

## **QUALITY RISK ANALYSIS AT SAMPLING STATIONS CROSSED BY ONE MONITORED PRODUCT AND AN UNMONITORED FLOW**

Anna Rotondo  
John Geraghty  
Paul Young

Enterprise Process Research Centre,  
School of Mechanical and Manufacturing Engineering, Dublin City University  
Dublin 9, IRELAND

### **ABSTRACT**

When inspection economies are implemented in multi-product, multi-stage, parallel processing manufacturing systems, there exists a significant risk of losing control of the monitoring efficacy of the sampling strategy adopted. For a product-based sampling decision limited to a particular station in a production segment, the randomness of the departure process and the merging of different product flows at the machines of the different stations subvert the regularity of deterministic sampling. The risk of not regularly monitoring any machine in the segment can be measured in terms of maximum number of consecutive unsampled items. In this study, the distribution of this measure at sampling station machines is developed for a production scenario characterized by one monitored product and an unmonitored flow and compared with the behavior of the same measure at non-sampling station machines. The prediction models illustrated prove fundamental pragmatic tools for quality management involved in sampling strategy-related decisions.

### **1 INTRODUCTION**

The underlying complexities that govern a multi-stage serial-parallel system represent a fundamental obstacle to a straightforward investigation on the effectiveness of quality control strategies used to monitor a production process. Although common in industry, this type of system has only been object of a limited number of studies (Jin, Li, and Tsung 2010). The majority of literature in the field of quality control of complex manufacturing systems dwells on the analysis of inspection allocation and sampling frequency optimization for multi-stage serial systems (Van Volssem, Dullaert, and Van Landeghem 2007; Vaghefi and Sarhangian 2009). In balancing productivity and quality control effort, several parallel machines are often assigned to a production station with a single control monitor used to examine a sample of the line flow creating complex defect propagation dynamics (Zeng and Zhou 2007; Jiang et al. 2012). For these systems, stream of variation and state space models have been developed to identify optimal chart allocation strategies (Jin, Li, and Tsung 2010; Liu, MacKay, and Steiner 2008; Li and Wang 2011).

Introduced in the field of Statistical Process Control (SPC), the ‘Quality Risk’ concept originally referred to the risk incurred by manufacturers and customers due to classification errors during quality inspections (Montgomery 2008, Kaya and Özer 2009); however, it can assume a broader application. This study, examines the risk of not receiving information on the quality status of any machine operating in a system with the desired frequency and is measured in terms of number of consecutive unsampled items processed by that machine. In semiconductor manufacturing facilities, this measure is analogous to the concept of “wafers at risk” (W@R) that is the number of wafers processed at a machine since the last inspected wafer processed at that machine. Various research studies focus on the analysis of W@R in relation to optimal dispatching rules (Rodriguez-Verjan et al. 2012), control plan design (Rodriguez-Verjan, Dauzère-Pérès, and Pinaton 2011) and defect inspection capacity (Rodriguez-Verjan, Dauzère-Pérès, and Pinaton 2012) when dynamic sampling strategies are implemented (Nduhura Munga et al. 2012). The

quality risk measure investigated in this study generalizes the W@R concept to an item level. An item can refer to any element that flows through a system at any aggregated level (e.g. single element or batch). Also, W@R specifically refers to the semiconductor industry, whereas the results found in this study can be applied to any production system similar the one described in this paper.

Analysis of the material flow dynamics is fundamental for a thorough understanding of the mechanism that transforms the properties of a global sampling strategy of a production line to the specific case of a particular machine. In order to minimize the impact that quality inspections have in terms of cost, resource and time on a production process (Tirkel, Reshef, and Rabinowitz 2009), a particular station in a production segment can be chosen to implement a deterministic sampling scheme. This choice is sometimes dictated by feasibility matters as destructive tests or time constraints can cause damages to the items which are eventually scrapped. The quality measurements taken at the sampling station are used to infer the quality status of the entire production segment. Final inspection can be considered a particular implementation of this sampling concept, where inspection at the last station in the production line prevents shipment of defective product and consequent external failure costs (Freeman 2008). Alternatively, the sampling station can be chosen from among the several stations operating in the production segment based on the relevance that a station has in terms of costs or value added to the processed items (Van Volsem, Dullaert, and Van Landeghem 2007; Vaghefi and Sarhangian 2009).

Figure 1 exemplifies the typical layout of the class of multi-product, multi-stage, parallel processing manufacturing systems analyzed here. Each station consists of a number of identical machines used to perform different operations as determined by both the type and the production stage of the product as the underlying system has re-entrant flow. Serial flow products, as is Product *A* in Figure 1, visit all the stations of the segment consecutively, and this flow forms the line segment being monitored. Product *A* items are sampled regularly as they leave the machines of the last station (the sampling station) for inspection. A deterministic sampling interval is set at each machine in the sampling station, so that every fifth (say) product *A* item processed is routed to the inspection station. Other products or other stages of product *A* (serial flow products) being processed at these stations are not part of the monitored flow segment, however they have equal priority in the queues for processing and thus disrupt the monitored flow.

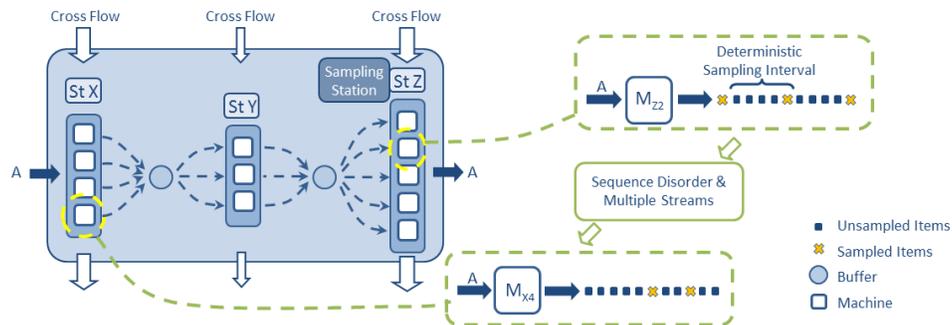


Figure 1: System layout and sampling strategy characteristics at the machines of both non-sampling and sampling stations.

When the global monitoring frequency is analyzed, significant differences exist between the station where sampling is assigned and the upstream stations. For the non-sampling stations, two dynamic flow effects interact to determine the pattern of items sampled from each machine. The presence of parallel machines operating in a station and the lack of deterministic routing policies are responsible for the multiple stream effect (Liu, MacKay, and Steiner 2008). Due to different operations required and non-deterministic routing policies, the cycle time through a segment can vary from item to item. This results in a disorder effect for the sequence of items between consecutive production stations. As a consequence, the sequence of measurements obtained at the sampling station does not necessarily reflect the evolution

of the quality status of the machines in the non-sampling stations. The detection of deteriorating trends at any machine is therefore made more difficult. If bi-univocal relationships between the machines of different stations are not defined by the routing policy, the measurements obtained at a specific machine in the sampling station cannot be exclusively referred to a particular machine operating in a non-sampling station. In production systems affected by the sequence disorder and multiple stream effects, ignoring the flow dynamics will cause misleading interpretations of the quality data at the non-sampling stations (Fan et al. 2000) as an apparently deterministic sampling strategy becomes random memory-less sampling. Recourse to data mining techniques (Fan et al. 2001), graphical (Zeng and Zhou 2007) and knowledge-based approaches (Guo et al. 1996) prove helpful for extracting useful knowledge from the quality measurements available.

Likewise, there are significant complexities associated with quality monitoring at a sampling station when more than one monitored product flows or only a portion of the work processed is monitored; the latter is the case for product A in Figure 1. The presence of an unmonitored flow (cross-flow) at the station introduces sampling randomness when the global flow is considered, which means a higher sampling variability and eventually a higher ‘quality risk’.

This study examines the number of unsampled items between consecutive samples at the machines of a sampling station for a production flow scenario that considers one monitored product and an unmonitored flow. The random behavior of this measure is compared across the different stations of the production segment. An analytical prediction model for the distribution of the number of consecutive unsampled items is developed. This represents a fundamental step for the prediction and, hence, the control of the quality risk in a production segment monitored by means of a determinist product-oriented sampling strategy. The risk of not continuously monitoring a particular machine can be effectively measured as the maximum number of items not consecutively sampled at a given confidence level. This measure is based on the percentile of the distribution of the number of unsampled items corresponding with the confidence level set as a reference risk level. Alternatively, the possibility of predicting the distribution of the number of unsampled items also proves useful for the determination of the parameters of a sampling strategy able to satisfy quality risk constraints. In either cases, the prediction models constitute fundamental pragmatic tools that can be used by quality management in real serial-parallel systems to make optimal decisions on sampling strategy-related issues in very short time.

The development of prediction models for the distribution of the number of consecutive unsampled items at a machine is based on a preliminary study that investigates the impact of some production design related factors on both the number of consecutive unsampled items and the corresponding time between consecutive samples (Rotondo, Young, and Geraghty 2012). The intersection between quality and productivity has been recognized as an important research topic as decisions taken in one often impact adversely on the other (Inman et al. 2003; Owen and Blumenfeld 2008; Colledani and Tolio 2011; Ioannidis 2012; Wang et al. 2013).

The remainder of this paper is organized as follows. Section 2 describes the characteristics of the manufacturing system on which the analysis has been focused; Section 3 gives insight into the experimental plan. Section 4 introduces some preliminary results; Section 5 describes the development of the prediction model. Conclusions are drawn in Section 6.

## **2 SIMULATION MODEL**

In order to investigate the effects of several productivity-related factors on the monitoring efficacy of the sampling plan adopted, observations of the performance metrics associated with various production scenarios were needed. Since the variations required to the system configuration proved difficult to implement in the real system without interrupting production or incurring in considerable investment costs, a simulation model of the segment was developed. The simulation model reproduces the system behavior and the material flow dynamics using a high level of detail.

## 2.1 Description of the System

Semiconductor fabrication facilities consist of complex machines laid out to facilitate easy supply and removal of process gases, liquids and waste. Multiple product types/families, re-entrant flows, complex routing and scheduling policies characterize the semiconductor industry. Despite these complexities, the production flow can be reduced to a serial flow when viewed from an operation perspective. For routine quality control purposes this flow is usually divided into production segments that consist of a series of operations and consequently a number of production stations that are monitored by an inspection station. This viewpoint is adopted in Figure 1, where the horizontal lines represent the serial flow through the production segment and the vertical lines indicate the flows of other operations (cross flow) that require capacity at the station. Each machine in the segment is regularly shut down for preventive maintenance with a frequency and duration that depends on the type of intervention. Machines are also shut down for unscheduled operational and quality failures. A quality failure causes the increase of the production rate of defective parts (Gershwin and Schick 2007), while an operational failure causes a machine to stop items production (Kim and Gershwin 2008). The duration of the repair depends on the underlying cause.

Each station is provided with an upstream buffer, shared between the machines, where items waiting to be processed in that station are gathered. The departure of the items from each buffer is not regulated by any structured queuing policy; this results in random queuing times and generates item sequence disorders at the entrance and exit of each buffer. The items waiting in a buffer can be routed to any machine of the associated station based on the machine availability. The production flow complexity of the segment is dependent on both the number of stations and the number of machines in each station. In the production segment modeled two serial flows were identified (Products *A* and *B*) and cross flow was observed at each station.

The sampling strategy implemented in the system is product-oriented and deterministic for the serial flow. A deterministic sampling interval is set for each monitored product so that every given number of items of a product type processed at a machine in the sampling station, one is chosen as a sample. Monitored products are chosen among the serial flow products. The sampling plan is independently applied to all the sampling station machines. In the case study considered here, Product *A* is sampled with a sampling interval equal to three, that is every three Product *A* items processed at a machine in the sampling station, the third item is chosen as a sample. Product *B* is sampled with a sampling interval equal to five.

## 2.2 Simulation Model Development

The simulation model of the segment described above was developed in ExtendSim V6. Being an ad-hoc model used exclusively for research purposes, individual simulation models were developed for each scenario analyzed. However, by using a modular concept where stations and item generation sources were modeled as hierarchical blocks, the implementation of any required modification to the model structure was made relatively quickly. Station blocks include an item generation source to model the cross flow, a transportation and queuing time delay block, several machine blocks and a routing decision block which routes the cross flow towards the exit block and the serial flow towards the following station. In the sampling station, a sampling decision block is introduced after each machine block to correctly select the samples; based on product type, the processed items are routed towards different item counters so that when the counter reaches the sampling interval value, the item passing through the counter is marked for inspection and directed to the inspection station; all other items leave the system directly.

A clear distinction between serial flow and cross flow is represented in the model. The serial flow, products *A* and *B*, is introduced in the first station of the segment. Items of each serial flow product are generated at exponential time intervals with differing mean intervals as established from historical factory records. On entering a station, these items are placed in the station common buffer where they wait until a machine of that station becomes available to process them. On the contrary, the cross flow at each station is modeled as a hybrid flow in the sense that there is no distinction between product types and operations

needed. Based on the observation of real data, a global exponential cross flow inter-arrival process is considered at each station. The cross flow items are introduced at a particular station and join the queue in the buffer. After being processed by one of the machines they leave the system directly. Even though such cross flow items might visit more than one station in the real segment, for simulation purposes the cross flow at each station can be modeled independently. The cross flow simulates both the utilization of resource capacity by items not included in the monitored serial flow and the increase in quality risk due to the presence of an additional unmonitored flow, cross or serial, through the machines of a station. No capacity limit is imposed to the buffers available in the segment. This assumption is justified by the presence of large capacity buffers in the real plant where the segment modeled operates and that some of the time spent queuing in the model represents time spent in transit in the real system. In order to compensate for this assumption and avoid irregular WIP build-up, the queuing times were modeled as production delays imposed on each item waiting in the buffer rather than time resulting from waiting for an available machine. As with the real system, the items produced in the simulation model are fully traceable, with the relevant data recorded during the simulation run and then exported to MS Excel™ worksheets at the end.

Distributions based on historical data are used for the simulations of 6,000 hours of production. With a warm-up period of 1500 hours removed, 5 replications were conducted with the model each time an experimental scenario was investigated. Only 5 replications per scenario were required as there was only small variability (less than 1.5% coefficient of variation) of all relevant metrics across the different simulation runs (Carson 2004). The computation time required for each simulation run was about 3 minutes on a 2GHz processor. Data from the models was averaged across the 5 simulation runs on an event-by-event basis, providing a population of about 3,000 samples behind each reported statistic. The model was validated against historical data from the real system and management at the facility verified that the model results were credible. Particular attention was paid to the metrics considered in this study, and goodness-of-fit tests performed on the distributions of the metrics obtained from both the historical data and the simulation results proved that the model was able to generate accurate results (Sargent 2010). To ensure that the modeling assumptions were valid other results (e.g. the impact of imposing queuing times as processing delays on the effective queuing times) were also analyzed. This validation analysis proved that the production flow was correctly simulated. For confidentiality reasons, the numerical aspects of this specific validation analysis are not reported here.

### 3 EXPERIMENTAL DESIGN

To develop the prediction model of the distribution on the number of consecutive unsampled items at the machines of a sampling station, eight different scenarios were simulated. As presented in Table 1, these scenarios differ from each other with respect to the values of three factors: the line speed, the line configuration and the sampling intervals of the two monitored product types.

Table 1: Parameters settings for the scenarios simulated.

Scenario	WIP turn	Number of Machines							Product Sampling Intervals	
		Factor Level	In detail						A	B
			St 1	St 2	St 3	St 4	St 5	Insp.		
1	Low	Low	4	4	3	3	5	4	Large	
2									Small	
3	High		Large							
4			Small							
5	Low	High	5	5	3	4	6	5	Large	
6									Small	
7	High		Large							
8			Small							

The experimental plan was initially developed as a  $2^3$  factorial plan, with each of the parameters having two different levels. In the plant, line speed is measured in terms of WIP-turn, which is a measure of how fast the WIP in a line is turned over. Using this concept, variations of line speed can be obtained by controlling the queuing time. In order to apply queuing time reductions in a manner representative of the real segment modeled, two different strategies had to be adopted. For scenarios 3 & 4, where the station capacities are the same as the actual system (scenarios 1 and 2), queuing time reductions could only be made by optimizing the inventory release strategy. Here the increased line speed is a consequence of lower inventory levels and lower inter-arrival rates. In other words, the system is made leaner. For scenarios 7 and 8, the line speed increase can be achieved by exploiting the additional capacity available (e.g. extra number of machines) in the stations and, for these scenarios, the station utilization significantly increases with respect to the utilization observed in scenarios 5 and 6. Both strategies are able to generate a 35% reduction in queuing time with respect to the Low WIP-turn scenarios and, as a consequence, 35% line speed increase. Even though this is not orthodox from a statistical perspective, the choice of modifying the WIP-turn using two different strategies for different levels of other factors is justified by the purpose of the Design of Experiments (DOE) and the Analysis of Variance (ANOVA), that is gaining a better understanding of the impact of some factors on the two performance measures investigated. An effective analysis of the results obtained cannot prescind the consideration of the complex nature of the WIP-turn and its correlation with the system capacity, which is also another factor investigated in the experimental plan. Moreover, the prediction models for the average values of the performance measures are not regression models generated by fitting the experimental results according to the standard DOE approach as, in this case, the correlation between the factors and the twofold nature of the WIP-turn variation strategies would have invalidated the results. The prediction model development is based on the evidence of the factors relevance given by the DOE results.

The line configuration was changed with respect to the number of parallel machines operating in a station. Two sets of sampling intervals were chosen for the two monitored products, A & B. The low level for the sampling intervals indicate that the products are sampled more frequently; as for the initial level the sampling intervals of the two products were different they were also kept different for the low level; specifically, sampling intervals of 2 and 4 were set for products A and B, respectively.

#### 4 PRELIMINARY ANALYSIS

The number of consecutive unsampled items and the time between consecutive samples were calculated from the data obtained from the simulation runs. Different combinations of variables, related to the factors investigated, were considered so that a straightforward and meaningful mathematical formulation of the relationship between the performance measures and the variables could be obtained. The number of samples per unit time and the number of machines in a station proved the best combination for the average number of unsampled items (Figure 2).

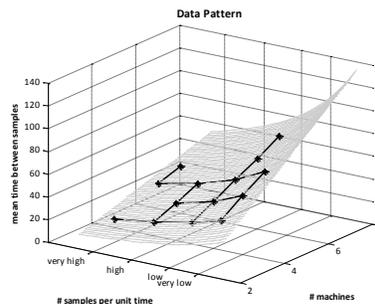


Figure 2: Mean time between samples vs. number of machines and number of samples per unit time.

A critical analysis of the factorial plots supported the identification of the factors impacting most on both the measures so that mathematical formulae could be derived. In this case, the ANOVA did not prove useful in providing further information on the nature of the relationships as the variability of the responses within the same factor level proved very low and the Mean Square Error was approximated to zero.

The number of samples per unit time,  $\# samples_{unit\ time}$ , is related to the sampling interval (number of unsampled items plus the sample at the sampling station machines),  $f_i$ , and the inter-arrival time (*int. arr. time<sub>i</sub>*) of the monitored product types (*i*)

$$\# samples_{unit\ time} = \sum_i \left( \frac{1}{int.\ arr.\ time_i} \times \frac{1}{f_i} \right)$$

In Figure 2 the starred points represent the average values obtained for scenarios characterized by the same variable combination. The black lines connect points that present the same *x* coordinate (number of machines) or the same *y* coordinate (number of samples per unit time), hence, they run parallel to the vertical Cartesian planes. In order to clarify the data pattern, an interpolated surface has been represented in Figure 2. The projection of the black lines onto the vertical planes clarifies the nature of the relationship between the average time between samples and the two variables. The patterns found can be intuitively justified. The average time between samples,  $\Delta T_{samples}$  (Equation (1)), is proportional to the number of machines in a station, *m*, as there exists a linear relationship between the inter-arrival time at a machine level and the inter-arrival time at a station level. For obvious reasons, the average time between samples is also inversely proportional to the number of samples per unit time; this is because the inverse of the number of samples per unit time is multiplied by the number of machines in a station, the delay caused by the spread of the overall flow across the machines of the station is taken into account and a measure of how often a sample is processed at a machine level is obtained

$$\Delta T_{samples} = \frac{m}{\# samples_{unit\ time}} \tag{1}$$

In a scenario where only monitored products are processed (i.e. no cross-flow), the number of samples per unit time does not vary between stations. Hence, the mean time between samples from a particular machine depends only on the number of parallel machines in the station. As a result, the line speed itself has no impact on the rate at which the quality issues are detected. This constitutes a quite counter-intuitive result as management had expected a faster flow due to increased system capacity would result in earlier detection. On the contrary, Equation (1) suggests that the introduction of extra machines in a station delays the quality information for that station and hence increases the quality risk. Moreover, the presence of an extra machine in a station increases the number of sources of quality issues. More compact formulations for the average time between samples can be obtained by expressing the number of samples per unit time with respect to its parameters

The average number of consecutive unsampled items can be derived by dividing the average time between samples by the global inter-arrival time at a machine level. The global inter-arrival time expresses the inter-arrival time between items at each machine. A basic formulation for the average number of consecutive unsampled items is

$$\# items_{between\ samples} = \frac{1}{\left[ \sum_i \left( \frac{1}{int.\ arr.\ time_i} * \frac{1}{f_i} \right) \right] * int.\ arr.\ time_{global}} - 1 \tag{2}$$

where the sub-index *i* refers to each monitored product and  $f_i$  represents the sampling interval of the *i*th product. The unit decrement in Equation (2) represents the sampled item. The presence of the global inter-arrival time (*int. arr. time<sub>global</sub>*) in Equation (2) indicates that the average number of consecutive unsam-

pled items is affected by the presence of any unmonitored product processed by the station. The unmonitored flow acts as a scaling factor on the average number of consecutive unsampled items as it modifies the global inter-arrival time and, as a consequence, the sampled fraction at a station. Equation (2) also suggests that the average number of consecutive unsampled items is not affected by the number of machines operating in a station.

Although the mean values presented in this section are quite easy to assess and provide long term evaluation of the sampling plan monitoring efficacy, they give no specific guarantee that a machine will be sampled over a specific period of time. Hence, it is only the distribution of these metrics that allows quantification of the sampling risk by selecting a particular confidence interval to provide a measure of the maximum number of items at risk.

## 5 PREDICTION MODEL DEVELOPMENT

Although the formulae in Equation (1) and (2) are valid for the mean values of both sampling and non-sampling stations, further analyses showed that the shape of the distributions differ according to the nature of the station. For the non-sampling stations, the distribution tends to follow an exponentially decreasing pattern. Whereas for the sampling station, the distribution shape appears less regular with a peak in the region of the mean value (Figure 3). The distributions reported in Figure 3 refer to a production scenario characterized by one monitored product type and a serial unmonitored flow.

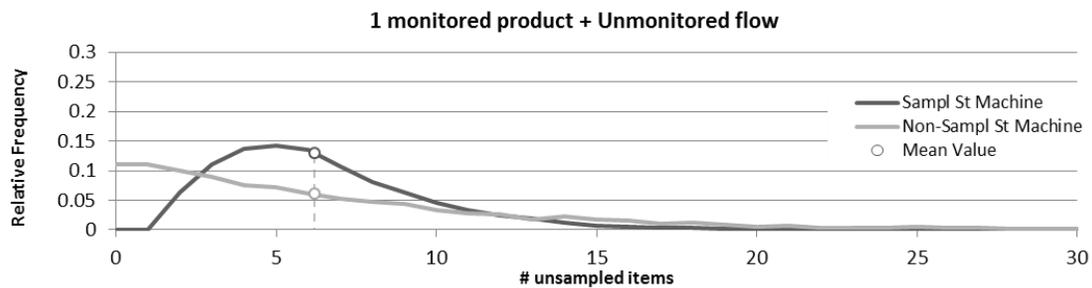


Figure 3: Comparison between the distribution of the number of consecutive unsampled items at a sampling station machine and a non-sampling station machine.

The presence of an unmonitored flow merely modifies the scale of the horizontal axis, and does not affect the shape of the distribution in a non-sampling station. On the contrary, the shape of the distribution of the number of unsampled items in a sampling station machine is strongly dependent on the composition of the production flow.

In the non-sampling stations, samples are not deliberately chosen and the sequence disorder and multiple stream effects are responsible for determining the loss of deterministic sampling patterns. This is particularly evident when only monitored product is modeled and no unmonitored flow exists. In this extreme example, shown in Figure 4, the distribution of the number of consecutive unsampled items is derived for the same scenario as in Figure 3 except that the unmonitored flow is ignored. In this case, for the machines of the sampling station the distribution of the number of unsampled items degenerates into a deterministic value, that is the sampling interval set for that product minus one. On the contrary, in the non-sampling stations, the variation of the order with which items move along the segment, together with the lack of deterministic routing policies, subverts the regularity of the sampling plan and the distribution presents an exponential shape even for this relatively simple scenario (Figure 4). Indeed, previous analyses showed that the number of consecutive unsampled items at the non-sampling station machines follows a geometric distribution with proportion equal to the inverse of the average number of consecutive unsampled items (Equation (2)) plus one (Rotondo, Young, and Geraghty 2012). The geometric model proves robust not only to variations of the product mix but also to variations of the location and quantity

of the quality information sources (e.g. non-sampling station common to two crossing segments independently monitored) and the randomness of the serial flow (e.g. some non-sampling stations can be skipped by the serial flow).

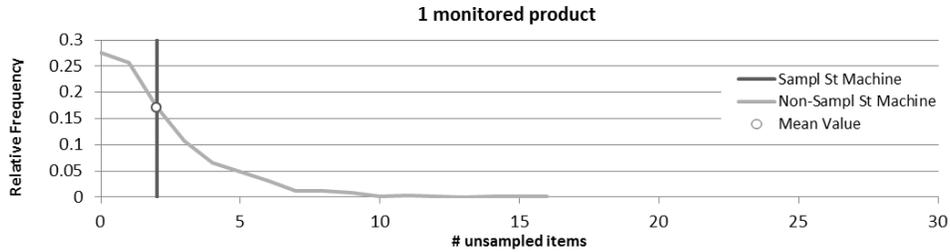


Figure 4: Distribution of the number of consecutive unsampled items at a non-sampling station machine in presence of one monitored product type.

When the sampling station machines are analyzed, the randomness of the number of consecutive unsampled items is caused by merging flows characterized by random departure. A fundamental proof of this effect is given by the production scenario investigated here. The introduction of an unmonitored product flow when one product type is monitored, has the interesting effect of turning the degenerative distribution into an unlimited domain distribution (Figure 3). This happens because, due to the randomness of the departure process, there is no theoretical limit to the number of unmonitored items that can be processed between two consecutive monitored items.

There are fundamental theoretical elements which support the analysis of the distribution of the number of consecutive unsampled items at the sampling station machines for this particular production scenario. Firstly, the sequence of monitored and unmonitored items processed at a machine is independent. The lack of dispatching rules for regulating the sequence of product types to be processed by any machine in the sampling station contributes to the random item sequence. This is also supported by the fact that the inter-arrival time distributions for the different product types are exponential and mutually independent. The memory-less property of the exponential distribution suggests that the sequence in which items arrive at a station and are processed is actually random. Secondly, the probability that an item processed by a machine at a given time is either monitored or unmonitored is constant over time as the inter-arrival process does not change; this means that if the items arriving at a station were organized into two separate categories, the arrival of an item belonging to either of the categories could be considered as the observation of a Bernoulli event. Finally, there is a deterministic number of monitored items to be processed before a sample is chosen. These three elements suggest that the number of consecutive unsampled items at a sampling station machine follows a negative binomial distribution. Indeed, a negative binomial distribution generally describes the number of failures before the  $r^{\text{th}}$  success in a sequence of independent Bernoulli trials with probability  $p$  of success. In the case analyzed here, a failure consists of processing an unmonitored item, a success would be processing a monitored item and, in particular, the  $r^{\text{th}}$  success coincides with processing an item chosen as a sample. The monitored volume fraction, which is expressed by the following equation, represents the probability of success:

$$\text{Monitored volume fraction} = \sum_i \frac{\text{int. arr. time}_{\text{global}}}{\text{int. arr. time}_i}$$

where  $i$  refers to the monitored products. The other parameter of the distribution,  $r$ , is given by the sampling interval set for the monitored product.

The negative binomial model validation was performed by comparing the model predictions against the simulation results for various production scenarios. For the purpose of the analysis illustrated in this section, the presence of product B was omitted in the simulation model. Combinations of several sampling intervals and monitored volume fractions were considered. Very high accuracy was obtained for all

the scenarios simulated (Figure 5). The average absolute error, calculated as the arithmetic average of the absolute difference between the predicted frequency and the frequency obtained from the simulation results, proved always lower than 0.4%. The Pearson's chi square test was also applied and confirmed the high goodness-of-fit of the negative binomial model for all the scenarios investigated.

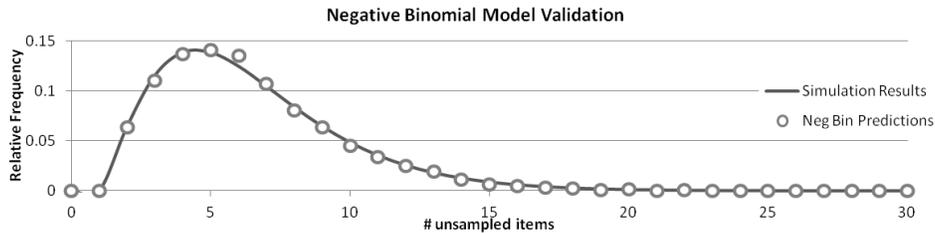


Figure 5: Validation of the negative binomial model against simulation results.

## 6 CONCLUSIONS

Being motivated by a real industrial case observed in the semiconductor industry, the analysis presented here has focused on the monitoring frequency of items rather than machines. The main contribution of this work consists in the demonstration that a simulation study allows the generation of much simpler way of calculating the impact of changes of a system settings. This was realized through the development of sampling related - decision support tools that do not require considerable amount of data and provide quality management with immediate and practical results.

The randomness of the departure process for the different products introduces complexities which make the assessment of the efficacy of a sampling policy complex. In particular, the distribution of the key indicator for the effectiveness has been shown to be more important than the average values. The development of the appropriate distribution could only be achieved through the use of simulation modeling. The randomness of the number of consecutive unsampled items is closely related to the merging of independently random production flows and follows a negative binomial distribution. While the model developed here is specifically valid for the production scenario considered during the analysis, its integration with other prediction models will expand its applicability to more complex production flow scenarios.

The prediction model developed can be used as an effective tool to facilitate decision makers during the analyses of sampling strategy related issues. Indeed, it can support the quantification of the quality risk associated with a sampling strategy in terms of the maximum number of unsampled items between consecutive samples. This quality risk measure can be calculated as the percentile of the distribution corresponding with a cumulative probability equal to the risk level set as a reference. With an inverse and more laborious procedure, the model can also support the development of a sampling strategy able to satisfy quality risk conditions at predetermined risk confidence levels.

## ACKNOWLEDGMENTS

The research leading to these results has been funded under the National Development Plan 2008-2013 by the Irish Research Council for Science Engineering and Technology and carried out in collaboration with the Enterprise Ireland & IDA sponsored Technology Centre, The Irish Centre for Manufacturing Research.

## REFERENCES

Carson, J.S. 2004. "Introduction to Modeling and Simulation." In *Proceedings of 2004 Winter Simulation Conference*, edited by J. Smith, B. Peters, R.G. Ingalls, M.D. Rossetti, 9-16. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

- Colledani, M., and T. Tolio. 2011. "Integrated Analysis of Quality and Production Logistics Performance in Manufacturing Lines." *International Journal of Production Research* 49 (2): 485-519.
- Fan, C. M., R. S. Guo, S. C. Chang, and C. S. Wei. 2000. "SHEWMA: an End-of-Line SPC Scheme Using Wafer Acceptance Test Data." *IEEE Transactions on Semiconductor Manufacturing* 13(3): 344-58.
- Fan, C. M., R. S. Guo, A. Chen, K. C. Hsu, and C. S. Wei. 2001. "Data Mining and Fault Diagnosis Based on Wafer Acceptance Test Data and In-Line Manufacturing Data." In *Proceedings of the 2001 IEEE International Semiconductor Manufacturing Symposium*, 171-174.
- Freeman, J. M. 2008. "The Case for Quality Costing Simulation." *The TQM Journal* 20(5): 476-87.
- Gershwin, S. B., and I. C. Schick. 2007. "A Taxonomy of Quality/Quantity Issues in Manufacturing Systems." In *Proceedings of the Sixth Conference on the Analysis of Manufacturing Systems*. Lunteren, the Netherlands.
- Guo, R. S., C. K. Tsai, J. H. Lee, and S. C. Chang. 1996. "Intelligent Process Diagnosis Based on End-of-Line Electrical Test Data." In *Proceeding of the 1996 Electronics Manufacturing Technology Symposium*, 347-354.
- Inman, R. R., D. E. Blumenfeld, N. Huang, and J. Li. 2003. "Designing Production Systems for Quality: Research Opportunities from an Automotive Industry Perspective." *International Journal of Production Research* 41 (9): 1953-1971.
- Ioannidis, S. 2013. "Joint Production and Quality Control in Production Systems with Two Customer Classes and Lost Sales." *IIE Transactions* 45(6): 605-616.
- Jiang, P., F. Jia, Y. Wang, and M. Zheng. 2012. "Real-Time Quality Monitoring and Predicting Model Based on Error Propagation Networks for Multistage Machining Processes." *Journal of Intelligent Manufacturing* (iFirst October 2012): 1-18. DOI 10.1007/s10845-012-0703-0.
- Jin, M., Y. Li, and F. Tsung. 2010. "Chart Allocation Strategy for Serial-Parallel Multistage Manufacturing Processes." *IIE Transactions* 42 (8): 577-588.
- Kaya, M., and Ö. Özer. 2009. "Quality Risk in Outsourcing: Non-Contractible Product Quality and Private Quality Cost Information." *Naval Research Logistics* 56 (7): 669-685.
- Kim, J., and S. B. Gershwin. 2008. "Analysis of Long Flow Lines with Quality and Operational Failures." *IIE Transactions* 40 (3): 284-296.
- Li, Y. T., and H. Wang. 2011. "Chart Allocation Strategy for Serial-Parallel Multistage Manufacturing Processes with Multiple Faults." *Journal of the Chinese Institute of Industrial Engineers* 28 (7): 493-503.
- Liu, X., R. J. MacKay, and S. Steiner. 2008. "Monitoring Multiple Stream Processes." *Quality Engineering* 20 (3): 296-308.
- Montgomery, D. C. 2008. *Introduction to Statistical Process Control*. 5th ed. John Wiley & Sons, Inc.
- Nduhura Munga, J., S. Dauzère-Pérès, P. Vialletelle, and C. Yugma. 2012. "Industrial Implementation of a Dynamic Sampling Algorithm in Semiconductor Manufacturing: Approach And Challenges." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspace, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 2151-9. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Owen, J., and D. Blumenfeld. 2008. "Effects of Operating Speed on Production Quality and Throughput." *International Journal of Production Research* 46 (24): 7039-7056.
- Rodriguez-Verjan, G. L., S. Dauzère-Pérès, and J. Pinaton. 2012. "A mathematical model for estimating defect inspection capacity with a dynamic control strategy." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspace, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 2092-2100. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Rodriguez-Verjan, G. L., S. Dauzère-Pérès, and J. Pinaton. 2011. "Impact of Control Plan Design on Tool Risk Management: a Simulation Study in Semiconductor Manufacturing." In *Proceedings of the 2011*

- Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 1918–1925. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Rodriguez-Verjan, G. L., E. Tartiere, J. Pinaton, S. Dauzere-Peres, and A. Thieullen. 2012. "Dispatching of Lots to Dynamically Reduce the Wafers at Risk in Semiconductor Manufacturing." In *Proceedings of the 2012 IEEE International Conference on Automation Science and Engineering (CASE)*, 920-923.
- Rotondo, A., P. Young, and J. Geraghty. 2012. "Quality Risk Prediction at a Non-Sampling Station Machine in a Multi-Product, Multi-Stage, Parallel Processing Manufacturing System Subjected to Sequence Disorder and Multiple Stream Effects." *Annals of Operations Research* (iFirst May 2012): 1-23. DOI 10.1007/s10479-012-1145-y.
- Sargent, R. G. 2010. "Verification and Validation of Simulation Models." In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hagan and E. Yücesan, 130-143. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Tirkel, I., N. Reshef, and G. Rabinowitz. 2009. "In-Line Inspection Impact on Cycle Time and Yield." *IEEE Transactions on Semiconductor Manufacturing* 22 (4): 491-498.
- Vaghefi, A., and V. Sarhangian. 2009. "Contribution of Simulation to the Optimization of Inspection Plans for Multi-Stage Manufacturing Systems." *Computers & Industrial Engineering* 57: 1226-1234.
- Van Volsem, S., W. Dullaert, and H. Van Landeghem. 2007. "An Evolutionary Algorithm and Discrete Event Simulation for Optimizing Inspection Strategies for Multi-Stage Processes." *European Journal of Operational Research* 179 (3): 621-633.
- Wang, J., J. Li, J. Arinez, and S. Biller. 2013. "Quality Bottleneck Transitions in Flexible Manufacturing Systems with Batch Productions." *IIE Transactions* 45 (2): 190-205.
- Zeng, L., and S. Zhou. 2007. "Inferring the Interactions in Complex Manufacturing Processes Using Graphical Models." *Technometrics* 49 (4): 373-381.

## AUTHOR BIOGRAPHIES

**ANNA ROTONDO** is a postdoctoral research in the School of Mechanical and Manufacturing Engineering at Dublin City University. She holds a MS degree in management engineering from Politecnico di Bari (Italy) and received her PhD in manufacturing engineering from Dublin City University, Ireland. Her research interests are in modeling, analysis and optimization of complex manufacturing systems, with particular application to semiconductor manufacturing. Her e-mail address is [anna.rotondo@dcu.ie](mailto:anna.rotondo@dcu.ie).

**JOHN GERAGHTY** is a lecturer in the School of Mechanical and Manufacturing Engineering at Dublin City University, Ireland. He holds a B.Eng., a M.Eng. (by research) and a PhD in Industrial Engineering from the University of Limerick, Ireland. The focus of his research for both his M.Eng. and PhD theses was on inventory and production control strategies in manufacturing systems subject to variability in cycle time and demand distributions. He is a founding member of the Enterprise Process Research Centre at DCU. His e-mail address is [john.geraghty@dcu.ie](mailto:john.geraghty@dcu.ie).

**PAUL YOUNG** is the director of the Enterprise Process Research Centre at Dublin City University, Ireland. His PhD (Trinity College Dublin, 1991) concerned condition monitoring of turning. The design, analysis, modeling and monitoring of manufacturing processes was the main focus of his Post-Doctoral research work with the Advanced Manufacturing Research Centre in University College Dublin. In 1998 he was appointed to the faculty in Dublin City University where, since 2001 his research is mainly in the application of modeling and analysis to improve the performance of complex manufacturing systems, while maintaining some machine design and modeling. His e-mail address is [paul.young@dcu.ie](mailto:paul.young@dcu.ie).