

# Adaptive Lasso-Based Robust Multivariate Process Monitoring and Diagnostics

Xiulin Xie

xx23d@fsu.edu

Department of Statistics



## Background

Big data often take the form of **Data Streams** with observations of certain processes collected sequentially over time. Among many different purposes, one major goal to collect and analyze big data is to **Monitor** the longitudinal status of the related processes; Do these data stream remain stable over time, or do they exhibit significant variations? Furthermore, upon detecting an unusual variation, it's crucial to **Diagnose** which specific variables have shifted and identify the onset of these shifts.

## Limitation of Existing methods

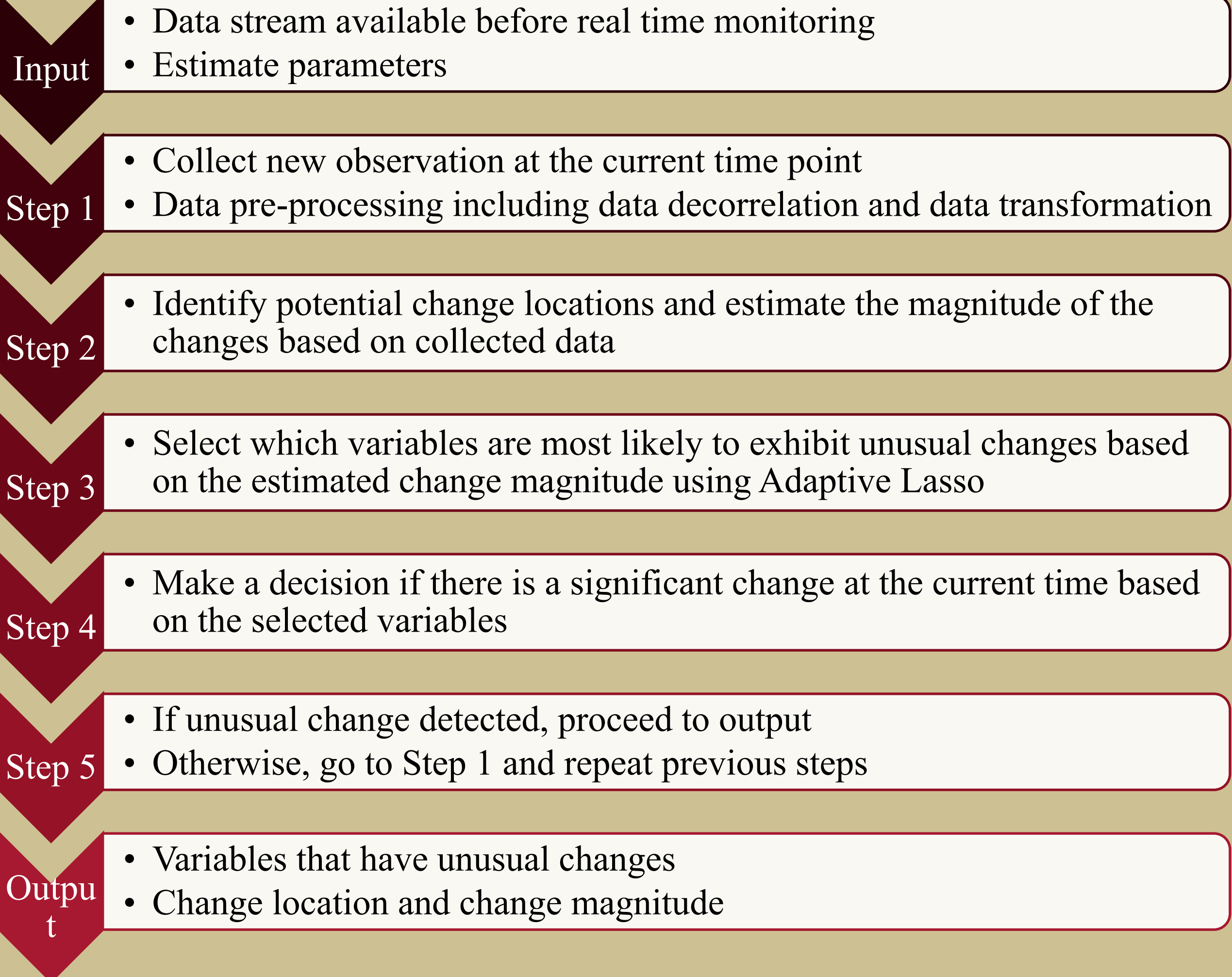
- Many existing methods are designed for process observations are independent and identically distributed with parametric distribution.
- Traditional methods are effective only when all variables undergo an unusual change (i.e., shift), but not effective for detecting sparse shift.
- Primarily focus on the swift detection of these shifts but lack diagnostics.

## Project Objective

Develop a general process monitoring and diagnostics framework

- Implement a robust method for change point and shift estimation in a multivariate serially correlated process using **Recursive Data Decorrelation** and **Data Transformation** techniques.
- Design a detection method that is effective in identifying sparse shifts using **Adaptive Least Absolute Shrinkage and Selection Operator (Lasso)**.
- Provide **Real-time Diagnostics** to offer estimates of both shift's position and identify the shifted variables using variable selection method.

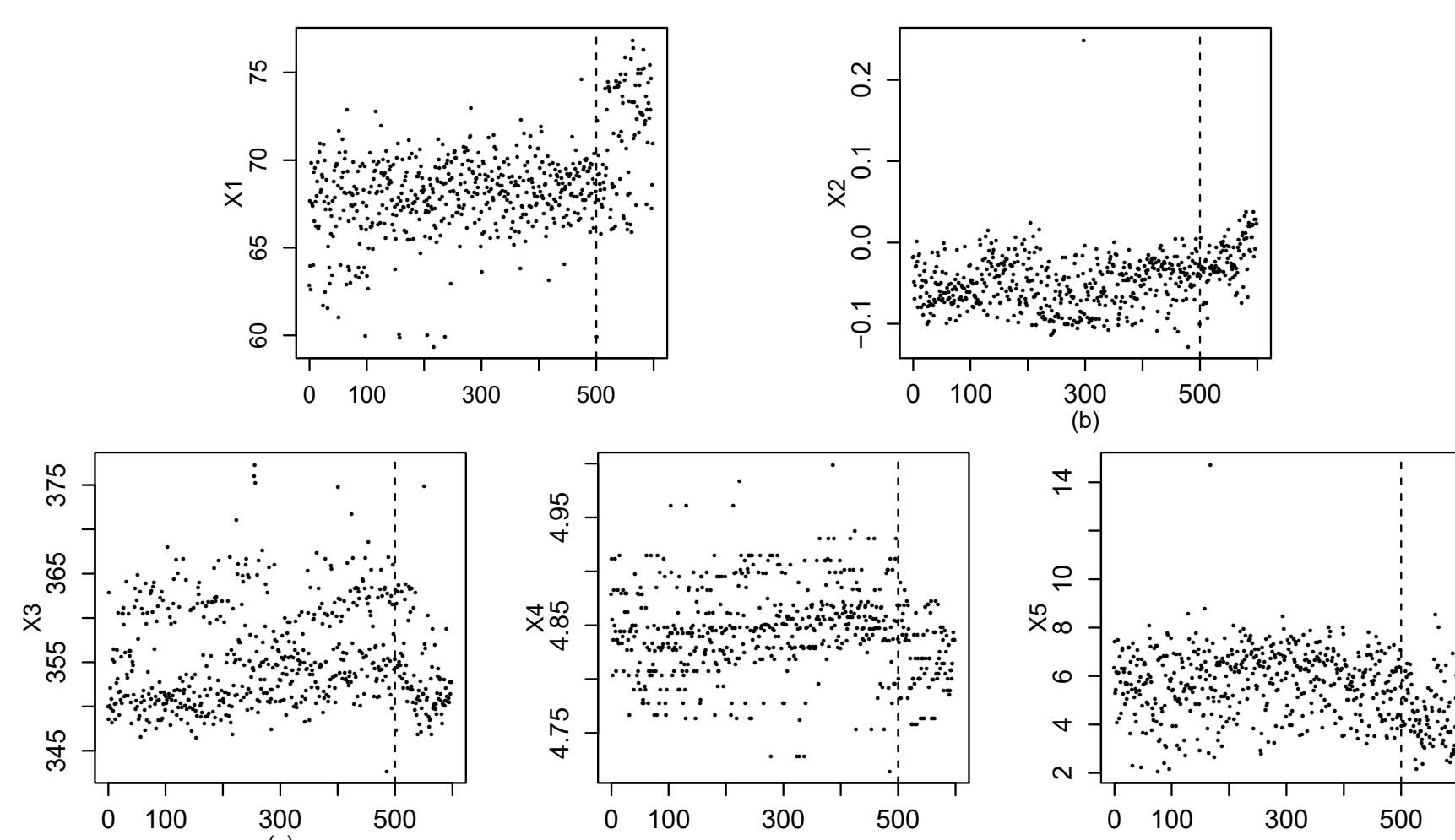
## Proposed Framework



## Application

The dataset contains observations of various extracted quality characteristics of semiconductor products sampled from the manufacturing process. It is crucial to monitor such production processes and detect any unusual changes.

- Training Data: first 500 observations
- Online Process Monitoring begins from the 501st observation.
- Signal Time: given at the 520th observation.
- Change location: occur at the 516th observation.
- Shifted Variables: 1st, 3rd, and 5th variables.



## Future Works

Infectious Disease Surveillance: Focus on the early detection of infectious diseases, which presents additional challenges such as spatial-temporal correlation and insufficient training data.