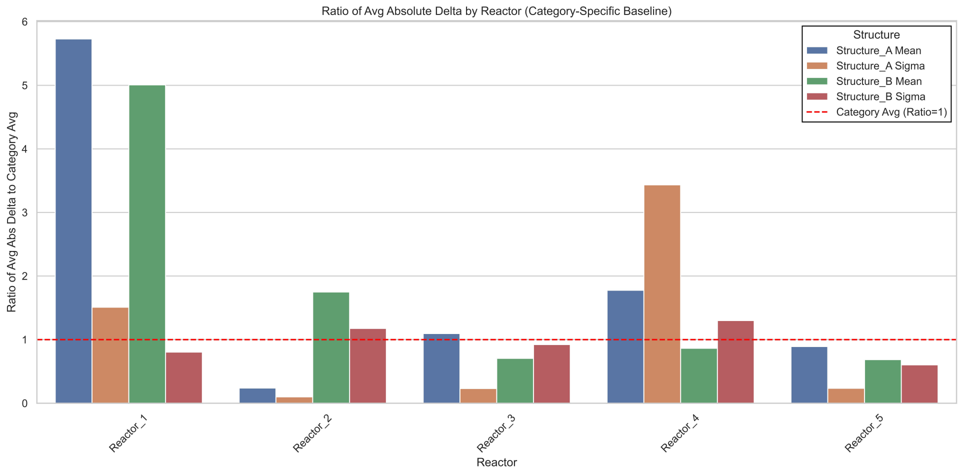


Figure 4: Bar Graphics

2 Correlative Advance Statistical Process Control (SPC) Workflow

This section explains the Python code implementing the *Correlative Advance Statistical Process Control (SPC)* method for reactor performance monitoring.

The code processes tool data, performs statistical analysis, and generates visualizations to detect process drift and variability across multiple reactors.

2.1 Data Loading and Setup

- CSV data is loaded into a Pandas `DataFrame`.
- The file name encodes parameters such as slope, drift, reactor excluded, and scale factor.
- Reactors are sorted numerically to ensure consistent plotting order.
- The structures and types of measurements analyzed are:
  - Structure A – Mean
  - Structure A – Sigma
  - Structure B – Mean
  - Structure B – Sigma

2.2 Levey–Jennings Chart Generation

- For each Structure–Type combination, data is filtered and plotted as a **Levey–Jennings (LJ) control chart**.
- The LJ chart includes:
  - Mean and $\pm 1\sigma$, $\pm 2\sigma$, $\pm 3\sigma$ control lines.
  - Scatter points colored by reactor.
  - Linear regression lines:
    1. **All data points** – solid line.
    2. **Last n points** – dashed line (captures recent trends).
- A specified reactor can be excluded from control limit calculations but still included in regression.

2.3 Regression Statistics Collection

- The slope and intercept are computed for each regression line.
- The last-n regression is useful for detecting recent drifts that may be masked in full data analysis.
- All regression stats are stored in a global list and later merged with control limit data.

2.4 Merging Control Limits with Regression Data

- Mean, standard deviation, and $\pm\sigma$ thresholds are calculated for each Structure–Type–Reactor combination.
- `Delta_Regression` is computed as:

$$\Delta_{\text{Regression}} = \text{Slope\_All} - (\text{CL\_Mean} - \text{Intercept\_All})$$

- This measures deviation between regression trend and the baseline process mean.

2.5 Category-Specific SPC Ratio Analysis

- The average absolute `Delta_Regression` is calculated per reactor.
- For each Structure–Type category, the category-specific average (excluding the chosen reactor) is computed.
- The ratio:

$$\text{Ratio\_to\_Category\_Avg} = \frac{\text{Avg\_Abs\_Delta\_Regression}}{\text{Category\_Avg\_Abs\_Delta}}$$

  is used to normalize performance across structures and types.

2.6 Visualization Outputs

The code produces:

1. **Bar Plots** – Comparing reactor performance ratios by category.
2. **Heatmaps** – Showing relative deviations across reactors.
3. **Scatter Plots** – Highlighting structure and type relationships.

2.7 Data Saving

- All intermediate and final data tables are saved as CSV files.
- File names are generated using `f-strings` to include run parameters (e.g., slope, drift, reactor exclusion).

3 Conclusion

This workflow combines SPC, regression analysis, and multi-reactor correlation to:

- Identify drifting tools early.
- Quantify deviations relative to process baselines.
- Provide actionable insights for process engineers.

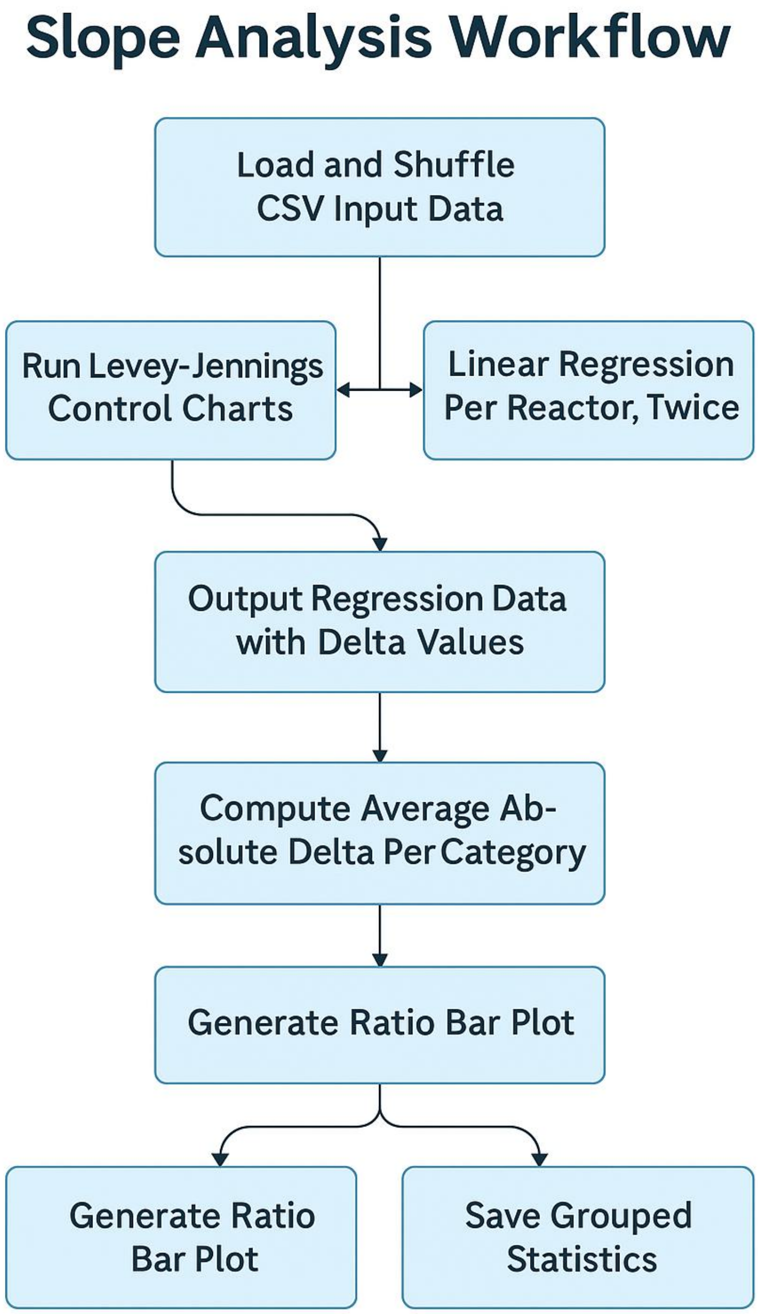


Figure 5: Code Full Workflow

References

[1] Stanley Levey and E. R. Jennings. The use of control charts in the clinical laboratory. *American Journal of Clinical Pathology*, 20(11):1059–1066, 1950.

[2] Walter A. Shewhart and William Edwards Deming. *Statistical Method from the Viewpoint of Quality Control*. Dover Publications, New York, paperback edition, 1986.