Data-driven Bottleneck Detection in Manufacturing Systems: A Statistical Approach

Chunlong Yu¹ and Andrea Matta²

Abstract—Data-driven bottleneck detection has received an increasing interest during the recent years. This approach locates the throughput bottleneck of manufacturing systems based on indicators derived from measured machine performance metrics. However, the variability in manufacturing systems may affect the quality of bottleneck indicators, leading to possible inaccurate detection results. This paper presents a statistical framework to decrease the data-driven detection inaccuracy caused by system variability. The proposed statistical framework is numerically verified to be spectacularly effective in decreasing the wrong bottleneck identifications in production lines.

I. INTRODUCTION

Throughput is the most relevant metric to evaluate the efficiency of a production system. However, the system throughput is significantly constrained by the bottleneck. More specifically, the bottleneck was recently defined as the machine to which the overall system throughput has the largest sensitivity [1]. Hence, to improve the system throughput, it is necessary to allocate additional production resources to the bottleneck machine. Unfortunately, locating the bottleneck in complex manufacturing systems is not easy, because a direct measure of the throughput sensitivity does not exist in the factory floor.

During the last two decades, several methods for detecting bottleneck in factory floors have been proposed in the literature. These make use of the on-field collected data for obtaining indirect measures of the throughput sensitivity. Roser et al. [2]-[4] proposed the Active period method by utilizing a new classification of machine states in “active” and “inactive”, the machine having the longest uninterrupted “active” duration is considered as the so-called momentary bottleneck; then, the machine accounting for the largest proportion of time being a momentary bottleneck is detected as the bottleneck. Leporis et al. [5] proposed the Criticality method using an indicator that combines the information of machine utilization, machine starvation and blockage, and duration the machines stay in waiting for labor resources. Betterton et al. [6] proposed the TVT method using the station interdeparture time variance (ITV) as a measure for bottleneck detection. Kuo et al. [8] and Chiang et al. [9] [10] proposed the Arrow method, Li et al. [1] [7] proposed the Turning point method. Both of these two methods are based on the analysis of machine starvation and blockage probabilities, relevant details will be provided in section II.

Generally, the bottleneck detection methods can be implemented coupled with analytical models or simulation models for the parameters estimation. However, the drawbacks, e.g., low accuracy of analytical models and long-developing time of simulation models, limit the wide application of model-based methodology in complex systems [1]. Actually, there is an increasing tendency to detect the bottleneck without building an analytical or simulation model but merely with real-time data collected from the manufacturing systems. Such detection approach is known as Data-driven Bottleneck Detection [1].

The data-driven approach has several advantages, but its accuracy is closely related to the variability of the data coming from the field because of manufacturing system randomness. Indeed, variability can be introduced by unscheduled downtime of machines, process time variation, machine setups, recycle, etc. [11]. These uncertainties render machine performance metrics (e.g., buffer level, machine blockage and starvation time) behaving as random variables, with an underlying joint distribution that put all of them in correlation. Therefore, estimation errors are inevitably related to the data-driven bottleneck detection approach, because the machine performance metrics are evaluated by using a finite stream of online records. This may result in possible unreliable bottleneck indicators and, finally, inaccurate detections for the manufacturing companies.

This paper proposes a Statistical Framework (SF) for data-driven detection methods. The SF assesses the reliability of bottleneck detection results, rejecting the proposal of bottleneck when there is no statistical evidence. The main result is the decrease of wrong bottleneck identifications.

The rest of this paper is organized as follows. Section II offers a brief review of two relevant bottleneck detection methods. Section III describes the proposed SF. Section IV illustrates the application of the proposed SF to some numerical cases. Section V discusses some possible problems of the SF. Conclusion and future works are presented in section VI.

II. BOTTLENECK DETECTION METHODS

The proposed SF is general enough to be applied to any bottleneck detection method. However, to make clearer the description of the framework in the next section, we review two relevant bottleneck detection methods. These methods will also be considered in the numerical application of the framework in section IV.

¹Chunlong Yu is with the Department of Mechanical Engineering, Politecnico di Milano, Milan, 20156, Italy (e-mail: chunlong.yu@polimi.it).
²Andrea Matta is with the Department of Industrial Engineering & Management, School of Mechanical Engineering, Shanghai Jiao Tong University (SJTU), Shanghai, 200240, P.R. China. (e-mail: matta@sjtu.edu.cn)
A. Arrow method

The Arrow method (AM) developed in [8]-[10] is a system-theoretic approach that utilizes the machine blockage and starvation probabilities to indicate the bottleneck location. This approach is analytically verified to be consistent to the sensitivity-based bottleneck definition in serial production lines with two machines and one buffer. The method was then expanded to long lines with unreliable machines and finite capacity buffers. The bottleneck is detected with the following rules:

**Rule 1:** Let $m_b$ and $m_s$ be the blockage and starvation probabilities of machine $m_i$, $M$ be the number of machines in the serial production line. If the following condition holds:

$$m_b > m_{s,i+1} : i = 1,\ldots,M-1,$$

the bottleneck is downstream of $m_i$, and an arrow is directed from $m_i$ to $m_{i+1}$. If the following inequality holds:

$$m_b < m_{s,i+1} : i = 2,\ldots,M,$$

the bottleneck is upstream of $m_i$, and an arrow is directed from $m_{i+1}$ to $m_i$. Then, the machine with no departing arrows is detected as the system bottleneck.

**Rule 2:** If multiple machines are detected as the bottleneck using Rule 1, the machine with the highest bottleneck severity is the bottleneck. The bottleneck severity is defined as:

$$S_i = m_{s,i} - m_b, \quad S_M = m_{b,M-1} - m_{s,M}$$

**Rule 3:** With the same notation of $m_b$, $m_s$, and $M$ described before, machine $m_j$ is the turning point if all of the following inequalities are satisfied:

$$m_b - m_s > 0 : i = 1,\ldots,j-1, j \neq 1, j \neq M,$$

$$m_b - m_s < 0 : i = j+1,\ldots,M, j \neq 1, j \neq M,$$

$$m_b + m_s < m_{b,j} + m_{s,j+1}, j \neq 1, j \neq M,$$

$$m_b + m_s < m_{b,j+1} + m_{s,j}, j \neq 1, j \neq M.$$

If $j = 1$:

$$m_b - m_s > 0 \quad \text{and} \quad m_b - m_s < 0$$

$$m_b + m_s < m_{b,1} + m_{s}, j \neq 1, j \neq M$$

If $j = M$:

$$m_{b,M-1} - m_{s,M-1} > 0 \quad \text{and} \quad m_{b,M} - m_{s,M} < 0$$

$$m_{b,M} + m_{s,M} < m_{b,M-1} + m_{s,M-1}$$

Then, the turning point is detected as the bottleneck.

**Rule 4:** If there are multiple turning points, the machine with the maximum bottleneck index is the bottleneck. The bottleneck index is defined as:

$$I_i = \frac{m_{s,i}}{m_b + m_s}, \quad I_M = \frac{m_{b,M-1}}{m_b + m_s}$$

$$I_i = \frac{m_{b,i} + m_{s,i+1}}{m_b + m_s} : i = 2,\ldots,M-1$$

III. STATISTICAL FRAMEWORK

Bottleneck detection methods can be described as a logical procedure to judge a machine being a bottleneck based on machine performance metrics estimated from on-field data in the production system. Here, a ‘logical procedure’ can be described as the action to verify a set of mathematical conditions under which a machine will be detected as the bottleneck.

The accuracy of the bottleneck detection method is not only affected by whether the developed bottleneck conditions can correctly reflect the essence of the bottleneck, but it is also related to whether the machine performance metrics involved in the bottleneck conditions are correctly estimated from the on-field data and precisely describe the nature of the machines. Suppose that $Q$ is a machine performance metric to be estimated using the real-time data record of length $t$. The conventional way for estimation is the sample mean $\bar{Q}$, but this is not always accurate when $t$ is not long enough or the variance of $Q$ is not small. Such estimation error, if not properly taken into consideration, can lead, firstly, to the wrong judgment of the bottleneck condition involving $Q$ and, then, to a wrong bottleneck detection result. To avoid this problem, an indicator able to give information about the reliability of the detection result, is introduced.

A. Indicator of detection reliability

For bottleneck detection methods, let $L_i$ be the set of bottleneck conditions used to judge machine $m_i$ is the bottleneck. Machine $m_i$ will be detected as the bottleneck only when all the bottleneck conditions in $L_i$ are satisfied. As we summarize, in the bottleneck detection methods developed in the last two decades [1]-[10], any bottleneck condition $l_{ij} \in L_i$ (i for $i^{th}$ machine, j for $j^{th}$ bottleneck condition in $L_i$), can be generally formalized into a comparison between two variables deriving from machine performance metrics, as:

$$X_{i,j} (Q_1, Q_2, \ldots, Q_n) < Y_{i,j} (Q_1, Q_2, \ldots, Q_n)$$

Here, $X_{i,j}$ and $Y_{i,j}$ can be described in general as functions of all available machine performance metrics in the system, let us call $X_{i,j}$ and $Y_{i,j}$ as the **Bottleneck Indicators**. Each bottleneck condition has its own bottleneck indicators. For example, the bottleneck indicators of each bottleneck condition in AM and TPM can be easily derived from the description in section II.

Judging whether $l_{ij}$ is satisfied in long-term perspective is actually to verify whether the mean of the random variable $X_{i,j}$ is smaller than $Y_{i,j}$. Actual methods verify this inequality in a deterministic sense neglecting randomness of $X_{i,j}$ and $Y_{i,j}$.
Instead of simply using the sample means of $X_{ij}$ and $Y_{ij}$, we propose the following hypothesis test:

$$H_0 : \mu_{X_{ij}} = \mu_{Y_{ij}}$$

$$H_1 : \mu_{X_{ij}} < \mu_{Y_{ij}}$$

Here, $\mu_{X_{ij}}$ and $\mu_{Y_{ij}}$ represent the true mean of $X_{ij}$ and $Y_{ij}$, respectively.

Standard statistical methods can be adopted to test the above assumptions of $H_0$ and $H_1$. Here, let us denote the p-value of the hypothesis test in $l_{ij}$ as $P_{ij}$. A low $P_{ij}$ indicates a low probability that $H_0$ holds and therefore, a high probability that the $H_1$, i.e., the bottleneck condition $l_{ij}$, is true. Hence, the probability that the bottleneck condition $l_{ij}$ is true is obtained by:

$$\Pr(l_{ij} \text{ is true}) = 1 - P_{ij}$$

Assume that all the $n$ bottleneck conditions in $L_i$ are independent each other, the overall probability that $L_i$ is satisfied is:

$$R_i = \prod_{j=1}^{n} \Pr(l_{ij} \text{ is true})$$

Probability $R_i$ can be used as an indicator for showing how significantly the bottleneck conditions are satisfied. Here we use the term “detection reliability” to represent this probability. More specifically, $R_i$ is defined as the probability that a detection result indicating machine $m_i$ as the bottleneck is reliable.

**B. Batching**

An important assumption made by many of the statistical tests is that the observations are an independent sample from some underlying distribution. However, the data independence in manufacturing system is not always assured. To generate independent observations for the bottleneck indicators $X_{ij}$ and $Y_{ij}$, the batch means technique [12] can be applied.

Assume that $t$ is the length of real time records in which the random variable $Z$ is measured, divide the resulting observations $z_1, z_2, \ldots, z_w$ into $w$ batches of length $k$. Let the mean value of observations in the $s$th batch be $\bar{Z}(s)$, the batch means can be obtained: $Z(1), Z(2), \ldots, Z(w)$. Similarly, we can obtain $\bar{X}_{i,j}(s)$ and $\bar{Y}_{i,j}(s)$, $\forall i,j$.

The batch size of $\bar{X}_{i,j}(s)$ and $\bar{Y}_{i,j}(s)$ should be chosen carefully. It should be large enough to guarantee the independence of batch means, also be small enough to render the number of batches sufficiently large to have an acceptable statistical power in the hypothesis test. In our experiments, for simplicity, we use an identical batch size for all the datasets of $\bar{X}_{i,j}(s)$ and $\bar{Y}_{i,j}(s)$. Firstly, this batch size is chosen small enough to guarantee a specified statistical power (90%) in the hypothesis test of each bottleneck condition. Then, we test the independence of all datasets of $\bar{X}_{i,j}(s)$ and $\bar{Y}_{i,j}(s)$ using the Rank version of Von Neumann’s ratio test [13]. If any dataset is tested to be not independent, we increase the batch size gradually until independent datasets are obtained. If independent dataset cannot be obtained till its batch size reaches a certain fraction (1/5 in our case, considering the minimum $w$ for independence test) of the data length, the records length is considered to be too short and hence longer records are necessary.

**C. Proposed procedure**

The developed $R_i$ can be integrated into the conventional bottleneck detection procedure to reduce inaccurate detection results, with the scheme depicted in Fig. 1. If machine $m_i$ is detected as the bottleneck with the conventional approach based on real-time records, but with a low value $R_i$, it means the detection result may be an extreme realization of bottleneck indicators and it is, therefore, unreliable. A user-defined threshold can be set according to the required precision in practice. Detected results having $R_i$ below the threshold will be rejected. After that, to increase the reliability of the detected result, longer records have to be collected.

With this procedure, if the wrong detected bottlenecks can be rejected, and the true bottleneck can be accepted within an acceptable data length, then the bottleneck detection inaccuracy will be reduced efficiently.

**IV. NUMERICAL ANALYSIS**

The presented SF has been verified using discrete event simulations performed with Arena® simulation software. The performance of the developed SF is studied in two configurations of a production line composed by five machines and four intermediate buffers.

The parameters of the production lines are listed in Table I. The cycle time of machine is normally distributed with mean listed in the table and coefficient of variation equal to 0.2. All machines are supposed to have random failures, particularly, the time to failure and the time to repair are exponentially distributed with MTTF and MTTR equal to 300 and 50 for all of the machines. All buffers have identical capacity of 50 part
TABLE I. MACHINE PARAMETERS (MINUTES)

<table>
<thead>
<tr>
<th>Line A</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine</td>
<td>(m_1)</td>
<td>(m_2)</td>
<td>(m_3)</td>
<td>(m_4)</td>
<td>(m_5)</td>
</tr>
<tr>
<td>Cycle Time</td>
<td>2.5</td>
<td>3.0</td>
<td>2.5</td>
<td>2.8</td>
<td>2.5</td>
</tr>
<tr>
<td>Line B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycle Time</td>
<td>3.0</td>
<td>2.95</td>
<td>2.9</td>
<td>2.95</td>
<td>3.0</td>
</tr>
</tbody>
</table>

slots. Two configurations are considered in this paper: Line A contains one or more bottlenecks whereas Line B contains no bottleneck. They have identical layout and machine/buffer parameters, except the mean cycle time for each machine.

A. Line A: Production line with bottleneck

To locate the true bottleneck in the production line, we estimate the sensitivity of the system throughput to single machine throughput using the finite difference method [10]. We first run a simulation with 100 replications to evaluate the mean system throughput as the reference value. Then we increase the throughput of each machine once per turn by reducing their nominal cycle time by a small amount, say 5%, and run simulation to evaluate the mean throughput of the improved line. Each simulation runs 100 replications and has a length of 40 days (1 day=1440 min) following a warmup period of the same length. The mean throughput values of the improved lines are compared to each other by using ANOVA and Tukey test. According to bottleneck definition, the bottleneck is supposed to be the machine whose cycle time reduction leads to the largest improvement in the system mean throughput.

The mean throughputs of the unimproved/improved production lines and the corresponding 95% confidence intervals over 100 replications are as in Table II. For Line A, the ANOVA test (p-value<0.001) indicates that the throughput of the improved lines are not statistically equal. Particularly, the Tukey test indicates that the system throughput with improved \(m_2\) is significantly higher than the one with improved other machines, which means \(m_2\) is the true bottleneck.

Now we detect the bottleneck in the production line using bottleneck detection methods. A warmup period of 40 days is performed, the steady-state of the production line has been verified using the Welch graphical procedure based on the moving average of the system throughput [12]. Then, we make 500 experiments, each has a simulation length of 2 days. The machine blockage and starvation records are used to drive the AM and TPM. The results are given in Table III. AM and TPM locate the true bottleneck, i.e., \(m_2\), in merely 345 and 335 experiments, respectively. As a conclusion, they are not able to detect the true bottleneck in all the experiments.

We calculate the detection reliability \(R_i\) of each detected result of AM and TPM according to the SF described in section III. In our cases, the Mann-Whitney U Test [14] is adopted to test the hypothesis of each bottleneck condition. The calculated \(R_i\) are depicted in Fig. 2 and Fig. 3, respectively. As in Fig. 2, the detection reliability of \(m_2\) by AM is distributed with a median (95.3%) higher than that of all the other machines. Similarly, in Fig. 3, the \(R_i\) by TPM has the largest median (90.6%). This may imply that \(R_i\) is able to distinguish the correct bottleneck detection result \((m_2)\) from the wrong ones. To filter out the unreliable detection results, let us set the rejection threshold as 0.95, which means the detection result with a \(R_i\) below 0.95 will not be accepted. Moreover, if independent batch means could not be obtained in the calculation of \(R_i\), the detection result is not accepted.

TABLE II. MEAN SYSTEM THROUGHPUT AND 95% CONFIDENCE INTERVAL OF UNIMPROVED/IMPROVED LINES (PARTS/DAY)

<table>
<thead>
<tr>
<th>Line A</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved Machine</td>
<td>none</td>
<td>(m_1)</td>
<td>(m_2)</td>
<td>(m_3)</td>
<td>(m_4)</td>
</tr>
<tr>
<td>Throughput</td>
<td>388.4</td>
<td>390.0</td>
<td>398.6</td>
<td>391.3</td>
<td>391.0</td>
</tr>
<tr>
<td>95% CI</td>
<td>(387.0; 389.8)</td>
<td>(388.8; 391.2)</td>
<td>(397.1; 400.0)</td>
<td>(390.0; 392.5)</td>
<td>(389.6; 392.3)</td>
</tr>
<tr>
<td>Line B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Throughput</td>
<td>368.3</td>
<td>371.1</td>
<td>370.8</td>
<td>369.8</td>
<td>370.6</td>
</tr>
<tr>
<td>95% CI</td>
<td>(367.0; 369.5)</td>
<td>(369.8; 372.4)</td>
<td>(369.6; 372.1)</td>
<td>(368.6; 371.0)</td>
<td>(369.5; 371.8)</td>
</tr>
</tbody>
</table>

FIGURE 2. Line A: Boxplot of the \(R_i\) (AM)

FIGURE 3. Line A: Boxplot of the \(R_i\) (TPM)

TABLE III. LINE A: BOTTLENECK DETECTION RESULT WITH AM AND TPM IN 500 EXPERIMENTS (T = 2 DAYS)

| Methods | Frequency of being detected as the bottleneck |
|---|---|---|---|---|---|---|
| AM | \(m_1\) | \(m_2\) | \(m_3\) | \(m_4\) | \(m_5\) |
| TPM |  |  |  |  |  |
|  | 8 | 335 | 11 | 120 | 17 |

a. In 9 special cases of the 500 experiments, no bottleneck is detected by the TPM.
since an accurate \( R_t \) cannot be obtained. Table IV gives the results after applying the rejection criteria.

Comparing the results in Table III and in Table IV, the majority of the incorrect detection results is rejected by the proposed SF. AM and TPM, coupled with the SF, give wrong detection results only in 27 and 21 cases out of 500 experiments, respectively. However, a portion of the correct detection results (53% for AM, 61% for TPM) is rejected as well because of no statistical evidence. Similar results were obtained in other experiments reported in this paper.

To assure that the correct results can be finally accepted with longer records length, we performed 500 experiments each with simulation length varying from 2 days to 30 days, and detect the bottleneck using AM and TPM. The developed SF was applied with a 0.95 rejection threshold. In Fig. 4 and Fig. 5 the detection results are illustrated. Here, the correct/wrong identification probability (abbr. \( P_c/P_w \)) was calculated by the ratio of the number of experiments in which the bottleneck was correctly/incorrectly identified divided by the total number of experiments. As in Fig. 4, the \( P_c \) of the AM without SF starts from 31% and decreases to 1% when the records length equals 30 days. When the SF was applied, the maximum \( P_c \) is only 5% with 2 day records, this probability decreases continuously with increasing record length and goes below 2% with record length longer than 6 days. On the other hand, the \( P_w \) with SF increases gradually with increasing record length, reaching 83% when the record length is equal to 30 days. As a consequence, by increasing the record length the detection reliability of the correct detection can be improved, and the correct detection results will be accepted when the input record length becomes long enough. Similarly, as shown in Fig. 5, the \( P_w \) of TPM with SF is suppressed below or equal to 1% with record length longer than 6 days, the \( P_w \) with SF increases from 26% to 74% when records of 30 days are used.

### Table IV. Line A: Bottleneck detection results after applying the SF with 0.95 rejection threshold ( \( T = 2 \) days)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Frequency of being detected as the bottleneck</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( m_1 )</td>
</tr>
<tr>
<td>AM</td>
<td>162</td>
</tr>
<tr>
<td>TPM</td>
<td>130</td>
</tr>
</tbody>
</table>

Therefore, it can be concluded that the developed SF is capable in rejecting the wrong bottleneck detection results efficiently for both AM and TPM. Moreover, the correct bottleneck detection result can be obtained by the SF by increasing the input records length.

**B. Line B: Production line with no bottleneck**

Another concern in the bottleneck detection is that, in the case that the production lines do not have any bottleneck, is the data-driven bottleneck detection procedure able to give detection result as “no bottleneck”? Let us consider Line B shown in Table I, in which the machines have different cycle times characterized with a “bowl” shape. According to the “bowl phenomenon”, the production line with no bottleneck can be realized with a “bowl” shape cycle time distribution rather than a “flat” shape [15]. We performed the throughput sensitivity estimation as described before and obtained the mean system throughput of the unimproved/improved lines shown in Table II. ANOVA (p-value=0.45) and Tukey test indicate that indeed no bottleneck exists in this line. Similarly to the analysis for Line A, we performed bottleneck detection with AM and TPM in 500 experiments, with the input record length varying from 2 to 30 days. The results are depicted in Fig. 6 and Fig. 7.

Any bottleneck that is detected in Line B (with no bottleneck) is considered wrong. As seen in Fig. 6 and Fig. 7, without the use of SF, AM always detects a wrong bottleneck in all the experiments with all record lengths, while the wrong identification probability (\( P_w \)) for TPM has a reducing tendency as the record length increases, yet quite slowly. Indeed, for AM, no matter how the arrow directions are, there always must be a detected bottleneck in the system, according to Rule 1. Whereas in TPM, the conditions to be a turning point, i.e., Rule 3, are more difficult to satisfy in the line without any bottleneck. Indeed, as the estimation error reduces with increasing record length, the probability to find a turning point in Line B becomes lower. On the contrary, with the SF, the probability to detect a wrong bottleneck in Line B is reduced significantly for both AM and TPM, because most of the wrong results have low detection reliability and thus they are rejected. Consequently, most of the unnecessary improvements performed in production lines with no bottleneck can be avoided when the SF is applied.
V. DISCUSSION

One problem for the proposed SF is that the indicator $R_i$ is based on the assumption that the bottleneck conditions in $L_i$ are independent one from the other. However, this is not always true because the bottleneck indicators involved in different bottleneck conditions may be correlated with each other to certain extent. Hence, the calculated $R_i$ may deviate from its true value, and the severity of such error depends on the complex effects of the correlations.

Another problem for the SF is that, it may reject a bottleneck detection result even when it is correct. In such case, it implies that the input on-field records are not long enough to make the correct result reliable enough. Although the correct result can be accepted by using longer records, the detection cost increases and efficiency becomes lower. However, it should be noticed that, in scenarios where the bottlenecks are needed to be improved by considerable amount of investments, e.g., purchasing new machines, introducing new technologies, detecting an incorrect bottleneck would usually bring much lost to the manufacturing companies. In this case, i.e. when there is a high penalty for improving a non-bottleneck machine, it is therefore not a bad choice to wait for more data and guarantee a correct detection result. On the other hand, if the bottleneck improving activities require no or few additional investments, such as assigning higher repairing priority for the bottleneck machine, it is recommended to apply the SF with a moderate rejection threshold to compromise between accuracy and efficiency.

VI. CONCLUSION

This research proposes a statistical framework (SF) to spectacularly decrease the inaccuracy of data-driven bottleneck detection procedures. The presented SF has many advantages. Firstly, it is easy to implement and needs no additional information but only the on-field records used by the adopted bottleneck detection method. Secondly, the SF is versatile and applicable to all bottleneck detection methods proposed in the literature. Furthermore, the framework is not limited to avoid wrong detection results in production lines with bottlenecks, but it can also prevent misleading detection results in production lines with no bottleneck.

Further studies will be carried out to improve the SF by taking into account the correlations between bottleneck conditions. The SF will also be extended to become able to rank bottlenecks and to evaluate the probability of a subset of machines containing the true bottleneck. The SF will be applied on different bottleneck detection methods (with different number of logical conditions) and tested in manufacturing systems with more complex layouts that can contain multiple bottlenecks.

REFERENCES


![Figure 6](image1.png) 
Figure 6. Line B: Bottleneck detection results with AM

![Figure 7](image2.png) 
Figure 7. Line B: Bottleneck detection results with TPM