

International Journal of Advanced Operations Management

ISSN online: 1758-9398 - ISSN print: 1758-938X

<https://www.inderscience.com/ijaom>

Defect detection while setting up an assembly line - analytical approach to reduce the N-dimensional solution space

Malolan Sundararaman

DOI: [10.1504/IJAOM.2025.10069920](https://doi.org/10.1504/IJAOM.2025.10069920)

Article History:

Received:	29 March 2023
Last revised:	02 October 2024
Accepted:	04 October 2024
Published online:	25 March 2025

Defect detection while setting up an assembly line – analytical approach to reduce the N-dimensional solution space

Malolan Sundararaman

Department of Management Studies,
National Institute of Technology Tiruchirappalli,
Tiruchirappalli – 620015, India
Email: malolan@nitt.edu

Abstract: While setting up an assembly line, challenge of defect detection is twofold. The first is determination of possible defects and second is examination of a newly setup assembly line to identify the occurrence of possible defects. Here, a two-step analytical approach to address this managerial problem is proposed. In the first step, past setup data are utilised to identify: 1) list of possible defects; 2) list of control parameters, along with the possible operating levels, to construct an N-dimensional solution space. A bootstrapping approach generates bootstrapped sampling distributions for each sub-solution space. From this, probability of occurrence for a specific defect when control parameters are operated at a particular level is identified. In the second step, an optimisation model determines the ideal sub-solution space, ensuring all defects are identified with an acceptable level of certainty. Data from a developed experimental design is used to demonstrate the proposed approach empirically.

Keywords: defect detection; assembly line; bootstrapping; optimisation; probability mass function; PMF.

Reference to this paper should be made as follows: Sundararaman, M. (2025) 'Defect detection while setting up an assembly line – analytical approach to reduce the N-dimensional solution space', *Int. J. Advanced Operations Management*, Vol. 16, No. 1, pp.109–134.

Biographical notes: Malolan Sundararaman is an Assistant Professor in the Department of Management Studies at the National Institute of Technology Tiruchirappalli, India. He is highly passionate about problems in the area of operations management, applied data analytics, Industry 4.0 and sustainable environment policies. Therefore, his research primarily focuses on the application of analytics for solving business problems, resulting in both resource savings and economic benefits. He obtained his PhD from the Department of Management Studies, Indian Institute of Science Bangalore. His PhD thesis addresses the development of production and inventory plans for an automobile manufacturer when returns are incorporated into the production system through the adoption of remanufacturing practices. His research is publications in referred journals such as *Operations Management Research*, *OPSEARCH*, *Resources Conservation and Recycling Advances* and *Environmental Science and Pollution Research*.

1 Introduction

Mass manufactures always focus on defect detection as an essential aspect of assembly lines. From a management perspective, defect identification before the sale of products leads to significant savings in terms of warranty costs and, more importantly customer satisfaction. However, the costs of conducting tests for defect detection are expensive. Conventionally, samples from the production are used to determine the quality of a production lot (Wetherill, 1977). Past literature has also considered concepts of accelerated life testing to establish the life of products and equipment (Zhu et al., 2021; Woo, 2016; Shin and Lee, 2008). These types of analysis are suitable for mature assembly lines operating at full scale. However, the setup of an assembly line is done in stages. Here, it is imperative to determine the possible defects in each stage and rectify them, to ensure they are avoided when producing the product to scale. Therefore, this study focuses on the problem of defect detection in the stage-by-stage setup of an assembly line.

In a stage-by-stage approach for setting up an assembly line, each stage represents a ramp-up in production. Ramp-up is the process of incrementally increasing the quantity/volume of output obtained from the assembly line. The last stage of this process is the final volume that the assembly line will achieve, thereby operating at full scale. The main reason for considering a stage-by-stage approach for setting up an assembly line is to identify possible manufacturing defects and rectify them before moving to the next stage by ramping up. In this fashion, maximum possible defects are identified and rectified before the final ramp-up. It is to be noted that products from the final stage are sold to customers. The products produced at the intermediate stages are not sold to customers, but are solely for internal testing purposes.

An example of a stage-by-stage ramp-up would be the setup of an automobile (car) assembly line. The assembly line to assemble a car is setup before the launch of a new car. One important requirement of the assembly line is its ability to assemble cars without manufacturing defects. This ensures that during mass production, cars adhere to targeted driving comforts. For this purpose, production volume on the assembly line is generally ramped-up in stages.

At each stage, a specific number of test cars are produced and a road test is performed. The road test involves driving test cars for pre-specified test distances on actual roads. The actual roads are in target regions where the car is expected to be sold. Local drivers from the target region are hired to drive the test cars. The drives are rotated such that every car is driven by every driver within the total test distance. When a test car is driven by a driver, a keen eye is maintained to identify possible manufacturing defects (such as engine noise in the cabin, steering vibrations, door not closing, etc.). If a defect occurs, the nature and distance of a defect occurrence are recorded in a database. Thus, the road test aims to identify possible defects a customer may experience when driving the car. Once a defect is identified, the root cause may be rectified before the subsequent production ramp-up. In this fashion, defects are reduced when moving from one stage to another. However, certain defects may arise due to inherent design elements. These will be persistent throughout the car's lifecycle.

From a management perspective, the road test is significant due to the costs involved. For every launch (either new or facelift) and at every stage, the company has to produce cars which are tested and then disposed. This leads to incurrence of the production and disposal costs. Additionally, in each stage of the road-testing process, test cars are driven

by drives up to the specified test kilometres. This creates the cost components of salary and overheads for drives, and fuel cost for the specified drive distance (test kilometres). The total cost for the road test activity is the summation of all these costs. The purpose of the road test activity is to identify possible defects in a car and rectify the identified defects before the final stage of setup.

Thus, any reduction in either the number of test cars or the drive distance per test vehicle (test kilometres) will result in significant savings. This in turn will lead to the managerial benefit of improved profitability and competitiveness. However, this reduction should be realised without compromising the ability of the road test to identify the possible defects in each stage.

Therefore, for each stage in a general stage-by-stage setup of an assembly line, it is required to identify the best solution space from an N -dimensional solution space. In the N -dimensional space, each defect takes up one dimension. Hence, if there are ' N_1 ' defects, the remaining $N - N_1$ dimensions are the production-related operational variables (in the previous example, it was the number of test cars to produce and the drive distance per test car). The best solution space will identify all possible defects (N_1) in combination with minimal values for the other $N - N_1$ dimensions.

Accordingly, in this study, a bootstrapping approach is proposed to identify the best solution space for defect detection. The best solution space is an N -dimensional probability space consisting of the N_1 defect dimensions and the other $N - N_1$ production parameter levels (controllable parameters). The approach aims to reduce the solution space for each defect by generating the bootstrapped sampling distribution. These bootstrapped distributions will assist in determining the statistical defect distribution parameters, leading to the observation of possible plateaus in the cumulative probability distributions for each identified defect when considering combinations of the other $N - N_1$ production parameter levels. An ideal combination for the other $N - N_1$ dimensions can be suggested from the different plateaus for various defects. The suggestions can be aggregated, and a single combination can be determined that satisfies the requirements of capturing all defects with a specified probability. It is noted that the proposed approach is applicable to any manufacturing/ assembly industry where setup of the production facility is done in stages, where each stage is a ramp-up of the previous stage.

In the next section, past literature is explored to establish the newness of the problem addressed in this work. Section 3 presents a detailed description of the new focused problem with an example. The mathematical constructs for the described problem are presented in Section 4. In Section 5, an experimental design is proposed to simulated data which is used to empirically illustrate the mathematical construct presented in Section 4. Section 6 highlights the managerial insights observed for the simulated data. The study concludes in Section 7.

2 Literature review

The works of Chen et al. (2021) and Li et al. (2021a) have extensively reviewed past literature on defect detection in the manufacturing space. Chen et al. (2021) review the literature on defect detection in additive manufacturing technologies. The study describes the different defect inspection technologies, their applications and architectures in

additive manufacturing. Li et al. (2021a) provide a survey of the literature on defect detection in textile manufacturing. These two studies can be used to classify general defect detection literature into traditional and learning-based algorithms. Accordingly, modern research has extensively focused on learning-based artificial intelligence and machine learning algorithms for defect detection (Singh et al., 2023; Singh and Desai, 2022; Suo et al., 2022; Tulbure et al., 2022; Khan et al., 2021; Li et al., 2021b; Liu et al., 2021; Zhang et al., 2021; Czimmermann et al., 2020; Ren et al., 2019; Yang et al., 2019). Such studies focus on defect identification in fully operational production systems through real-time or near real-time monitoring. They do not address the initial setup of production systems.

Literature that focuses on traditional algorithms also addresses the issue of defect detection from the standpoint of established production systems. Among these, image processing techniques are the most common and have been extensively used to detect defects for various products (Kuo and Nursyahid, 2022; Wang and Cheung, 2021; Gao et al., 2020; Akhtar et al., 2019; Arnal et al., 2017; Lee and Vachtsevanos, 2002). Apart from image processing, statistical techniques such as sampling and bootstrapping are also adopted for defect detection.

The book by Wetherill (1977), establishes the extensive use of sampling for quality inspection in the production process from as early as 1950. Therefore, only some exclusive works which have led to different use of sampling in production systems are presented here. Originally, optimal inspection policies were assumed to hold only for IID Bernoulli product characteristics with single-sample lot acceptance-sampling plans. Later, Wiel and Vardeman (1994) establish that results from single-sample lot acceptance-sampling plans hold more generally, and over a much wider class of inspection plans. The study by Nurani et al. (1994) is one initial work that explores the development of a methodology to determine optimal sampling strategies for in-line defect inspection. The methodology determines the optimal sampling strategies to identify trade-offs between the costs of sampling and cost of defective dyes. Nurani et al. (1996) extend the previous work to be a more comprehensive cost-based methodology for defect inspection, the number of inspected lots and the number of inspected wafers per lot for a wafer fabrication process. The methodology proposed in the study assists in allocating in-line inspection resources based on the distribution of defects across the wafer layers, considering relevant cost components such as the cost of running out of control and wafer inspection. Tsarouhas and Liberopoulos (2004) explore the use of statistical analysis of failure data of equipment obtained from an automated pizza production line. The study identifies the optimal allocation of the total number of visits a quality inspector has to undertake to different workstations on the production line so as to minimise the total production time of undetected, defective products.

The contemporary works of Elhabashy et al. (2021) and Wang et al. (2021) on sampling provide a new perspective on its adoption. Elhabashy et al. (2021) focus on using random sampling to determine the attributes to be monitored to detect cyber-physical manufacturing attacks. The study highlights the usefulness of randomness in sampling plans of multivariate control charts to design more resilient quality control tools. The sampling plans developed in the study did not require a large number of variables for inspecting if the system has had a cyber-attack or not. The study concludes by indicating the cost-benefit of using the proposed technique to detect cycle attacks. Wang et al. (2021) propose a dynamic sampling algorithm based on cost-risk assessment. The proposed algorithm considers measurement cost, product and process risks at the

same time. The study focuses on dynamically selecting lots with lower measurement costs and higher risks from a large number of produced lots, in a semiconductor manufacturing process, for scheduling at the measurement site. It is shown that the proposed algorithm is effective in terms of reducing sample ratio and cycle time, and also reduces the required number of sampling lots and interval time for catch execution. Comparisons for effectiveness in terms of cost and catch execution of defects are established for standard sampling procedures and the proposed algorithm.

In bootstrapping literature, Gine and Zinn (1990) characterise the consistency of bootstrapping and establish that no local uniformity of the central limit theorem is needed for bootstrap to work for a large class of statistics. Jiang and Wang (2001) address the concern of small sample sizes for effective defect classification by machine vision algorithms, particularly for surface defect detection. In the study, a bootstrapping sampling technique is adopted to generate more defect samples. By the adoption of bootstrapping, statistical features of the defects are determined which enhances the accuracy of classification algorithms. The study by Chung and Lee (2001) corrects the convergence error in the construction of bootstrap confidence bounds. The study establishes that optimal bootstrapped sample size yields intervals that are shorter in length and with greater stability, compared to competitors of similar convergence accuracy. Phu et al. (2018) adopt a bootstrapping approach to guide the choice of the number of healthy subjects required for the visual field normative database. The study concludes that the ground truth statistics for visual field-testing sensitivities can be approximated using set sizes that are significantly smaller than the original cohort. One can explore and identify more studies that have adopted bootstrapping for defect detection. However, the studies consider the problem of identifying defects in outputs produced by established production systems.

The current study takes a cue from the works of Jiang and Wang (2001), Chung and Lee (2001) and Phu et al. (2018) in the bootstrapping space and from Elhabashy et al. (2021) and Wang et al. (2021) in the sampling space, to support the decisions of setting up an assembly line by reducing the N -dimensional space for the occurrence of defects. Accordingly, the problem of focus in this work is explained in detail in the next section.

3 Problem description

Setting up an assembly line involves long-term investments with minimal flexibility for rapid modifications. When the assembly lines are put in place, the aim is to ensure that produced products adhere to quality standards with minimal defects. In this study, the focus is on manufacturing defects caused by elements in the production system that can be observed by customers. The presence of such manufacturing defects causes significant losses in terms of customer satisfaction and warranty costs. Thus, it is imperative to identify and rectify elements which cause such defects. This identification is a herculean challenge, especially in the initial phase of establishing the assembly lines. Hence, most assembly lines adopt a stage-by-stage approach to reach maximum production capacity. In each stage, a pre-fixed number of products are produced and examined. The examination leads to the identification of defects, which are then further analysed to identify root causes leading to corrections/ calibrations of the assembly line. As the assembly line moves from one stage to the next, capacity to produce products is

increased. This mode of increase is referred to as ‘ramping-up’. Undertaking the above activity ensures that as the volume of production increases, defects are reduced, ultimately producing defect-free products to scale for customer sales.

A real-life application of setting up an assembly line in this fashion is the automobile industry. Generally, when a new car is launched, the assembly line is setup by procuring and installing required new machinery from external sources. These new pieces of machinery are sequenced and performed in a specific order to obtain assembled cars. When an existing car is updated (facelifted), the existing assembly line undergoes modifications. The modifications could be the addition of new machines, replacement of existing machines with new machines, the performance of additional operations in existing machines, etc. In either case, setup of a new assembly line or modifications of an existing assembly line, installed machines are individually tested and tuned for performance.

Once the assembly line is ready with tuned machines, the assembly of cars begins. However, before full-fledged assembly reaches maximum throughput, production is progressively stepped-up so that all machines work in synergy. Generally, a four-stage approach is adopted to gradually increase production volumes. Nevertheless, every automobile company has internal standards based on experience that pre-specify:

- 1 the number of stages
- 2 the number of cars to produce in each stage.

For the general scenario of four stages, all cars produced in the first three stages are test cars. Among these three stages, all test cars produced in the first two stages are scrapped after the necessary tests have been performed. In the third stage, test cars can be repaired and sold to interested parties at a price lower than the actual selling price. They may also be scrapped if a sale is not possible. The cars produced in the final stage are for actual market sale. Of the cars for final market sale, a few are taken for test purposes. The cars taken for test purposes are later salvaged by using them as test-drive vehicles, display models in showrooms, etc. Test cars of the fourth stage are not sold at retail price to end customers.

The above example clearly illustrates a practical application of the stage-by-stage setup of an assembly line. The main challenge in such a setup is to identify how many cars to produce in each stage for identifying defects. This, as illustrated in the example, is specified based on experience. Nevertheless, this study proposes a statistical approach to identify the number of cars to be produced in each stage. The approach aims to identify the number of cars to be produced in each stage so that all possible manufacturing defects can be identified. For this, setups of past cars with similar assembly lines can be exploited to determine the probability distribution of the possible defects. This probability distribution would be in an N -dimensional space, which can be divided into ‘ N_1 ’ dimensions for the defects and ‘ N_2 ’ dimensions for the control parameters. Using a bootstrapping approach, bootstrapped sampling distribution can be determined for each defect under different combinations of ‘ N_2 ’ control parameters. Using the determined bootstrapped sampling distribution, the probability mass function (PMF) and cumulative distribution function (CDF) for identifying a defect at a particular combination of the ‘ N_2 ’ control parameters is computed. Once the PMF and CDF for the ‘ N_1 ’ defects are identified, from the entire N -dimensional solution space, an ideal combination of remaining ‘ N_2 ’ control parameters can be determined so that all ‘ N_1 ’ defects can be

captured with a desired level of probability. For the determination of the ideal combination of control parameters, an optimisation model is proposed. The objective is to minimise the costs of the test. The optimisation is performed while ensuring a specified desired minimum level of capture probability for each defect.

Returning to the illustrative example of setting up an assembly line to assemble cars, tests cars produced at various stages are used for several test purposes. One among them is the road test, which focuses on identifying manufacturing defects from a customer perspective. In the road test, a select number of test cars are exclusively driven on actual road conditions in which the cars are to be sold. The test design is such that, in each stage, drivers who drive the cars are of different drive characteristics to mimic actual target customers. Moreover, in each stage, all drivers drive all test cars of that stage. The total distance a test car must be driven in a particular stage is pre-specified by the company. The company pays drivers a fixed salary and provides fixed routes to drive the respective test cars. The company also incurs fuel costs for all test cars for the duration of the road test.

The drivers are trained to have a keen eye for identifying manufacturing defects in the test cars. Here, manufacturing defects refer to those a target customer may notice when driving the car after purchase. Some examples of such manufacturing defects are:

- 1 engine noise in the cabin
- 2 AC vent not working
- 3 steering vibrations
- 4 headlights not working
- 5 slow pickup, etc.

Past service complaints about other similar cars manufactured by the company provide the basis for identifying possible defects. In addition to this, drivers may also identify defects based on their driving experience.

When a driver identifies a defect in a particular test car, the information: defect, kilometre at which the defect was observed and test car in which it was observed are recorded. After completing the road test for a particular stage, the recorded information is compiled from various drivers. These are addressed appropriately (a production issue is rectified; a design issue is catalogued and provided as feedback to the design team) before the next ramp-up. The defects identified in test cars of the final stage are catalogued and addressed either during service or in the next vehicle launch.

Hence, the road test identifies possible manufacturing defects that may cause discomfort to a customer. The aim is to ensure that these defects are identified in advance (during road test of earlier stages of setup) so that they can be rectified in successive ramp-up stages thereby avoiding their occurrence within the warranty period of the car. The regular warranty period, for most cars sold in India, is two years or 40,000 kilometres (whichever is earliest) which may be extended by additional plans for up to four years or 100,000 km (whichever is earliest) (Paul, 2016). In an ideal scenario, if a detected manufacturing defect does not occur in a customer's car within the regular warranty period, then the road test is considered a success. However, some detected manufacturing defects occur in the customer's vehicle. This is because the usage of a car will create wear and tear leading to the occurrence of manufacturing defects. Thus, the

road test is considered a success if all manufacturing defects reported by the customer had been noticed in the road test. Therefore, a car manufacturer can achieve significant savings in cost and time from the road test if the occurrence of different manufacturing defects is found using lesser resources (reducing the number of test cars or reducing the drive distance per test car or both). These savings should be achieved without compromising the ability of the road test to identify the occurrence of a defect.

To identify when different defects occur, the probability of a defect happening in a random number of test cars when driven up to a certain distance is to be found. However, given the limited availability of data, a small number of test cars and very few road tests (1 or 2) in a year, the probability distribution for defect occurrence cannot be obtained by sampling the population. Nevertheless, it is possible to obtain the bootstrapped distribution for the occurrence of a defect from a single past road test. As road tests are designed experiments, which are blocked for driver and distance characteristics, it is possible to obtain the PMF for sub-solution spaces of every defect identified in the past. The sub-solution space is the combination of the number of test cars and the drive distance per test car. For every such sub-solution space, it is possible to obtain bootstrapped samples using which the PMF for that sub-solution space can be determined. With the PMF for different sub-solution spaces, it is possible to establish the CDF for a defect. The above procedure can be adapted to generate the CDF for different defects, from which an appropriate number of test cars and the drive distance for each test car can be determined using an optimisation model. The selection should achieve the requirement of identifying all the defects with the best (minimal) cost and time.

Therefore, this study proposes a developed bootstrapping approach to estimate the PMF and CDF of different defects for different possible sub-solution spaces. Once the PMF and CDF are estimated for different defects, a developed optimisation model determines the ideal sub-solution space. The ideal sub-solution space ensures the identification of all defects with a necessary degree of capture probability while incurring the minimal possible cost for the road test. In the next section, a general mathematical construct for determining the PMF and CDF of the sub-solution spaces and the proposed optimisation model to identify the best sub-solution space are presented.

4 General mathematical construct for the problem

To mathematically formulate the problem described in Section 3, it is assumed that defects are not only independent of each other but also across the different stages. Furthermore, the control parameters are assumed to be a discrete set with each parameter having discrete levels. With these assumptions, it is possible to visualise an N -dimensional space consisting of defects dimensions and control parameter dimensions. If there are ' N_1 ' total defects and ' N_2 ' total number of control parameters, then total number of dimensions (N) equals $N_1 + N_2$. Let ' D ' denote the set of defects $\{d \mid d \in D, D = \{1, \dots, N_1\}\}$ and ' C ' the set of control parameters, $\{c \mid c \in C, C = \{1, \dots, N_2\}\}$. Each of the N_2 control parameters can have different levels. Let set ' L ' denote the total number of levels for each control parameter in C , then $\{l \mid l \in L, L = \{l_1, \dots, l_{N_2}\}\}$. With these, the ' CP ' matrix denotes different possible levels for each control parameter- c . Then $\{cp_{jc} \mid cp_{jc} \in CP\}$, cp_{jc} indicates the level- j for control parameter- c . It is noted that level- j for control parameter- c is in the range of $[1 \text{ to } l_c]$. Thus, CP is an irregular matrix consisting of N_2 columns representing the control parameters, with rows in each column

representing the levels for the control parameter. A generic matrix representation of CP is given in (1)

$$CP = \begin{bmatrix} cp_{1,1} & \cdots & cp_{N_2,1} \\ \vdots & \ddots & \vdots \\ cp_{1,l_1} & \cdots & cp_{N_2,l_{N_2}} \end{bmatrix} \quad (1)$$

Using these notations, the aim is to determine the probability of capturing the defects in set- D , when each control parameter in set- C is operated at a particular level in matrix CP. Therefore, for each individual defect- d , the PMF for various combinations of the control parameters needs to be computed. This is denoted as $PMF_{d,cp_{j_1,1},cp_{j_2,2},\dots,cp_{j_{N_2},N_2}}$ where $\{d \mid d \in D, D = \{1, \dots, N_1\}\}, \{cp_{j_1,1}, \dots, cp_{j_{N_2},N_2} \mid cp_{j,c} \in CP \text{ such that } 0 < j_1 \leq l_1, \dots, 0 < j_{N_2} \leq l_{N_2}\}$.

For effective communication of the above formulation, consider a small example where there are two defects $D = \{1, 2\}$ which need to be determined by two control parameters $C = \{1, 2\}$. Say, the first control parameter has 2 levels and the second control parameter has three levels, then $L = \{2, 3\}$. Given these, the CP matrix is represented in (2)

$$CP = \begin{bmatrix} cp_{1,1} & cp_{2,1} \\ cp_{1,2} & cp_{2,2} \\ - & cp_{2,3} \end{bmatrix} \quad (2)$$

For this, the total number of PMFs that will be determined for defect-1 is 6, and these are as follows:

- $PMF_{1,cp_{1,1},cp_{1,2}}$ PMF for defect-1 when control parameter-1 is in level-1 and control parameter-2 is in level-1
- $PMF_{1,cp_{1,1},cp_{2,2}}$ PMF for defect-1 when control parameter-1 is in level-1 and control parameter-2 is in level-2
- $PMF_{1,cp_{1,1},cp_{3,2}}$ PMF for defect-1 when control parameter-1 is in level-1 and control parameter-2 is in level-3
- $PMF_{1,cp_{2,1},cp_{1,2}}$ PMF for defect-1 when control parameter-1 is in level-2 and control parameter-2 is in level-1
- $PMF_{1,cp_{2,1},cp_{2,2}}$ PMF for defect-1 when control parameter-1 is in level-2 and control parameter-2 is in level-2
- $PMF_{1,cp_{2,1},cp_{3,2}}$ PMF for defect-1 when control parameter-1 is in level-2 and control parameter-2 is in level-3.

Similarly, there are six PMFs for defect-2. Therefore, the total number of PMFs to be computed for a general problem can be defined as given by formula in (3)

$$\text{Total number of PMFs} = (l_1 * l_2 * l_3 * \dots * l_{N_2}) * N_1 \quad (3)$$

For computing these PMFs, the past test data on similar products are utilised. The challenge is that the past data is a single sample. For determining the PMF for each defect in set- D , at various level combinations in set- L for the N_2 control parameters in set- C , a bootstrapping approach is proposed. Here, for a specific combination of levels for all the control parameters, bootstrapped samples are taken from the past sample. For each bootstrapped sample, the occurrence or non-occurrence of various defects is noted. The procedure is repeated a large number of times. Next, the respective probabilities for the different defects are computed as the ratio of the number of occurrences to the total number of bootstrapped samples drawn. This provides an estimate of the probability to capture a defect when the different control parameters are operated at a specific combination of levels. These determined probabilities are the PMFs of interest. Using these probabilities, the CDFs can be computed as specified in (4)

$$\begin{aligned} & PMF_{d, cp_{j1,1}, cp_{j2,2}, \dots, cp_{jN_2, N_2}} \\ & \text{number of samples in which defect} - d \text{ was noted when the control} \\ & \text{parameters are at a specified combination} \\ = & \frac{\text{Total numbers of samples taken at the specified}}{\text{combinations of control parameter}} \\ CDF_{d, cp_{j,c}, \dots, cp_{j',c'}} &= \sum_{a=i'} \sum_{b \leq j'} PMF_{d, cp_{j1,1}, \dots, cp_{b,a}} \forall cp_{j,c} \text{ and } cp_{j',c'} \in CP \end{aligned} \quad (4)$$

For the small example used for effective communication previously, the different CDFs for the first defect would be as specified in the equations below:

$$\begin{aligned} CDF_{1, cp_{1,1}, cp_{1,2}} &= PMF_{1, cp_{1,1}, cp_{1,2}} \\ CDF_{1, cp_{1,1}, cp_{2,2}} &= PMF_{1, cp_{1,1}, cp_{1,2}} + PMF_{1, cp_{1,1}, cp_{2,2}} \\ CDF_{1, cp_{1,1}, cp_{3,2}} &= PMF_{1, cp_{1,1}, cp_{1,2}} + PMF_{1, cp_{1,1}, cp_{2,2}} + PMF_{1, cp_{1,1}, cp_{3,2}} \\ CDF_{1, cp_{2,1}, cp_{1,2}} &= PMF_{1, cp_{1,1}, cp_{1,2}} + PMF_{1, cp_{2,1}, cp_{1,2}} \\ CDF_{1, cp_{2,1}, cp_{2,2}} &= PMF_{1, cp_{1,1}, cp_{1,2}} + PMF_{1, cp_{2,1}, cp_{1,2}} + PMF_{1, cp_{1,1}, cp_{2,2}} + PMF_{1, cp_{2,1}, cp_{2,2}} \\ CDF_{1, cp_{2,1}, cp_{3,2}} &= PMF_{1, cp_{1,1}, cp_{1,2}} + PMF_{1, cp_{2,1}, cp_{1,2}} + PMF_{1, cp_{1,1}, cp_{2,2}} + PMF_{1, cp_{2,1}, cp_{2,2}} \\ & \quad + PMF_{1, cp_{1,1}, cp_{3,2}} + PMF_{1, cp_{2,1}, cp_{3,2}} \end{aligned}$$

Similarly, there are six CDFs for defect-2. Therefore, the total number of CDFs to be computed for a problem can be defined by the same formula specified in (3) that is used to compute the total number of PMFs.

Once the PMFs and CDFs are computed, they can be used to determine the ideal level combinations for the different control parameters, such that all the defects can be identified with the desired level of capture probability.

The determination of the ideal combination for the control parameters is challenging, given the large number (counts) of PMFs and CDFs. To address this, a binary linear programming optimisation model is developed. In this model, the objective is to minimise the costs associated with each level for the various control parameters. The

model is constrained with a lower bound on the desired probability of capturing the different defects. Let $P_{cp_{j,c}, \dots, cp_{j',d}}$ indicate the price associated with performing the defect detection process when control parameter- c is operated in level- j and control parameter- i' at level- j' . Let $Y_{cp_{j,c}, \dots, cp_{j',d}}$ be a binary variable indicating the combination in which the test is going to be performed (level of operation for a control parameter- c is j and level of operating control parameter- i' is j'). Thus, if $Y_{cp_{j,c}, \dots, cp_{j',d}}$ is 1 then the test is performed with control parameter- c operating at level- j and control parameter- i' operating at level- j' . It is 0 otherwise. Therefore, the objective of reducing the cost associated with performing the defective detection can be mathematically represented as given in (5)

$$\text{Min } Z = \sum_{CP} P_{cp_{j,c}, \dots, cp_{j',d}} * Y_{cp_{j,c}, \dots, cp_{j',d}} \quad (5)$$

This objective is constrained by specifying the required level of confidence with which each defect- d must be identified. If the desired level of confidence for a defect- d is denoted by ' X_d ', then the constraint is mathematically represented as given in (6). Moreover, a control parameter can be operated at only of its levels, and there can be only combination (a level identified for every control parameter) for this test. This constraint is mathematically incorporated as given in (7). The readers may note that the suffix 'CP' used in 'summation-terms' of (5), (6) and (7) refers to the elements in the CP-Matric defined earlier as $\{cp_{j_1,1}, \dots, cp_{j_{N_2}, N_2} \mid cp_{j,c} \in CP \text{ such that } 0 < j_1 \leq l_1, \dots, 0 < j_{N_2} \leq l_{N_2}\}$

$$\sum_{CP} P_{cp_{j,c}, \dots, cp_{j',d}} * Y_{cp_{j,c}, \dots, cp_{j',d}} * CDF_{d, cp_{j,c}, \dots, cp_{j',d}} \geq X_d \quad \forall d \quad (6)$$

$$\sum_{CP} Y_{cp_{j,c}, \dots, cp_{j',d}} = 1 \quad (7)$$

Solving this optimisation model, will result in a solution which suggests the best operating combination for the control parameters to identify all possible defects. This mathematical construct is empirically demonstrated using simulated data generated from a proposed experimental design in the next section.

5 Empirical demonstration using simulated data

A numerical instance is simulated through a proposed experimental design to illustrate the mathematical construct in Section 4. For the numerical example, the setup of a vehicle assembly line through four stages is considered. In each stage, a certain number of test vehicles are to be assembled and driven on the road to identify possible manufacturing defects through the road test. Here, in each stage, there are two control parameters. The control parameters are the production related aspects that can be controlled by the manufacturer. Accordingly, the first control parameter is the number of test vehicles to be assembled in a stage. The second control parameter is the drive distance for each test vehicle. Past data on the setup of an assembly act as a single sample which provides the upper limit on the two control parameters for every stage. The past data also has the information of defects and when they were identified. From this single sample, it is possible to identify the list of possible defects. Moreover, it is also conceivable to create sub-levels for the control parameters. From this past data, the aim is

to identify the likelihood of an error happening at a particular level combination for the two control parameters.

Hence, as an illustration, consider a scenario where 20 thinkable defects are to be identified through the two control parameters (number of test vehicles to produce and drive distance for each test vehicle). For past data, considering an assembly line on which 25 test vehicles were assembled at a given stage, each was driven for a distance of 3,000 kilometres to identify defects. It is noted that the numbers '20 defects', '25 test vehicles' and '3000 kilometres of drive distance per test vehicle' are assumed without loss of generality and for the purpose of demonstration. These can be any numbers based on actual past data. Furthermore, without loss of generality, for this illustration, the levels for the two control parameters ('Number of test vehicles assembled' and 'Drive distance for each test vehicle'), are developed considering that vehicles be assembled only in multiples of 5 and the running distance must be in multiples of 500 kilometres. This will create 5 levels for control parameter-1 'Number of test vehicles assembled' (which are 5, 10, 15, 20 and 25 vehicles) and six levels for control parameter-2 'Drive distance per test vehicle' (which are < 500 km, 501 km–1,000 km, 1,001 km–1,500 km, 1,501 km–2,000 km, 2,001 km–2,500 km and $\geq 2,501$ km).

To simulate the data for the above configuration, three variables are randomly generated. The first is the distance at which the defect is likely to occur. The second is the likeliness of the defect occurring in a particular vehicle when it is driven for a specified distance (which is identified in the previous variable). The third is the likeliness of a driver to observe the defect if it occurs in the vehicle at the specified kilometre. For effective communication, the notations 'Var1', 'Var2' and 'Var3' are used to represent the three variables.

To generate these variables, a 3-step experimental design is proposed as described below

- Step 1 Without loss of generality, to simulate Var1, the defects D1 to D10, D11 to D17 and D18 to D20 are assumed to occur at distances that follow the three normal distributions $N(400, 2,500)$, $N(700, 10,000)$ and $N(2,000, 40,000)$ respectively. Accordingly, for each vehicle-defect combination, a random number between ± 6 Sigma of the respective normal distribution is generated. This provides the distance at which the defect may occur in a particular vehicle.

For example: For defect D12, the distribution is $N(700, 10,000)$. This indicates that the minimum distance range (minus six sigma limit) where D12 can be observed is 100 km and the maximum distance range (plus six sigma limit) is 1,300 km. Within this range, a random number is generated for each vehicle for identifying the defect D12. As an example, say for vehicle-1, the obtained number for Var1 is 1,250 km.

- Step 2 Var1 and Var3 are categorical variables with two levels. Thus, a uniform distribution between 0 and 1 is used to stimulate these variables. To stimulate the probability that the defect would have occurred in that particular vehicle at a specified distance level, a random number between 0 and 1 is generated. If this number is greater than or equal to 0.5 then the defect would have occurred at that distance in that vehicle (Var2 = 1). Else it would not have occurred (Var2 = 0). Similarly, Var3, a random number between 0 and 1 is generated and if it is greater than or equal to 0.5, then the defect is identified by the driver

driving the vehicle (Var3 = 1). Else the driver does not identify the defect (Var3 = 0). It is noted that the proposed experimental design considers Var2 and Var3 for clearly distinguishing the instances where the defect would have occurred in the vehicle (Var2 = 1), but was not observed by the driver (Var3 = 0).

- Step 3 Once the three variables Var1, Var2 and Var3 are simulated. The final result is combined by multiplication. If the multiplied value is non-zero, then the control parameter-2 is set to the mapped level for that specified test vehicle. Else, it is set to zero.

For example, say in test vehicle-1 for defect D12 Var1 is 1250, Var2 is 1 and Var3 is 0. Then the corresponding level entry in Table 1 is obtained by mapping $1250 * 1 * 1 = 1,250$ with the level '1,001 to 1,500 km' of control parameter-2.

Therefore, through the above experimental design, a level for control parameter-2 is assigned for every defect-vehicle combination. A data simulation based on the above experimental design is presented in Table 1.

From Table 1, it can be understood that only defects D3, D5, D12, D15, D19 and D20 are identified when vehicle V1 is driven. The other defects are not identified by the vehicle. The readers may explore the simulated data further to understand that the defects are spread across the distribution from which they were generated.

For the simulated data in Table 1, identifying the probability of capturing a defect when vehicles are driven to a certain distance is the same as identifying the expected value (mean) of the original distribution from which the defect was generated. Moreover, this probability can vary based on the combination of levels for the two control parameters. Thus, the aim is to identify the probabilities for every possible level combination of the two control parameters. In this illustration, the total number of combinations is 30 (five levels for control parameter-1 multiplied by the six levels for control parameter-2). The various combinations are presented in Table 2. These 30 combinations represent the sub-solution spaces for the problem.

In each combination, the PMF for the 20 defects are to be determined. For this purpose, repeated samples (with replacement) that satisfy the different combinations are taken from the original sample in Table 1. These sub-samples, from the original sample, are referred to as bootstrapped samples. In this illustration, for each combination, a total of 100 bootstrapped samples are generated. In each bootstrapped sample, the occurrence or non-occurrence of a defect is noted. R-software is used to generate the bootstrapped samples and determine the PMF and CDF. To facilitate further research, the developed R-code is made available on GitHub. On completion of the above bootstrapping operation for the 30 combinations, a ratio of total occurrence of a defect to the total number of bootstrapped samples computes the expected value of the bootstrapped distribution for each combination. This expected value is the PMF for the defect at that specified combination. For the sample data provided in Table 1, the determined PMF for different defects at the various combinations of the control parameters is provided in Table 3.

Table 1 A simulated input data generated according to the proposed experimental design

Vehicle in which defect is identified	Kilometre level at which the defect is identified																			
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20
V1	NI	NI	500	NI	1,000	NI	NI	NI	NI	NI	NI	1,500	NI	NI	1,500	NI	NI	NI	3,000	3,000
V2	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	2,500
V3	NI	NI	NI	NI	500	500	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
V4	NI	NI	NI	NI	NI	NI	NI	NI	500	NI	1,000	1,500	500	1,000	NI	NI	NI	NI	NI	NI
V5	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	1,000	NI	NI	NI	NI
V6	NI	NI	NI	NI	NI	NI	NI	1,000	NI	NI	500	500	NI	NI	NI	NI	NI	3,000	NI	NI
V7	1,000	1,000	500	500	NI	0	0	1,000	NI	NI	NI	NI	NI	NI	NI	NI	NI	2,000	NI	NI
V8	500	500	500	NI	NI	1,000	500	0	NI	NI	1,000	NI	NI	NI	500	NI	NI	NI	NI	2,500
V9	NI	500	500	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI	3,000	1,500
V10	1,000	NI	NI	NI	NI	NI	NI	NI	NI	1,000	NI	NI	NI	NI	500	NI	NI	NI	2,000	NI
V11	NI	NI	NI	NI	500	NI	NI	NI	1,000	NI	500	1,000	NI	NI	NI	0	NI	NI	2,000	3,000
V12	NI	NI	500	NI	500	NI	NI	NI	500	NI	NI	NI	500	500	0	500	NI	NI	NI	NI
V13	NI	NI	NI	NI	1,000	NI	NI	NI	NI	NI	NI	1,500	0	500	0	1,000	NI	NI	2,000	NI
V14	NI	1,000	NI	NI	500	NI	500	500	NI	500	NI	NI	1,000	0	0	0	500	NI	NI	NI
V15	NI	500	NI	NI	NI	NI	NI	500	NI	NI	NI	NI	1,500	NI	500	NI	NI	NI	1,500	3,000
V16	NI	500	NI	NI	NI	NI	NI	NI	500	NI	NI	NI	1,000	1,000	NI	NI	500	1,000	1,500	NI
V17	NI	NI	500	NI	NI	NI	NI	NI	NI	NI	500	NI	NI	NI	NI	500	NI	NI	NI	NI
V18	NI	NI	NI	500	500	NI	500	NI	NI	NI	NI	1,500	NI	NI	NI	1,000	NI	2,500	NI	NI
V19	NI	NI	NI	500	NI	NI	NI	NI	500	NI	NI	NI	NI	NI	1,000	NI	NI	2,500	NI	NI
V20	NI	NI	NI	NI	NI	500	NI	NI	500	1,000	1,500	0	NI	500	NI	1,500	NI	NI	NI	NI
V21	NI	NI	NI	500	1,000	500	NI	NI	NI	NI	NI	NI	NI	NI	NI	500	NI	NI	NI	NI
V22	NI	1,000	NI	1,000	NI	500	NI	500	NI	NI	500	NI	NI	NI	NI	NI	NI	1,500	NI	NI
V23	NI	NI	1,000	500	NI	NI	500	NI	NI	500	NI	1,500	NI	NI	NI	1,500	NI	1,500	2,000	NI
V24	NI	NI	NI	NI	NI	500	NI	NI	NI	500	NI	NI	NI	NI	NI	NI	1,500	NI	1,500	NI
V25	NI	NI	500	NI	NI	NI	NI	NI	NI	NI	NI	NI	1,000	NI	NI	NI	1,000	NI	NI	NI

Note: NI – not identified.

Table 2 Different combinations of control parameters

<i>Combination ID</i>	<i>Control parameter-1: number of vehicles to be produced (in count)</i>	<i>Control parameter-2: distance to be driven per vehicle (in kilometres)</i>
ID-1	5	500
ID-2	5	1,000
ID-3	5	1,500
ID-4	5	2,000
ID-5	5	2,500
ID-6	5	3,000
ID-7	10	500
ID-8	10	1,000
ID-9	10	1,500
ID-10	10	2,000
ID-11	10	2,500
ID-12	10	3,000
ID-13	15	500
ID-14	15	1,000
ID-15	15	1,500
ID-16	15	2,000
ID-17	15	2,500
ID-18	15	3,000
ID-19	20	500
ID-20	20	1,000
ID-21	20	1,500
ID-22	20	2,000
ID-23	20	2,500
ID-24	20	3,000
ID-25	25	500
ID-26	25	1,000
ID-27	25	1,500
ID-28	25	2,000
ID-29	25	2,500
ID-30	25	3,000

In Table 3, the values highlighted in ‘BOLD’ indicate the highest PMF (the first highest PMF in case of a tie) for the various defects at various possible levels of control parameter-1 (number of test vehicles to produce). On closer examination of these highest values, it can be seen that the values at each level of control parameter-1 match (very closely) the means of the various source distributions from which the data were simulated. This implies that the above bootstrapping approach can capture the expected value at which there is a maximum possibility for capturing the defect. Using Table 3, the computed CDF for each combination is presented in Table 4.

Table 3 PMF for different defects at various combinations of the control parameters

Combination		PMF for defect at a combination of control parameters																		
ID	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20
ID-1	0.3	0.6	0.8	0.6	0.6	0.6	0.5	0.5	0.7	0.3	0.7	0.9	0.3	0.4	0.4	0.4	0.3	0	0	0
ID-2	0.3	0.3	0	0.1	0.2	0.1	0	0.2	0.1	0.3	0.2	0.1	0.3	0.3	0.2	0.3	0.2	0.2	0	0
ID-3	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.1	0.1	0.2	0.3	0.5	0.2
ID-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0.3	0
ID-5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0.3
ID-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0.3
ID-7	0.5	0.8	1	0.9	0.9	0.8	0.8	0.7	0.9	0.8	0.8	0.4	0.4	0.8	0.7	0.8	0.6	0	0	0
ID-8	0.2	0.2	0	0	0.1	0.1	0	0.1	0.1	0.1	0.1	0.2	0.4	0.1	0.1	0.1	0.2	0.3	0	0
ID-9	0	0	0	0	0	0	0	0	0	0	0.1	0.4	0.1	0	0.1	0.1	0	0.3	0.7	0.4
ID-10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0.3	0
ID-11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.3
ID-12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.2
ID-13	0.7	0.9	1	1	1	1	0.9	0.8	1	0.9	0.9	0.5	0.7	0.9	0.9	0.8	0.7	0	0	0
ID-14	0.2	0.1	0	0	0	0	0	0.2	0	0.1	0.1	0.2	0.3	0.1	0	0.2	0.1	0.5	0	0
ID-15	0	0	0	0	0	0	0	0	0	0	0	0.3	0	0	0.1	0	0.1	0.3	0.8	0.6
ID-16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0.2	0
ID-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.3
ID-18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1
ID-19	0.8	1	1	1	1	1	1	0.9	1	0.9	1	0.6	0.8	0.9	0.9	0.9	0.9	0	0	0
ID-20	0.2	0	0	0	0	0	0	0.1	0	0.1	0	0.2	0.2	0.1	0.1	0.1	0.1	0.6	0	0
ID-21	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0	0	0	0	0.3	1	0.6
ID-22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ID-23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.3
ID-24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1
ID-25	0.9	1	1	1	1	1	1	0.9	1	0.9	1	0.6	0.9	1	1	1	0.9	0	0	0
ID-26	0.1	0	0	0	0	0	0	0.1	0	0.1	0	0.3	0.1	0	0	0	0.1	0.6	0	0
ID-27	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0.4	1	0.7
ID-28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ID-29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2
ID-30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1

Table 4 CDF for different defects at various combinations of the control parameters

Combination		CDF for defect at a combination of control parameters																		
ID	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20
ID-1	0.3	0.6	0.8	0.6	0.6	0.6	0.5	0.5	0.7	0.3	0.7	0.2	0.3	0.4	0.4	0.4	0.3	0	0	0
ID-2	0.6	0.9	0.8	0.7	0.8	0.7	0.5	0.7	0.8	0.6	0.8	0.4	0.6	0.7	0.6	0.7	0.5	0.2	0	0
ID-3	0.6	0.9	0.8	0.7	0.8	0.7	0.5	0.7	0.8	0.6	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.5	0.5	0.2
ID-4	0.6	0.9	0.8	0.7	0.8	0.7	0.5	0.7	0.8	0.6	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.6	0.8	0.2
ID-5	0.6	0.9	0.8	0.7	0.8	0.7	0.5	0.7	0.8	0.6	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.8	0.8	0.5
ID-6	0.6	0.9	0.8	0.7	0.8	0.7	0.5	0.7	0.8	0.6	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.8	0.9	0.8
ID-7	0.5	0.8	1	0.9	0.9	0.8	0.8	0.7	0.9	0.8	0.8	0.4	0.4	0.8	0.7	0.8	0.6	0	0	0
ID-8	0.7	1	1	0.9	1	0.9	0.8	0.8	1	0.9	0.9	0.6	0.8	0.9	0.8	0.9	0.8	0.3	0	0
ID-9	0.7	1	1	0.9	1	0.9	0.8	0.8	1	0.9	1	1	0.9	0.9	0.9	1	0.8	0.6	0.7	0.4
ID-10	0.7	1	1	0.9	1	0.9	0.8	0.8	1	0.9	1	1	0.9	0.9	0.9	1	0.8	0.8	1	0.4
ID-11	0.7	1	1	0.9	1	0.9	0.8	0.8	1	0.9	1	1	0.9	0.9	0.9	1	0.8	0.9	1	0.7
ID-12	0.7	1	1	0.9	1	0.9	0.8	0.8	1	0.9	1	1	0.9	0.9	0.9	1	0.8	1	1	0.9
ID-13	0.7	0.9	1	1	1	1	0.9	0.8	1	0.9	0.9	0.5	0.7	0.9	0.9	0.8	0.7	0	0	0
ID-14	0.9	1	1	1	1	1	0.9	1	1	1	1	0.7	1	1	0.9	1	0.8	0.5	0	0
ID-15	0.9	1	1	1	1	1	0.9	1	1	1	1	1	1	1	1	1	0.9	0.8	0.8	0.6
ID-16	0.9	1	1	1	1	1	0.9	1	1	1	1	1	1	1	1	1	0.9	0.9	1	0.6
ID-17	0.9	1	1	1	1	1	0.9	1	1	1	1	1	1	1	1	1	0.9	1	1	0.9
ID-18	0.9	1	1	1	1	1	0.9	1	1	1	1	1	1	1	1	1	0.9	1	1	1
ID-19	0.8	1	1	1	1	1	1	0.9	1	0.9	1	0.6	0.8	0.9	0.9	0.9	0.9	0	0	0
ID-20	1	1	1	1	1	1	1	1	1	1	1	0.8	1	1	1	1	1	0.6	0	0
ID-21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.9	1	0.6
ID-22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.9	1	0.6
ID-23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.9
ID-24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ID-25	0.9	1	1	1	1	1	1	0.9	1	0.9	1	0.6	0.9	1	1	1	0.9	0	0	0
ID-26	1	1	1	1	1	1	1	1	1	1	1	0.9	1	1	1	1	1	0.6	0	0
ID-27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.7
ID-28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.7
ID-29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.9
ID-30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

From Table 4, it is necessary to identify the best combination of control parameters that would result in the desired level for the probability of capturing a defect. For this purpose, a binary optimisation model is developed. In the model, the decision to choose a combination such that the different defects are identified with a probability that is at least that of the desired level. This translates to the mathematical requirement that the CDF must be greater than or equal to the desired level of probability for each error at the chosen combination. In the binary optimisation model, the objective function is the cost of conducting the test, and the aim is minimisation of cost. This cost is the combination of the number of test vehicles to produce and the drive distance per test vehicle. Without loss of generality, the cost for manufacturing a single vehicle is considered Rs. 15,000 and a cost of Rs. 10/km is considered to be incurred when the vehicle is driven. Accordingly, assuming a common desired probability of 0.9 (90% chance of occurrence) for all defects, the results in Table 4 are modelled as the binary model presented in Section 4 [equations (5), (6) and (7)] and solved in a LINGO 11.0 Solver. For the benefit of readers, the developed LINGO Set-Code to generate the proposed model is provided in GitHub. From the solution, for the simulated data the ideal combination of control parameters is identified to be 15 test vehicles to produced and driven for a distance of 2,500 kilometres per test vehicle. This will ensure that all 20 defects could have occurred in the road test at a given stage in the ramp-up process with a probability of 90%. Thus, control parameter-1 can be operated at its third level and control parameter-2 at its fifth level for achieving the degree of capture probability.

In the next section, the managerial insights of the proposed approach are presented.

6 Managerial insights

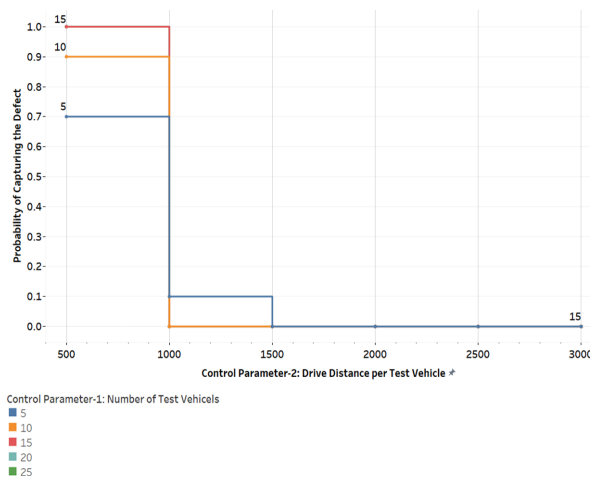
The inferences from the empirical illustration in Section 5 are first presented in brief.

- 1 The recommended combination of control parameters suggests the production of 15 vehicles and a drive distance of 2,500 kilometres per vehicle to ensure that all defects have a 90% chance of getting identified.
- 2 The reduction in operation level of the control parameters reduces the cost for the road test from Rs. 1,125,000 to Rs. 600,000. Thus a savings of Rs. 525,000 is realised.
- 3 A sample plot of the PMFs for defects 'D4', 'D13' and 'D20' respectively is presented in Figures 1, 2 and 3, respectively. This indicates the ability of the approach to identify the mean distance for the different defects, at different possible control parameter combinations. Furthermore, the figures also indicate the change in capture probability for various combination of control parameter levels (sub-solution spaces).
- 4 The CDF graph samples for 'D4', 'D13' and 'D20' in Figures 4, 5 and 6 emphasises that the capture probability of the maximum size probability space (25 vehicles driven for 3,000 kilometres each) is approximated by the sub-sample spaces generated by the proposed approach.

The above inferences lead to the following managerial insights:

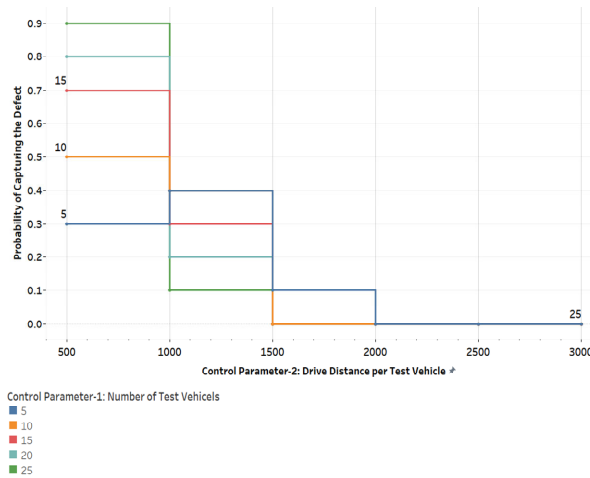
- 1 While setting up an assembly line, identifying an ideal sub-solution space leads to significant savings in cost and resources for the defect detection test.
- 2 The PMF for a defect help in identifying the control parameter combinations which yield same chance of capturing the defect. Hence, the management can do a deep dive to understand the impact of a control parameter on a defect.
- 3 It is possible that the CDF's for certain defects may plateau (reach maximum) after a certain combination of the control parameters. The plateaus indicate the different combination of control parameter levels over which the same capture probability may be achieved. This phenomenon is extremely interesting from a management perspective as it provides opportunity to incorporate further considerations for choice of operating levels.

Figure 1 PMF plot of defect 'D4' at various levels of control parameters 1 and 2 (see online version for colours)

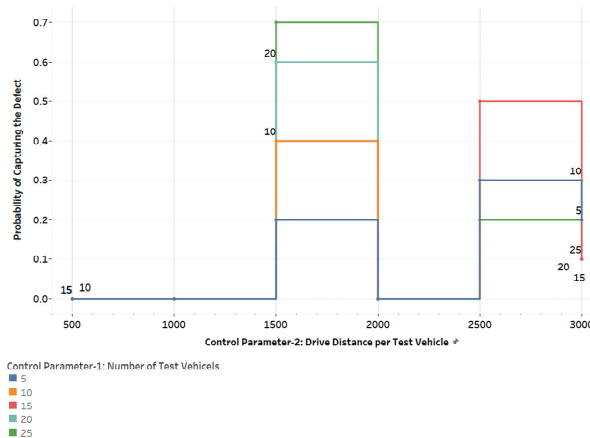


Notes: A Cartesian coordinate graph with control parameter-2 on the X-axis and probability of occurrence (PMF) on the Y-axis for defect 'D4' and lines of different colours for control parameter-1. A line graph drawn with the levels of control parameter-2 on the X-axis and the different probability of occurrence (PMF) in the Y-axis for defect 'D4'. Each line in the graph is for a particular level of control parameter-1. The points on each line indicate the probability of capturing the particular defect when control parameters are operated in that specific level combination.

It is imperative to note that the choice for operation level of a control parameter is greatly influenced by their respective cost. In the empirical example, given that the cost of manufacturing a vehicle is greater than the cost of driving the vehicle for additional distance, the results are always skewed towards first reducing the number of vehicles. Only after identifying the reduced number of vehicles, the distance is reduced. Therefore, the management has the power to emphasise the importance of a decision criteria based on the requirement during the choice phase by appropriately awarding the costs associated with the control parameter levels.

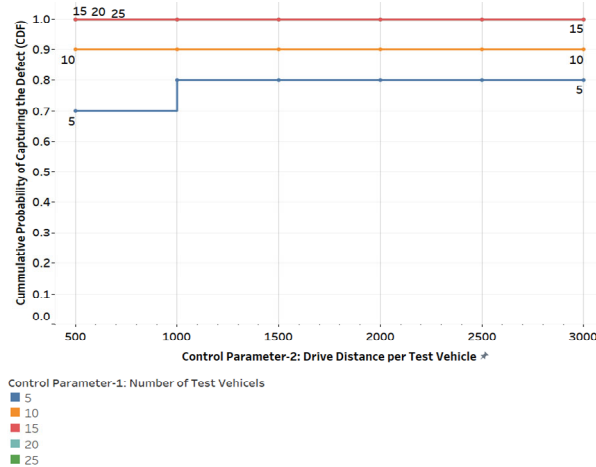
Figure 2 PMF plot of defect 'D13' at various levels of control parameters 1 and 2 (see online version for colours)

Notes: A Cartesian coordinate graph with control parameter-2 on the X-axis and probability of occurrence (PMF) on the Y-axis for defect 'D13' and lines of different colours for control parameter-1. A line graph drawn with the levels of control parameter-2 on the X-axis and the different probability of occurrence (PMF) in the Y-axis for defect 'D13'. Each line in the graph is for a particular level of control parameter-1. The points on each line indicate the probability of capturing the particular defect when control parameters are operated in that specific level combination.

Figure 3 PMF plot of defect 'D20' at various levels of control parameters 1 and 2 (see online version for colours)

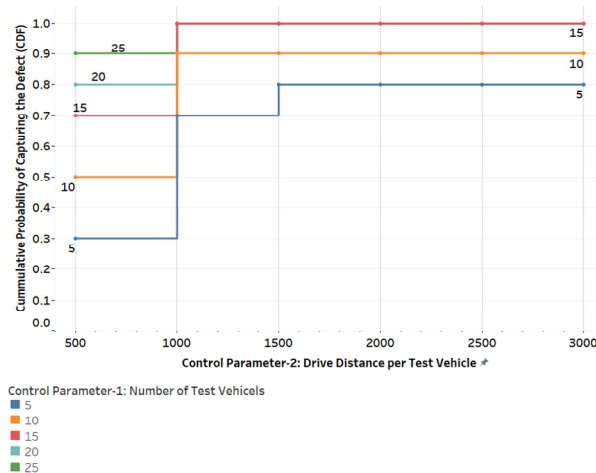
Notes: A Cartesian coordinate graph with control parameter-2 on the X-axis and probability of occurrence (PMF) on the Y-axis for defect 'D20' and lines of different colours for control parameter-1. A line graph drawn with the levels of control parameter-2 on the X-axis and the different probability of occurrence (PMF) in the Y-axis for defect 'D20'. Each line in the graph is for a particular level of control parameter-1. The points on each line indicate the probability of capturing the particular defect when control parameters are operated in that specific level combination.

Figure 4 CDF plot for defects 'D4' at various levels of control parameters 1 and 2 (see online version for colours)



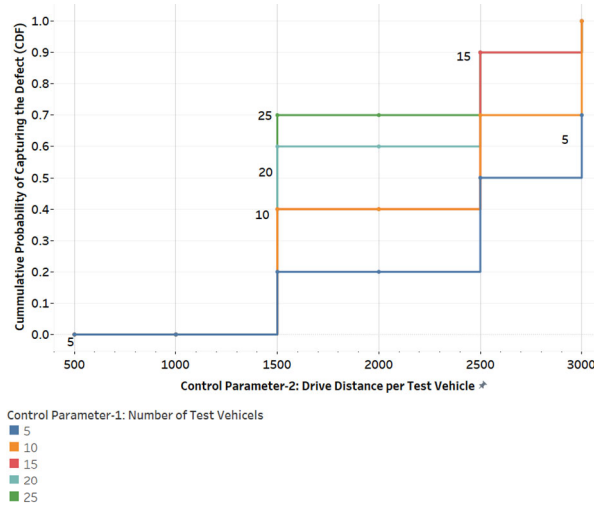
Notes: A Cartesian coordinate graph with control parameter-2 on the X-axis and CDF for defect 'D4' on the Y-axis and lines of different colours for control parameter-1. A line graph drawn with the levels of control parameter-2 on the X-axis and the CDF for defect 'D4' on the Y-axis. Each line in the graph is for a particular level of control parameter-1. The points on each line indicate the cumulative probability of capturing the particular defect when control parameters are operated in that specific level combination.

Figure 5 CDF plot for defects 'D13' at various levels of control parameters 1 and 2 (see online version for colours)



Notes: A Cartesian coordinate graph with control parameter-2 on the X-axis and CDF for defect 'D13' on the Y-axis and lines of different colours for control parameter-1. A line graph drawn with the levels of control parameter-2 on the X-axis and the CDF for defect 'D13' on the Y-axis. Each line in the graph is for a particular level of control parameter-1. The points on each line indicate the cumulative probability of capturing the particular defect when control parameters are operated in that specific level combination.

Figure 6 CDF plot for defects ‘D20’ at various levels of control parameters 1 and 2 (see online version for colours)



Notes: A Cartesian coordinate graph with control parameter-2 on the X-axis and CDF for defect ‘D20’ on the Y-axis and lines of different colours for control parameter-1. A line graph drawn with the levels of control parameter-2 on the X-axis and the CDF for defect ‘D20’ on the Y-axis. Each line in the graph is for a particular level of control parameter-1. The points on each line indicate the cumulative probability of capturing the particular defect when control parameters are operated in that specific level combination.

7 Conclusions

A business problem of defect detection while setting up an assembly line in stages is explored in this study. A stage in the setup of assembly lines refers to the incremental increase in capacity to produce products as compared to the previous stage. For this study, only manufacturing defects in a product which may be detected by the customer are considered. The current practice to identify such manufacturing defects is conducted by producing a certain number of products to be tested under specified conditions. The number of products to be produced and production conditions together constitute the control parameters for the test. At present, the operating levels for the control parameters are fixed based on management recommendations. Therefore, the primary focus of this study is to propose an analytical approach to recommend operating levels for the control parameters, such that defects are detected with a desired level of satisfaction.

Accordingly, the problem is addressed through the prism of identifying the solution space where past data on defects and control parameters can be utilised to determine the probability of capturing a specific defect at a particular operating level combination for the control parameters. The past data on possible defects specifies the occurrence or non-occurrence of different defects when the control parameters are operated at their maximum possible levels. From this, all sub-solution spaces (different combination of possible control parameter levels) are identified. For each identified sub-solution space, the probability of capturing each defect at that solution space is determined. This

determination of probability for each defect is conducted by generating bootstrapped sampling distributions for the given level combination of control parameters. The mean of the bootstrapped sampling distribution is the probability (PMF) for capturing the defect for a specific sub-solution space. The PMFs for a defect can be utilised to determine the CDF for the same defect when the level of a particular control parameter is changed and the levels of all other control parameters are held constant.

Once all defects PMFs across all sub-solution spaces have been determined, the question of what should be the sub-solution space is to be operated is addressed. Since there are a large number of possible combinations, a binary optimisation model is developed. The optimisation model takes the CDFs as input, along with the requirement of desired confidence for capturing each defect. The output of the model is a single sub-solution space which specifies the levels at which each of the control parameters need to operate in order to capture all defects at their desired confidence. The recommendation is made with the objective of minimising the test cost, which is the summation of costs associated with operating the control parameters at the specified level combination. The optimisation model is constrained by specifying the minimum probability that is expected for capturing each of the defects and that only possible combination of control parameter levels is permissible.

The proposed analytical approach is empirically demonstrated with data simulated through an experimental design. The demonstration establishes the usability, accuracy and economic effectiveness of the proposed approach. In addition, insights for the management is also presented.

The study, though comprehensive in assisting defect detection while setting up an assembly line in stages, has several future research prospects. Three most intuitive and interesting research directions are discussed here. The first is the possibility to relax the assumption that different defects are independent of each other. In most real-life scenarios, presence of one defect may cause or lead to another. This leads to a possibility of co-relation between defects. The second is that the study focuses on a stage in isolation as compared to other stages. This implies that a stage and the defects that occur in that stage are independent of the other stages. However, it may be possible to update the probability of occurrences of defects in future stages based on the occurrence of defects in the current stage. Thus, Bayesian based approaches may be explored to determine conditional probabilities for defects based on previous stages in different sub-solution spaces. It is noted that a Bayesian based approach may also be developed to address the first future scope, by identifying the conditional probability of capturing a defect given the occurrence of another defect. Such an approach may lead to better judgements and increased benefits. Finally, the CDFs and PMFs for the different defects may taper off at different combination of sub-solution spaces. This implies that it may be efficient to determine operation level combinations of sub-solution spaces, wherein the spaces can be dynamically reduced based on the capture of different defects. For this, the optimisation model of the current work can be modified such that different combinations of sub-solution spaces for different defects are suggested. This approach can lead to increased economic benefits. Additionally, two limitations to the current study are:

- 1 the exclusive focus on assembly lines which are setup in stages
- 2 the requirement of past data from similar assembly line setups to identify possible defects, possible control parameters and the levels for the control parameters.

Therefore, future research may explore possibilities to address the setup of any assembly line without the dependence for data from similar assembly lines.

Competing interests

I, the author, certify that I have no affiliations with or involvement in any organisation or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Data availability statement

It is confirmed that the data supporting the findings of this study are available within the article in Tables 1 through 4.

The codes developed as part of the study are available on GitHub: <https://github.com/DrMilo100/Defect-Detection-while-Setting-up-an-Assembly-Line-/tree/main>

References

- Akhtar, S., Tandiya, A., Moussa, M. and Tarry, C. (2019) 'An efficient automotive paint defect detection system', *Advances in Science, Technology and Engineering Systems Journal*, Vol. 4, No. 3, pp.171–182, DOI: <http://dx.doi.org/10.25046/aj040323>.
- Arnal, L., Solanes, J.E., Molina, J. and Tornero, J. (2017) 'Detecting dings and dents on specular car body surfaces based on optical flow', *Journal of Manufacturing Systems*, Vol. 45, pp.306–321, DOI: <https://doi.org/10.1016/J.JMSY.2017.07.006>.
- Chen, Y., Peng, X., Kong, L., Dong, G., Remani, A. and Leach, R. (2021) 'Defect inspection technologies for additive manufacturing', *International Journal of Extreme Manufacturing*, Vol. 3, pp.1–21, DOI: <https://doi.org/10.1088/2631-7990/abe0d0>.
- Chung, K. and Lee, S.M.S. (2001) 'Optimal bootstrap sample size in construction of percentile confidence bounds', *Scandinavian Journal of Statistics*, Vol. 28, No. 1, pp.225–239, DOI: <https://doi.org/10.1111/1467-9469.00233>.
- Czimmermann, T., Mario Milazzo, G.C., Chiurazzi, M., Roccella, S., Oddo, C.M. and Dario, P. (2020) 'Visual-based defect detection and classification approaches for industrial applications – a survey', *Sensors*, Vol. 25, DOI: <https://doi.org/10.3390/s20051459>.
- Elhabashy, A.H., Dastoorian, R., Wells, L.J. and Camelio, J.A. (2021) 'Random sampling strategies for multivariate statistical process control to detect cyber-physical manufacturing attacks', *Quality Engineering*, Vol. 33, No. 2, pp.300–317, DOI: <https://doi.org/10.1080/08982112.2020.1838541>.
- Gao, X., Zhou, X., Wang, C., Ma, N., Zhang, Y. and You, D. (2020) 'Skin depth and detection ability of magneto-optical imaging for weld defects in alternating magnetic field', *Journal of Manufacturing Systems*, Vol. 55, pp.44–55, DOI: <https://doi.org/10.1016/J.JMSY.2020.02.006>.
- Gine, E. and Zinn, J. (1990) 'Bootstrapping general empirical measures', *The Annals of Probability*, Vol. 18, No. 2, pp.851–869, DOI: <http://www.jstor.org/stable/2244320>.
- Jiang, B.C. and Wang, Y.M.W.C.C. (2001) 'Bootstrap sampling technique applied to the PCB golden fingers defect classification study', *International Journal of Production Research*, Vol. 39, No. 10, pp.2251–2230, DOI: <https://doi.org/10.1080/00207540110040015>.

- Khan, M.F., Alam, A., Siddiqui, M.A., Alam, M.S., Rafat, Y., Salik, N. and Al-Saidan, I. (2021) 'Real-time defect detection in 3D printing using machine learning', *Materials Today: Proceedings*, Vol. 42, pp.521–528, DOI: <https://doi.org/10.1016/j.matpr.2020.10.482>.
- Kuo, R.J. and Nursyahid, F.F. (2022) 'Foreign objects detection using deep learning techniques for graphic card assembly line', *Journal of Intelligent Manufacturing*, in press, DOI: <https://doi.org/10.1007/s10845-022-01980-7>.
- Lee, S. and Vachtsevanos, G. (2002) 'An application of rough set theory to defect detection of automotive glass', *Mathematics and Computers in Simulation*, Vol. 60, pp.225–231, DOI: [https://doi.org/10.1016/S0378-4754\(02\)00017-4](https://doi.org/10.1016/S0378-4754(02)00017-4).
- Li, C., Li, J., Li, Y., He, L., Fu, X. and Chen, J. (2021a) 'Fabric defect detection in textile manufacturing: a survey of the state of the art', *Security and Communication Networks*, Vol. 2021, pp.1–13, DOI: <https://doi.org/10.1155/2021/9948808>.
- Li, R., Jin, M. and Paquit, V.C. (2021b) 'Geometrical defect detection for additive manufacturing with machine learning models', *Materials & Design*, Vol. 206, pp.1–11, DOI: <https://doi.org/10.1016/J.MATDES.2021.109726>.
- Liu, J., Guo, F., Gao, H., Li, M., Zhang, Y. and Zhou, H. (2021) 'Defect detection of injection molding products on small datasets using transfer learning', *Journal of Manufacturing Processes*, Vol. 70, pp.400–413, DOI: <https://doi.org/10.1016/j.jmapro.2021.08.034>.
- Nurani, R.K., Akella, R. and Strojwas, A.J. (1994) 'Development of an optimal sampling strategy for wafer inspection', *International Symposium on Semiconductor Manufacturing, Extended Abstracts of ISSM*, pp.143–146, DOI: 10.1109/ISSM.1994.729441.
- Nurani, R.K., Akella, R. and Strojwas, A.J. (1996) 'In-line defect damping methodology in yield management: an integrated framework', *IEEE Transactions on Semiconductor Manufacturing*, Vol. 9, pp.506–517, DOI: 10.1109/66.542166.
- Paul, K. (2016) 'Car warranty: a guide to which brand provides the best warranty in India', *Drive Spark*, 29 July [online] <https://www.drivespark.com/four-wheelers/2016/car-warranty-india-brand-wise-list-016016.html> (accessed 22 September 2021).
- Phu, J., Bui, B.V., Kalloniatis, M. and Khuu, S.K. (2018) 'How many subjects are needed for a visual field normative database? A comparison of ground truth and bootstrapped statistics', *Translational Vision Science & Technology*, Vol. 7, pp.1–15, DOI: <https://doi.org/10.1167/tvst.7.2.1>.
- Ren, J., Ren, R., Green, M. and Huang, X. (2019) 'Defect detection from X-ray images using a three-stage deep learning algorithm', *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, pp.1–4, DOI: 10.1109/CCECE.2019.8861944.
- Shin, W. and Lee, S. (2008) 'A development of accelerated life test method for blower model for automobile using inverse power law model', *International Journal of Modern Physics B*, Vol. 22, pp.1074–1080, DOI: <https://doi.org/10.1142/S0217979208046347>.
- Singh, S.A., Kumar, A.S. and Desai, K.A. (2023) 'Comparative assessment of common pre-trained CNNs for vision-based surface defect detection of machined components', *Expert Systems with Applications*, Vol. 218, DOI: <https://doi.org/10.1016/j.eswa.2023.119623>.
- Singh, S.A. and Desai, K.A. (2022) 'Automated surface defect detection framework using machine vision and convolutional neural network', *Journal of Intelligent Manufacturing*, in press, DOI: <https://doi.org/10.1007/s10845-021-01878-w>.
- Suo, X., Liu, J., Dong, L., Shengfeng, C., Enhui, L. and Ning, C. (2022) 'A machine vision-based defect detection system for nuclear-fuel rod groove', *Journal of Intelligent Manufacturing*, Vol. 33, pp.1649–1663, DOI: <https://doi.org/10.1007/s10845-021-01746-7>.
- Tsarouhas, P. and Liberopoulos, G. (2004) 'Determining quality inspection frequency in an automated production line based on field failure data analysis', *Operational Research – An International Journal*, Vol. 4, No. 3, pp.305–315, DOI: <https://doi.org/10.1007/BF02944148>.
- Tulbure, A-A., Tulbure, A-A. and Dulf, E-H. (2022) 'A review on modern defect detection models using DCNNs – deep convolution neural networks', *Journal of Advanced Research*, Vol. 35, pp.33–48, DOI: <https://doi.org/10.1016/j.jare.2021.03.015>.

- Wang, M.R. and Cheung, C.F. (2021) 'CentreNet-based defect detection for additive manufacturing', *Expert Systems with Applications*, in press, DOI: <https://doi.org/10.1016/j.eswa.2021.116000>.
- Wang, S., Yan, S., Li, L., Kuo, C., Ai, J., Shen, Q., Wang, D., Ding, S. and Xia, Q. (2021) 'A dynamic sampling algorithm based on cost-risk assessment model in semiconductor manufacturing', *2021 China Semiconductor Technology International Conference (CSTIC)*, pp.1–4, DOI: 10.1109/CSTIC52283.2021.9461450.
- Wetherill, B.G. (1977) *Sampling Inspection and Quality Control*, Chapman and Hall Ltd., New York, DOI: 10.1007/978-1-4899-6858-6.
- Wiel, S.A.V. and Vardeman, S.B. (1994) 'A discussion of all-or-none inspection policies', *Technometrics*, Vol. 36, pp.102–109, DOI: <https://doi.org/10.1080/00401706.1994.10485405>.
- Woo, S. (2016) 'Reliability design of mechanical system and its parametric accelerated life testing', in Makhlof, A.S.H. and Aliokhazraei, M. (Eds.): *Handbook of Materials Failure Analysis with Case Studies from the Chemicals, Concrete and Power Industries*, pp.259–276, Elsevier Ltd., DOI: <https://doi.org/10.1016/B978-0-08-100116-5.00011-9>.
- Yang, J., Li, S., Wang, Z. and Yang, G. (2019) 'Real-time tiny part defect detection system in manufacturing using deep learning', *IEEE Access*, Vol. 7, pp.89278–89291, DOI: 10.1109/ACCESS.2019.2925561.
- Zhang, T., Ding, B., Zhao, X., Liu, G. and Pang, Z. (2021) 'LearningADD: machine learning based acoustic defect detection in factory automation', *Journal of Manufacturing Systems*, Vol. 60, pp.48–58, DOI: <https://doi.org/10.1016/j.jmsy.2021.04.005>.
- Zhu, X., Liu, K., He, M. and Balakrishnan, N. (2021) 'Reliability estimation for one-shot devices under cyclic accelerated life-testing', *Reliability Engineering and System Safety*, Vol. 212, pp.1427–2222, DOI: <https://doi.org/10.1016/j.res.2021.107595>.