A Bayesian Approach for Interpreting Mean Shifts in Multivariate Quality Control

Matthias H. Y. Tan and Jianjun Shi

Notes by Taylor Grimm

April 26, 2023

Background Information

There are many processes in different fields in which one or more variables are measured repeatedly over time in order to monitor the performance or quality of the process. When multiple variables are measured, the process can only effectively be monitored by jointly monitoring all variables of interest. This is the goal of multivariate statistical process monitoring (or control).

Various methods exist for monitoring the mean of multivariate processes, including Hotelling T^2 charts, multivariate cumulative sum (MCUSUM) charts, and multivariate exponentially weighted moving average (MEWMA) charts. These charts allow for identification of an observation as out-of-control when compared to previous in-control values. This is known as fault detection.

The task following fault detection is called fault isolation. Fault isolation deals with determining why an observation is flagged as out-of-control by the control chart. The purpose of this paper is to provide a Bayesian approach to identify i) which means have shifted and ii) the direction of the shifts.

In multivariate statistical process monitoring (MSPM), there are two phases we talk about: phase I and phase II.

- Phase I is the in-control training period where we define the control limits
- Phase II is when we monitor new observations for abnormal behavior using the control limits previously defined

Assumptions

Let $\mathbf{x} = (x_1, \ldots, x_p)^T$ be our variables of interest.

- We assume $\mathbf{x} \sim N_p(\boldsymbol{\mu_0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the in-control mean (phase I)
- Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ denote N in-control phase I observations, and n denote the sample size of the phase II sample suspected to be drawn from a common normal distribution different from $N_p(\boldsymbol{\mu_0}, \boldsymbol{\Sigma})$

- All samples are independent
- The covariance matrix $\mathbf{\Sigma}$ remains in control (does not change from phase I to phase II)

T^2 Control Chart

The standard T^2 statistic for a phase II sample can be written as

$$T^2 = n(\bar{\mathbf{x}}_f - \bar{\mathbf{x}})^T \widehat{\mathbf{\Sigma}}^{-1} (\bar{\mathbf{x}}_f - \bar{\mathbf{x}}),$$

where $\bar{\mathbf{x}}_f$ denotes the phase II sample mean (f stands for "future"), $\bar{\mathbf{x}}$ is the phase I sample mean, and $\widehat{\mathbf{\Sigma}}$ is the sample covariance matrix of the phase I data.

- If the phase I observations are independent and identically distributed as $N(\boldsymbol{\mu}_0, \mathbf{\Sigma})$, and $\bar{\mathbf{x}}_f \sim N(\boldsymbol{\mu}_0, \mathbf{\Sigma}/n)$, then

$$T^2 \sim \frac{(N+n)(N-1)p}{N(N-p)} F_{p,N-p}$$

- To control Type I error at α, the upper control limit (UCL) becomes the upper α quantile of this distribution
- If $T^2 >$ UCL for any phase II observations, we flag those observations as faults (abnormal behavior)
  - this indicates either a shift in the mean vector or the covariance matrix
  - since this paper assumes the covariance matrix stays in control, we focus on shifts in the mean vector

Bayesian Hierarchical Model

It can be shown that

$$\begin{aligned} \bar{\mathbf{x}} &\sim N(\boldsymbol{\mu}_0, \mathbf{\Sigma}/N) \\ \mathbf{S} &\sim W(\mathbf{\Sigma}, N-1), \end{aligned}$$

where $W(\mathbf{\Sigma}, N-1)$ denotes a Wishart distribution with scale matrix $\mathbf{\Sigma}$ and $N-1$ degrees of freedom, $\bar{\mathbf{x}}$ and $\mathbf{S}$ are independent, and $(\bar{\mathbf{x}}, \mathbf{S})$ is a sufficient statistic for $(\boldsymbol{\mu}_0, \mathbf{\Sigma})$.

- Above, $\mathbf{S}$ represents the sample dispersion matrix ($\mathbf{S} = \sum_{\mathbf{i=1}}^{\mathbf{N}} (\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^{\mathbf{T}}$).

For a phase II sample with sample mean $\bar{\mathbf{x}}_f \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$ and sample dispersion matrix $\mathbf{S_f} \sim \mathbf{W}(\boldsymbol{\Sigma}, \mathbf{n-1})$, we have the following likelihood function:

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}|\bar{\mathbf{x}}_f, \mathbf{S_f}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}n(\bar{\mathbf{x}}_f - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}_f - \boldsymbol{\mu})\right]$$
$$\times |\boldsymbol{\Sigma}|^{-\frac{n-1}{2}} \exp\left[-\frac{1}{2}\operatorname{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S_f})\right]$$

- Let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^T$ be indicator ariables such that $\delta_i = -1$ indicates that μ_i has decreased, $\delta_i = 0$ indicates that μ_i is unchanged, and $\delta_i = 1$ indicates that μ_i has increased.
- The parameters that will be used in the model are $\boldsymbol{\mu}, \boldsymbol{\delta}$, and $\boldsymbol{\Sigma}^{-1}$.
- Assume that $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}^{-1}$ are independent and that $p(\boldsymbol{\mu}|\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\mu}|\boldsymbol{\delta})$
  - Therefore, $p(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = p(\boldsymbol{\mu}|\boldsymbol{\delta})p(\boldsymbol{\delta})p(\boldsymbol{\Sigma}^{-1})$
- Assume $\boldsymbol{\mu}|\boldsymbol{\delta}$ has a multivariate normal distribution: $\boldsymbol{\mu}|\boldsymbol{\delta} \sim N(\boldsymbol{\Theta}_{\boldsymbol{\delta}}, \boldsymbol{\Psi}_{\boldsymbol{\delta}})$
- Let the following be the prior for $\boldsymbol{\Sigma}^{-1}$ (Wishart):

  $$p(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^t \exp\left[-\frac{1}{2}\operatorname{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right],$$

  where $t = (N-p-2)/2$. This comes from assigning the same prior to $\boldsymbol{\Sigma}^{-1}$ as a noninformative prior for $\boldsymbol{\Sigma}$ in phase I since we assume the covariance matrix to remain constant for this paper.

In summary, the proposed Bayesian hierarchical model consists of the following:

$$l(\bar{\mathbf{x}}_f, \mathbf{S_f}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^{\mathbf{-1}}),\ p(\boldsymbol{\mu}|\boldsymbol{\delta}),\ p(\boldsymbol{\Sigma}^{-1}), \text{ and } p(\boldsymbol{\delta})$$

We are primarily interested in $p(\boldsymbol{\delta}|\bar{\mathbf{x}}_f, \mathbf{S_f})$, which can be used to tell us the probability that a mean has shifted up, down, or remained unchanged.

Prior Distribution for Indicator Variables

We have the following prior for $\boldsymbol{\delta}$:

$$p(\boldsymbol{\delta}) = \prod_{i=1}^{p} p_{1i}^{I(\delta_i=-1)} p_{2i}^{I(\delta_i=0)} p_{3i}^{I(\delta_i=1)},$$

where p_{1i}, p_{2i}, and $p_{3i} = 1 - p_{1i} - p_{2i}$ are the prior probabilities that the ith mean shifted downward, remained in-control, and shifted upward, respectively.

Assuming no prior knowledge about relationships between the variables, we let the default choices of these prior probabilities be $\mathbf{p}_i = (p_{1i}, p_{2i}, p_{3i}) = (0.25, 0.5, 0.25)$.

Prior Distribution for the Mean

Recall that we said earlier that $\boldsymbol{\mu}|\boldsymbol{\delta} \sim N(\boldsymbol{\Theta}_{\boldsymbol{\delta}}, \boldsymbol{\Psi}_{\boldsymbol{\delta}})$.

We set the following:

$$\boldsymbol{\Theta}_{\boldsymbol{\delta}} = (\bar{x}_1 - I(\delta_1 = -1)c_{1d} + I(\delta_1 = 1)c_{1u}, \ldots, \bar{x}_p - I(\delta_p = -1)c_{pd} + I(\delta_p = 1)c_{pu})^T$$
$$\boldsymbol{\Psi}_{\boldsymbol{\delta}} = \text{diag}\left\{ \frac{[a_{1d}^{2I(\delta_1=-1)} a_{1u}^{2I(\delta_1=1)}]\hat{\sigma}_1^2}{N}, \ldots, \frac{[a_{pd}^{2I(\delta_p=-1)} a_{pu}^{2I(\delta_p=1)}]\hat{\sigma}_p^2}{N} \right\},$$

where $\hat{\sigma}_i$ is the phase I sample standard deviation for the ith variable (the ith diagonal element of $\hat{\boldsymbol{\Sigma}}$).

From the structure above, it follows that $\mu_1|\delta_1, \ldots, \mu_p|\delta_p$ are independently distributed with

$$\mu_i|(\delta_i = 0) \sim N\left(\bar{x}_i, \frac{\hat{\sigma}_i^2}{N}\right),$$
$$\mu_i|(\delta_i = -1) \sim N\left(\bar{x}_i - c_{id}, \frac{a_{id}^2\hat{\sigma}_i^2}{N}\right),$$
$$\mu_i|(\delta_i = +1) \sim N\left(\bar{x}_i + c_{id}, \frac{a_{id}^2\hat{\sigma}_i^2}{N}\right).$$

Note that $\mu_i|(\delta_i = 0) \sim N(\bar{x}_i, \hat{\sigma}_i^2/N)$ is solely determined by phase I data.

With these priors on $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$, we see that the only remaining task is to specify c_{id} and a_{id}. With no prior knowledge about potential mean shifts, we consider letting $c_{id} = c_{iu} = c_i = h\hat{\sigma}_i/\sqrt{n}$ and $a_{id} = a_{iu} = a_i$. Then, we have

$$\mu_i|\delta_i \sim N(\bar{x}_i + \delta_i h\hat{\sigma}_i/\sqrt{n}, a^{2|\delta_i|}\hat{\sigma}_i^2/N).$$

We can empirically estimate h and a by performing the following steps:

$$\text{Find } I = \{i : |\bar{x}_{fi} - \bar{x}_i|/\hat{\sigma}_i > 2\}$$

$$\text{Set } h = h^{EB} = \begin{cases} \sqrt{n}\sum_{i\in I} \frac{|\bar{x}_{fi} - \bar{x}|}{\hat{\sigma}_i} & \text{if } I \text{ is nonempty,} \\ 2 & \text{if } I \text{ is empty} \end{cases}$$

$$\text{Set } a = a^{EB} = \begin{cases} \max\left\{\sqrt{N}\,\text{stdev}\left\{\frac{|\bar{x}_{fi} - \bar{x}|}{\hat{\sigma}_i}, i \in I\right\}, \frac{h^{EB}}{2}\sqrt{\frac{N}{n}} - 1, 1\right\} & \text{if } |I| \geq 2, \\ \max\left\{\frac{h^{EB}}{2}\sqrt{\frac{N}{n}} - 1, 1\right\} & \text{if } |I| \leq 1 \end{cases}$$

Gibbs Sampling and Decision Rules for Mean Shifts

We have the following full conditional distributions:

$$p(\boldsymbol{\Sigma}^{-1}|\bar{\mathbf{x}}_f, \mathbf{S}_f, \boldsymbol{\delta}, \boldsymbol{\mu}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{n}{2}+t} \times \exp\left\{-\frac{1}{2}\,\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{S}_f + \mathbf{S} + n(\bar{\mathbf{x}}_f - \boldsymbol{\mu})(\bar{\mathbf{x}}_f - \boldsymbol{\mu})^T)\right]\right\} \tag{1}$$

$$p(\boldsymbol{\mu}|\bar{\mathbf{x}}_f, \mathbf{S}_f, \boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) \propto |\mathbf{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}|^{\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \mathbf{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}})^T \mathbf{V}^{-1}_{\boldsymbol{\delta},\boldsymbol{\Sigma}}(\boldsymbol{\mu} - \mathbf{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}})\right\}, \tag{2}$$

$$p(\boldsymbol{\delta}|\bar{\mathbf{x}}_f, \mathbf{S}_f, \boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Psi}_{\boldsymbol{\delta}}|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})^T \boldsymbol{\Psi}^{-1}_{\boldsymbol{\delta}}(\boldsymbol{\mu} - \boldsymbol{\theta}_{\boldsymbol{\delta}})\right\} p(\boldsymbol{\delta}), \tag{3}$$

where

$$\mathbf{g}_{\boldsymbol{\delta},\boldsymbol{\Sigma}} = (\boldsymbol{\Psi}^{-1}_{\boldsymbol{\delta}} + n\boldsymbol{\Sigma}^{-1})^{-1}(\boldsymbol{\Psi}^{-1}_{\boldsymbol{\delta}}\boldsymbol{\theta}_{\boldsymbol{\delta}} + n\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_f), \text{ and}$$

$$\mathbf{V}_{\boldsymbol{\delta},\boldsymbol{\Sigma}} = (\boldsymbol{\Psi}^{-1}_{\boldsymbol{\delta}} + n\boldsymbol{\Sigma}^{-1})^{-1}$$

Using the full conditional distributions given above, we can use Gibbs sampling to obtain samples from the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta})$. The Gibbs sampling algorithm is outlined in FIgure 1 below.

1. Start with $\boldsymbol{\mu}^0 = \bar{\mathbf{x}}$ and $\boldsymbol{\delta}^0 = (\delta_1^0, \ldots, \delta_p^0)^T$. Set $i = 1$.
2. Sample $(\boldsymbol{\Sigma}^{-1})^i$ from the Wishart distribution with scale matrix $\left(\mathbf{S}_f + \mathbf{S} + n(\bar{\mathbf{x}}_f - \boldsymbol{\mu}^{i-1})(\bar{\mathbf{x}}_f - \boldsymbol{\mu}^{i-1})^T\right)^{-1}$ and $v = n + 2t + p + 1 = N + n - 1$ degrees of freedom.
3. Sample $\boldsymbol{\mu}^i$ from a normal distribution with mean $\mathbf{g}_{\boldsymbol{\delta}^{i-1},\boldsymbol{\Sigma}^i} = \left(\boldsymbol{\psi}^{-1}_{\boldsymbol{\delta}^{i-1}} + n(\boldsymbol{\Sigma}^{-1})^i\right)^{-1}\left(\boldsymbol{\psi}^{-1}_{\boldsymbol{\delta}^{i-1}}\boldsymbol{\theta}_{\boldsymbol{\delta}^{i-1}} + n(\boldsymbol{\Sigma}^{-1})^i\bar{\mathbf{x}}_f\right)$, and covariance matrix $\mathbf{V}_{\boldsymbol{\delta}^{i-1},\boldsymbol{\Sigma}^i} = \left(\boldsymbol{\psi}^{-1}_{\boldsymbol{\delta}^{i-1}} + n(\boldsymbol{\Sigma}^{-1})^i\right)^{-1}$.
4. Define $\boldsymbol{\delta}_j^i(r) = (\delta_1^i, \ldots, \delta_{j-1}^i, r, \delta_{j+1}^{i-1}, \ldots, \delta_p^{i-1})^T$. For $j = 1, \ldots, p$, sample δ_j^i from the discrete distribution:

$$p(r) = \frac{\left|\boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(r)}\right|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(r)}\right)^T \boldsymbol{\psi}^{-1}_{\boldsymbol{\delta}_j^i(r)}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(r)}\right)\right\} p(\boldsymbol{\delta}_j^i(r))}{\sum_{q=-1}^{1}\left|\boldsymbol{\psi}_{\boldsymbol{\delta}_j^i(q)}\right|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(q)}\right)^T \boldsymbol{\psi}^{-1}_{\boldsymbol{\delta}_j^i(q)}\left(\boldsymbol{\mu}^i - \boldsymbol{\theta}_{\boldsymbol{\delta}_j^i(q)}\right)\right\} p\left(\boldsymbol{\delta}_j^i(q)\right)}, r \in \{-1,0,1\}$$

Figure 1: Gibbs sampling algorithm.

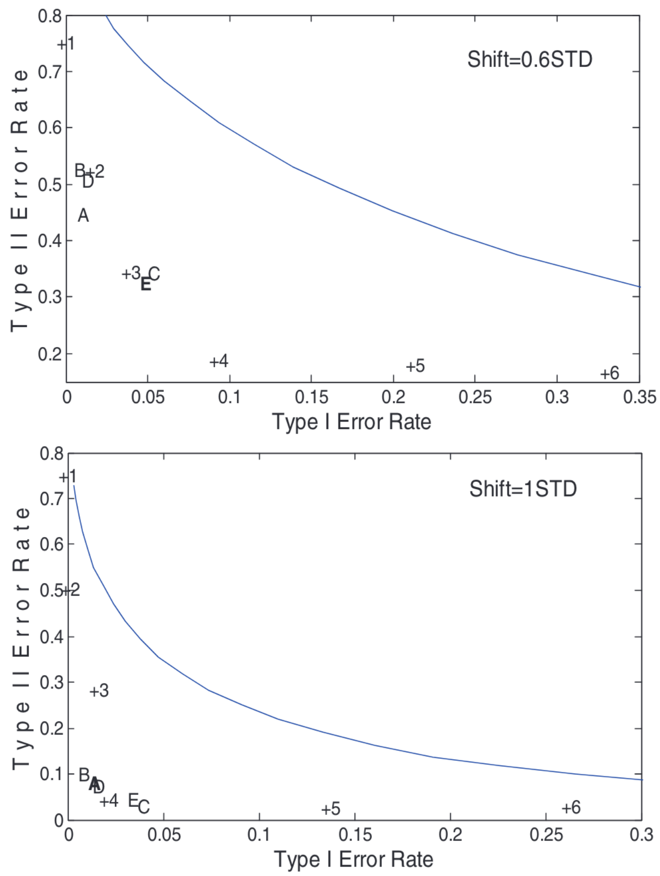
For each j, choose the value d_j of δ_j that appears most frequently (highest proportion) in the posterior samples, and make decision $(d_1, \ldots, d_p)$.

Example: Simulation Comparison

For this simulation, we have $p = 12$, $N = 90$, and $n = 6$ (12 variables, 90 phase I observations, and 6 out-of-control phase II observations).

Under different true shift sizes in the mean vector and different prior choices (A, B, C, and D), the proposed model type I and type II error rates were compared with WJPLM (an existing method) and the standard t-test. The proposed model with empirically selected prior hyperparameters is denoted with an E.

Plots of results under 3 sizes of mean shift are shown in Figure 2.



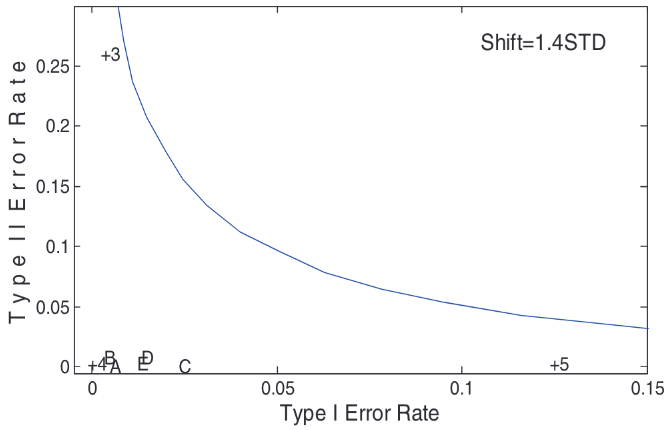


Figure 2: Type II error rate versus Type I error rate for the proposed approach, WJPLM, and t-test. The solid line is for the t-test, the symbol $+L$ is for the WJPLM with model size L, symbols A-D are for different prior distributions (different selections of h and b, and symbol E is for the Empirical Bayes approach.

Example: Fruit Juice Data

Process monitoring data from a fruit juice process. Data consists of concentrations, in microgram per standard volume, of $p = 11$ amino acids. We have $N = 19$ phase I (training) observations, and $n = 11$ phase II observations.

Taking $I = \{1, 2, 4, \ldots, 10\}$, we find that $h^{EB} = 3.58$ and $a^{EB} = 2.01$. Using 20,000 iterations (10,000 burn-in), the posterior mode of $\boldsymbol{\delta}$ is (0,0,0,1,0,0,0,1,-1,-1,0). This indicates that means 4 and 8 shifted upward, whereas means 9 and 10 shifted downward. Marginal posterior distributions of the δ_i's are shown in Figure 3. The T^2 chart for the fruit juice data is also below.

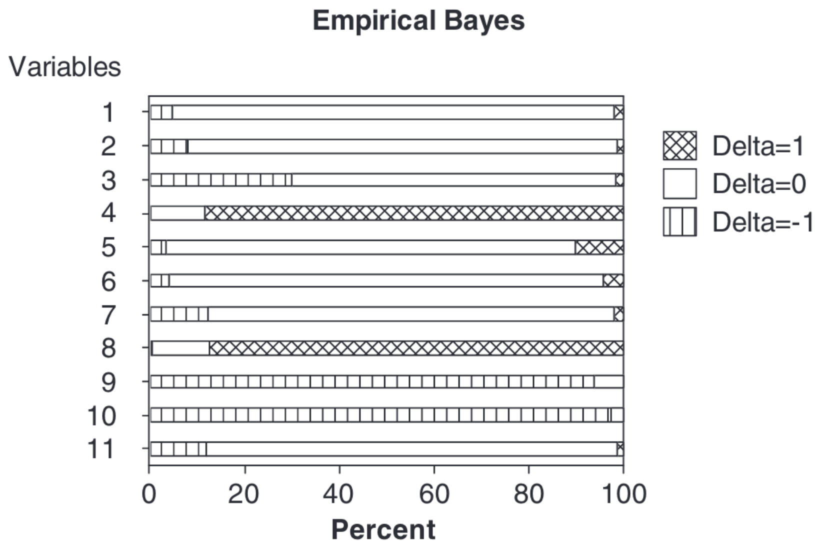


Figure 3: Marginal posterior distribution of each indicator for the EB method—fruit juice data.

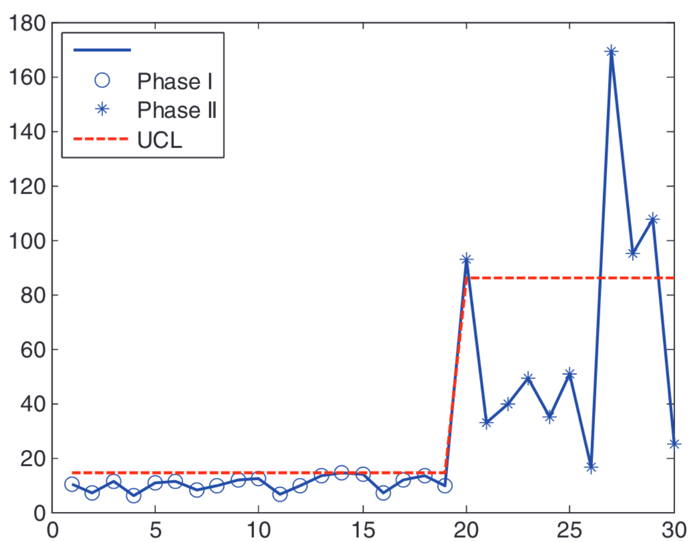


Figure 4: T^2 chart for fruit juice data.