Data Collection and Statistical Data Analysis in Preparation for Simulation of a Furniture Manufacturing Company

Wilson R. Nyemba, and Charles Mbohwa

***Abstract* — Systems that are simulated are generally made up of one or more elements that have some uncertainty associated with them. These systems may eventually evolve in a manner that is not completely predictable and thus become stochastic in nature. Simulation of stochastic systems requires that the variability of the elements in the system be characterized by probability distributions or concepts. An 'As-Is-Analysis' of the plant layout and product process flows was carried out at a furniture manufacturing company. Process data for four of their main products namely, pallets, baby tenders bunk beds and standard coffins was collected using a specially designed data sheet. An analysis of the product flow times was carried out by grouping the data into four time variables namely, material movement, processing and waiting (idle) times before and after processing at the active workstation. The distributions of these variables were obtained using graphical methods in which smooth distribution curves were generated. The gamma distribution with shape parameters of α= 3 and α=2 characterized the baby tenders while the product flow times for pallets were characterized by both the gamma distribution, α=3 and the exponential distribution with the mean varying between 14.95 and 271.78 seconds. All the data and analysis carried out produced useful information for input to the design of experiments for simulation.**

***Index Terms* — Data Collection, Manufacturing, Materials Flow, Statistical Analysis**

I. Introduction

Most industrial operations are generally not predictable in nature and in many cases decisions concerning these operations have to be made. Nevertheless these operations are composed of one or more elements that have uncertainty associated with them and are referred to as random or stochastic elements [1], [2]. The simulation of stochastic elements such as product processing time requires probability concepts to characterize them. The outputs from a simulation model are also probabilistic, and therefore statistical interpretations about them are usually required. An experiment is a well-defined procedure or process whose outcome is observable but is not known with certainty in advance, with sample space containing the set of all possible outcomes. If the sample space is finite or countably infinite, it is considered to be discrete otherwise it is continuous [3].

The research was carried out at a furniture manufacturing company in the Workington industrial area of Harare in Zimbabwe. The holding company consists of two main subsidiaries that specialize in the manufacture of a wide range of wooden products ranging from general furniture to wooden sheds, sharing machinery under the same complex but producing complementary products. Following the global recession that affected most countries around the world in 2008 [4], the company embarked on a scheme to revamp production and increase their market share owing to competition on the local market. Because of insufficient documentation on their manufacturing systems and production scheduling, worsened by the wide range of products and demands from customers, they prioritized the task to collect data on production for analysis to enable them to plan and schedule production to meet customer demands. The data collected would also be useful as input to simulation experiments to enable the company to predict performance, not only in the supply of their furniture but also on the performance of their resources. A team of engineering academics from the University of Zimbabwe set out to assist in this task, collecting the data over a period of a year, analyzing it through statistical analysis and then advise the company on the best possible mixes [5]. This was done as part of research at minimal or no cost to the company but with the ultimate objective of advising the company on the appropriate products to make at any given time. The research was based mainly in the plant with the team focusing on machine utilization, available skills, processing times and materials movement for the selected products, i.e. pallets, baby tenders, bunk beds and standard coffins as shown in the Appendix.

Statistical data analysis was carried out so that statistical inferences can be made on various manufacturing operations being carried out at this manufacturing company such that complex problems may be identified, studied and solved. The aim of the research was to determine the probability distributions for various product workflows. Organizing and grouping the collected data into various

Manuscript received March 07, 2016; revised March 27, 2016.

Wilson R. Nyemba is a Lecturer in Engineering Design and Manufacture, currently on sabbatical from the University of Zimbabwe and pursuing doctoral studies in Modelling, Process Flow and Simulation of Engineering Design and Manufacture at the University of Johannesburg, Department of Mechanical Engineering Science, Faculty of Engineering and the Built Environment, Auckland Park 2006, Johannesburg, South Africa. (Phone: +263 772 345 441, e-mail: nyemba@yahoo.com)

Charles Mbohwa is a Full Professor of Sustainability Engineering and Engineering Management at the University of Johannesburg, Auckland Park Bunting Road Campus, P. O. Box 524, Auckland Park 2006, Department of Quality and Operations Management, Johannesburg, South Africa. (e-mail: cmbohwa@uj.ac.za)

probability distributions in preparation for the design of experiments and planning for simulation was accomplished.

II. Data Collection and Statistical Analysis Review

Discrete Event Simulation (DES) models can represent stochastic behavior and this is one of its powerful capabilities, certainly by comparison with other techniques. In order to make use of this capability, the modeler must be able to specify the distribution associated with each stochastic process being represented in the model. This is one of the most difficult things for a manufacturing company to achieve, as it requires effort in gathering data and both skill and experience in preparing this data for simulation.

Statistical data analysis is a technique that involves the use of a set of methods to collect, analyse, present, and interpret data [3]. Statistical methods are used in a wide variety of industrial operations and help to identify, study, and solve many complex problems. In the business and economic world, these methods enable decision makers and managers to make informed and better decisions about uncertain situations. Statistical data analysis provides hands on experience to promote the use of statistical thinking and techniques to apply in order to make educated decisions in the business world. Studying a problem through the use of statistical data analysis usually involves four basic steps [3]:

- Defining the problem
- Collecting the data
- Analyzing the data and
- Reporting the results

A. Defining the Problem

An exact definition of the problem is imperative in order to obtain accurate data about it. It is extremely difficult to gather data without a clear definition of the problem. In the case study company, management realized that there was a good opportunity to expand the scope and product mix as there was constant demand from the customers. However, because of insufficient documentation and lack of production planning and scheduling, they were not able to meet customer demands. As such, they needed a system where the right products in the right quantities are produced at the right time. In addition, it was also necessary to collect data to be able to predict performance from the process flows.

B. Collecting the Data

An emphasis must be put on the importance of defining the population about which inferences are being sought to be made, all the requirements of sampling and experimental design must be met. Designing ways to collect data is an important task in statistical data analysis, paying particular attention to [1]:

- Population - a set of all the elements of interest in a study,
- Sample - a subset of the population.

Statistical inference refers to extending knowledge of information from a random sample of a population to the whole population [1], also sometimes referred to in mathematics as an ‘Inductive Reasoning’ that is, knowledge of whole from a particular [3]. Its main application is in hypotheses testing about a given population. The purpose of statistical inference is to obtain information about a population contained in a sample [6]. It is often difficult to test the entire population, so a sample is the only realistic way to obtain data because of time and cost constraints.

Data can be either quantitative or qualitative. Qualitative data are labels or names used to identify an attribute of each element. Quantitative data are always numeric and indicate either how much or how many. Data can also be collected from existing sources or obtained through observation and experimental studies designed to obtain new data [7]. In an experimental study, the variable of interest is identified, then one or more factors in the study are controlled so that data can be obtained about how the factors influence the variables [11]. In observational studies, no attempt is made to control or influence the variables of interest. A survey is perhaps the most common type of observational study [12].

C. Analyzing the Data

Statistical data analysis divides the methods for analyzing data into two categories: exploratory methods and confirmatory methods [3]. Exploratory methods are used to discover what the data seems to be saying by using simple arithmetic and easy-to-draw pictures to summarize the data. Confirmatory methods use ideas from probability theory in an attempt to answer specific questions. Probability is important in decision making because it provides a mechanism for measuring, expressing, and analyzing the uncertainties associated with future events [10].

D. Reporting the Results

Through inferences, an estimate distribution or test claims about the characteristics of a population can be obtained from a sample. The results may be reported in the form of a table, graph or set of percentages. Sometimes only a small collection (sample) will be available and would have been examined and not an entire population, the reported results must reflect the uncertainty through the use of probability statements and intervals of values.

E. Probability Distributions

Continuous distributions have the following 3 parameters [3]:

- Location γ
- Scale β
- Shape α

Location describes the center point of some distributions or lower endpoint for others. The scale parameter compresses or expands the distribution of *x* values, making the distribution more or less concentrated. The shape parameter changes the fundamental shape of the distribution [3]. The exponential and normal distributions do not have a shape parameter but the gamma does as shown in Fig. 1.

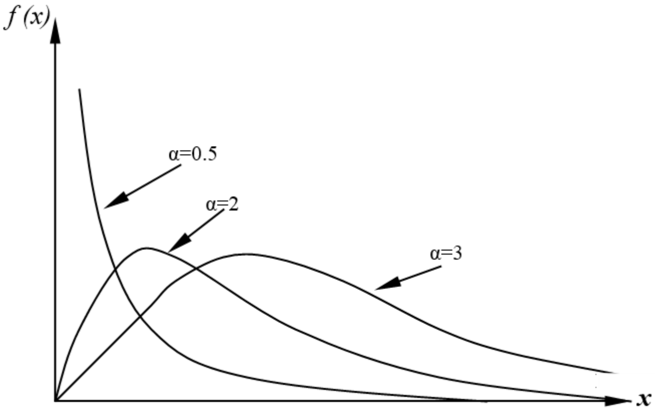

Fig. 1. Gamma Distributions with Different Shape Parameters

Gamma

This is a basic distribution of statistics for variables bounded at one side. This gives distribution of time required for exactly *k* independent events to occur, assuming the events take place at a constant rate and is often used in queuing theory, reliability, and other industrial applications. Example: Time between inventory restocking, time to failure for a system with standby components and time to complete some task [8].

Beta

This is a basic distribution of statistics for variables bounded at both sides. This is useful for both theoretical and applied problems in many areas. Uniform, right triangular, and parabolic distributions are special cases. Examples: distribution of daily percent yield in a manufacturing process; description of elapsed times to task completion [8]. Fig. 2 shows Beta distributions with different shape parameters.

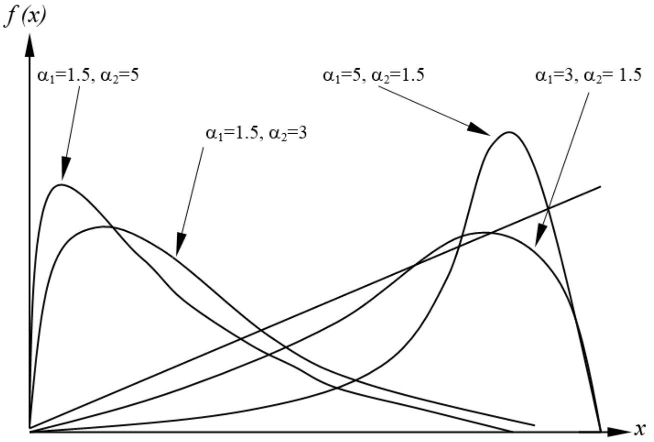

Fig. 2. Beta Distributions with Different Shape Parameters

III. Methodology and Detailed Case Study

Data collection is an important part of statistical data analysis. Useful statistical inferences can only be made from data collected properly. To come up with a suitable methodology for collecting data required a thorough understanding of the manufacturing processes being carried out at the furniture manufacturing company. After gathering the knowledge of manufacturing processes being carried out during the work study, a suitable data collection tool was developed that adequately met the objectives of the research. The main objective in this case was to collect product flow times. Following the detailed work study, sheets for collecting the data were designed and the process flows and times were recorded over the duration of the research.

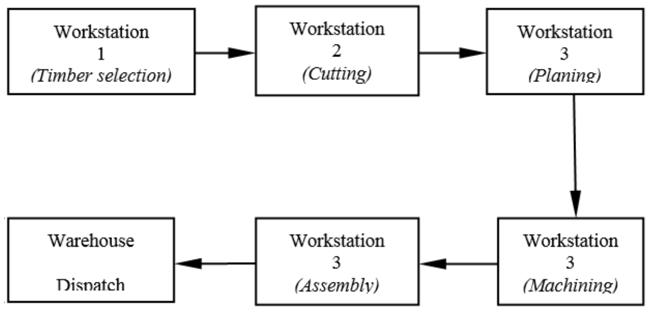

Fig. 3. General Process Flow at the Company

Fig. 3 shows a general production flow for the company with the selection of timber in the timber-yard (Workstation 1) followed by transportation by a forklift for cutting these to the required sizes depending on the product to be made, then planing and machining or carving the parts to the required designs. Material movement at this stage is usually by hand since these will be mainly small parts. The assembly section also included spray painting for some of the products that require this and finally to the Warehouse for dispatch. It was noted in general that at any given workstation the part arrives and waits to be processed before it is offloaded and waits for transportation to the next workstation. As such, in designing the data recording sheets the following was used and taken into account:

- a_t Material movement time from previous workstation to active workstation.
- b_t Waiting time before being processed at the active workstation.
- c_t Product processing time at the active workstation.
- d_t Waiting time after processing before moving to the next workstation

The instrument for time collection was designed in such a way that made it easy to collect time in the factory taking into account the 3 events of the product arriving from previous workstation, processing and leaving for the next workstation. The instrument for data collection was also designed such that the absolute time at which these events occur can easily be recorded. The differences between these times would then be calculated to give material movement times, processing times, waiting before and after processing times (idle time).

After selecting a product, the first thing was to establish its product assembly tree, consisting of all the respective number of components that make up a complete product. For example, pallets are composed of two components namely 2 legs and 3 endurite moldings. Its product assembly tree is illustrated in Fig. 4 showing the number of components that make up a single finished pallet. After establishing product assembly trees, the next task was to collect the time of producing a product whose assembly tree would have been established.

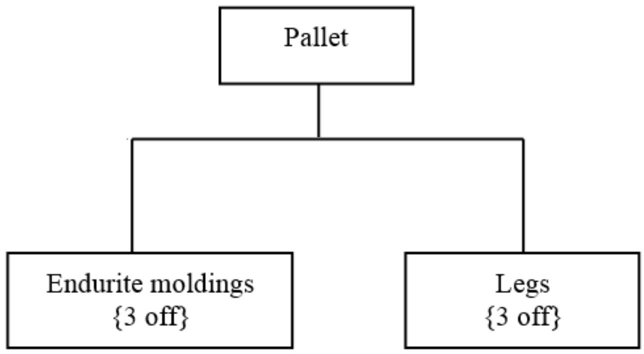

Fig. 4. Product Assembly Tree for a Pallet

IV. Data Analysis And Results

Graphical methods were used to analyse the data whereby frequency curves of various time flows were plotted to give different distribution shapes. The method used to obtain the distributions is the same with the one used for obtaining histograms but instead of histograms smooth curves were generated. These were preferred to histograms in showing the distributions because smooth curves are easier to visualize the shape of the distribution.

A. Skewness

The data was analyzed by checking the presence or lack of skewness. The method used to obtain the distributions (beta, exponential and so on) was to identify a family of distributions that closely matched the data being analyzed. Thus skewness was useful in determining the distributions of generated smooth curves. For example, lack of skewness implied a normal distribution. Some distributions were ruled out on the basis that they were skewed in a completely different direction to that of the generated smooth curves [9].

B. Frequency Distributions

Construction of frequency distributions was done using three basic steps:

- determining the class intervals
- entering the values into class intervals
- plotting the frequency distribution

Determining the Class Intervals

The class intervals were obtained by first identifying the maximum and minimum values in the data set. The difference between the two was then calculated to give the range which was then divided by an appropriate number to obtain a convenient class interval width.

Entering the Values into Class Intervals

After obtaining the class interval width, the limits of all class intervals were recorded. Intervals of equal width were used. The values of data were then placed in their respective class intervals.

Plotting Smooth Curves

The number of entries in each class interval were counted and recorded. This number of entries equal the frequency of the class interval. A smooth curve of frequency versus a class interval was finally drawn.

Using the technique described above, the analysis of data for pallets and baby tenders was carried out, for each and every component that makes up the final product. As a sample, the following figures 5 – 8 are a series of time distribution plots and curves for endurite moldings at Workstation 2 for each of the times as described earlier.

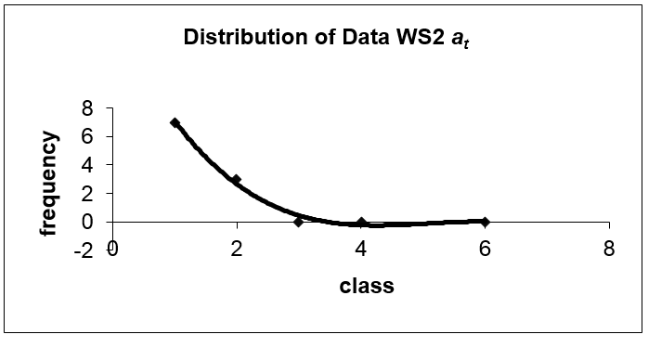

Fig. 5. Distribution of pallets' endurite moldings flow times at Workstation 2 (WS2): Approximated to exponential with a mean of 23.03 seconds

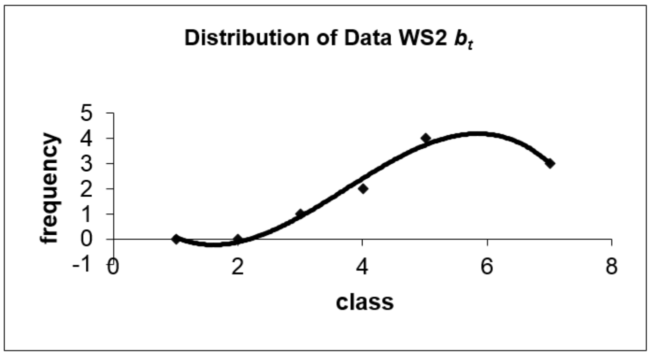

Fig. 6. Distribution of pallets' endurite moldings flow times at Workstation 2 (WS2): Negatively skewed. Approximated to a beta distribution with α_1=3, α_2 = 1.5

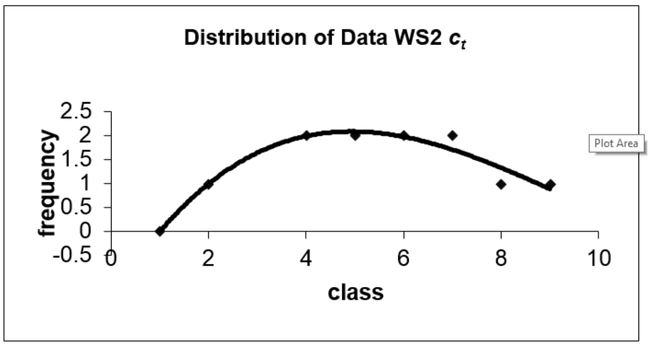

Fig. 7. Distribution of pallets' endurite moldings flow times at Workstation 2 (WS2): Positively skewed. Approximated to a gamma distribution with α=3

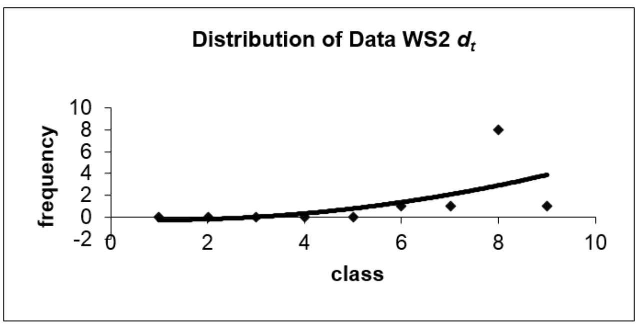

Fig. 8. Distribution of pallets' endurite moldings flow times at Workstation 2 (WS2): Negatively skewed. Approximated to a beta distribution with α_1=2, α_2 = 0.8

C. Utilization of Time across Workstations

Determining time distributions across 9 workstations in the production flow helped in planning and work scheduling. The time distribution revealed which workstations time is spent most or least. Fig 9 shows the time distributions during the production of short rails for bunk beds with the corresponding workstations in Table I. From this figure, it can be deduced that production following the sequence 1-2-3-4-5-6-7-8-9 started with a slower pace. It can be planned that production is scheduled such that workstation 8 is not loaded with much work as this will slow down production in the initial stages and hence significantly increase the time waiting before processing on this machine. This can be done by looking for any tasks that could be carried out on the products that do not require to necessarily pass through Workstation 8 first. These tasks could be done on the other workstations while a few other products are being processed at Workstation 8.

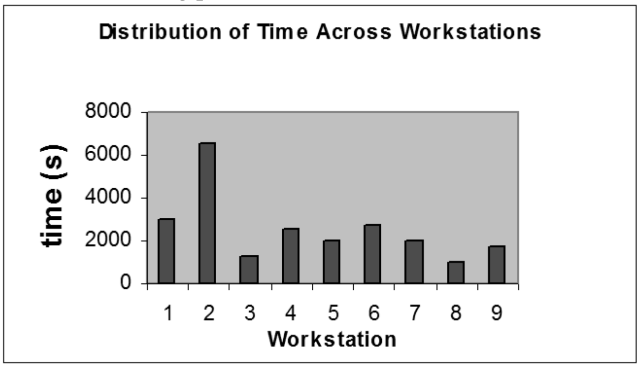

Fig. 9. Distribution of Time for Short Rails Production across Workstations

TABLE I
DESCRIPTION OF WORKSTATIONS

Workstation	Name
1	Timber yard {1}
2	4-cutter {6}
3	Cross-cut {12}
4	Tenoner {14}
5	Spindle moulder {13}
6	Dimension saw {13}
7	Belt sander {46}
8	Drum sander {35}
9	Joinery

The utilization of time during the manufacture of products followed almost the same pattern with material movement taking the least time. The time utilization for the production of short rails for bunk beds is shown in Fig. 10 to illustrate how time was shared among the four variables (a_t, b_t, c_t and d_t). Thus, processing time constituted 18% of production time and waiting before processing took 14% of the production time. In the above case, waiting after processing took most of the production time (66%). For bunk beds, the time spent on waiting before processing constituted 18% of the total production time. However for other products analyzed, this time is relatively high (about 33% for standard coffins). This pattern was representative of most products that the company manufactures.

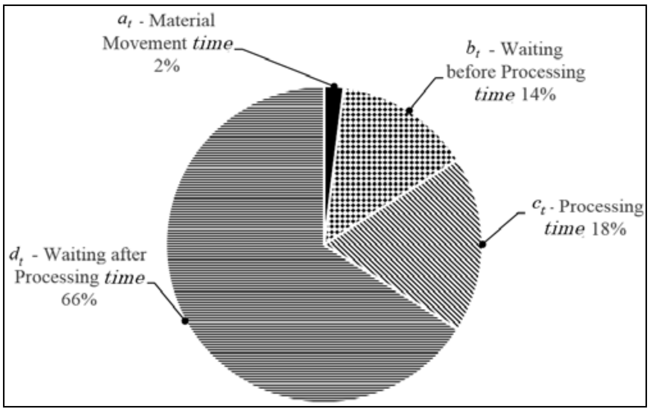

Fig. 10. Utilization of Time for Short Rails of Bunk Beds

V. DISCUSSION AND RECOMMENDATIONS

Product flow times are an important aspect of production. Monitoring these times will actually help in increasing the productivity of the company. Recommendations on how to improve productivity by altering the production system in use were made. Materials movement time is one variable studied in this research and all the flow times recorded showed that it took the least part of the production time. However, this time was greatly reduced when the machines were rearranged through optimization of the layout and process flows, such that backtracking was eliminated and all processes flowed in the same direction. Having clear gangways also greatly reduced material movement time. The distributions found to characterize this time variable were:

- Gamma (α=3 or 2) for baby tenders
- Gamma (α=3) and Exponential for pallets

It was recommended that in selecting input probability distributions for the simulation model for this variable, the gamma distribution (α=3) be selected for the baby tenders and the exponential distribution be selected for pallets.

Waiting (idle) time before processing was greatly reduced through proper work planning and scheduling following the reorganization of the plant. It was concluded that production of two or more products use the same machines, therefore there was a need for proper planning to decide on which product goes first through the machine shop. This reduced the amount of waiting time for others to be processed. Machine set up time was found to constitute only a small fraction of the time spent waiting for processing and in a significant number of cases this time was only a few seconds, which time results from a product waiting for the processing of other products to be completed or for the staff who are supposed to carry out the tasks but may be attending to other tasks. This time variable was found to be characterized by the gamma distribution (α=2 and α=3), beta distribution with {(α_1=3, α_2=1.5) and (α_1=1.5, α_2=3)} and exponential distributions at different workstations in the production of both pallets and baby tenders. It was recommended that for this time variable, a gamma distribution be selected as an input to the simulation model as it encompasses the other two distributions.

Processing time is another time parameter analyzed in this research and in most cases, it constituted the smallest fraction of the production time. There were only a few ways found feasible for reducing this time and among these ways are specialization and increasing the number of workers in time-consuming tasks such as hand sanding. Specialization also results in the workers being familiar and experienced with their daily tasks and be capable of performing these tasks in the shortest possible time. This time variable was also found to be characterized by the gamma distribution (α=3) for the production of both pallets and baby tenders and thus it was recommended to use this distribution as an input to the simulation model.

The time spent after processing on an active workstation was found in many cases to constitute the greatest part of production. This time resulted from waiting for the work on the next workstation to be completed. It also resulted from a product waiting to be carried by a forklift or worker who may be attending to a different task. Gamma and Beta distributions characterized this time variable for both pallets and baby tenders. It was recommended that a 'Goodness of Fit Test' be carried out on these two distributions before selecting a particular distribution as an input to the simulation model.

VI. Conclusion

Failure to adopt the most representative distribution can have a significant effect on the output from a simulation model. Therefore selecting the most appropriate distribution is important in building a valid simulation model. Consequently it was recommended that appropriate tests be carried out on the data to verify the validity of the selected distribution. Time utilization was greatly improved by careful planning and production scheduling following the optimization of the process flows and reorganization of the plant layout. However cost constraints may be a hindrance if actions like proper machine rearrangements are to be taken but these may have long term benefits.

The research focused on statistical data analysis of flow times of products that are being produced at the furniture manufacturing company. Data was collected for the company's four main and fast moving products namely pallets, bunk beds, baby tenders and standard coffins. For the other products, only a few readings were recorded mainly because of their limited production. However, these readings were sufficient to carry out the analysis of flow times as they were found to be consistent with the readings for the four main products. The flow times were grouped into four variables and various probability distributions of these variables were generated. Apart from informing management on the trends and what product mixes to pursue, these findings were also be used as input to a simulation model of the plant.

The probability distribution found to characterize the production of baby tenders was mostly the gamma distribution with a shape parameter of α=3. In a few cases the shape parameter was 2. The probability distributions found to characterize the production of pallets were beta with shape parameters (α_1=3, α_2=1.5) and in other cases (α_1=1.5, α_2=3). The exponential distribution with various means ranging from 14.95 seconds to 271.78 seconds was also found to characterize the production flows for pallets.

Appendix: Case Study Products

References

[1] Bloomfield, R., Mazhari, E., Hawkins, J. and Son Y.J., (2012), "Interoperability of manufacturing applications using the Core Manufacturing Simulation Data (CMSD) standard information model", *Computers & Industrial Engineering*, 62(2012), 1065–1079, Elsevier.

[2] Negahban, A. and Smith, J. S., (2014), "Simulation for manufacturing system design and operation: Literature review and analysis", *Journal of Manufacturing Systems*, 33(2014), 241–261, Elsevier.

[3] Mandel J., (1984), "The Statistical Analysis of Experimental Data", *National Bureau of Standards*, Dover Publications, Inc., Washington, D.C., ISBN 0-486-64666-1

[4] Bakrania, S. and B. Lucas (2009), "The Impact of the Financial Crisis on Conflict and State Fragility in Sub-Saharan Africa", *Governance and Social Development Resource Centre Issues Paper*, Available: http://www.gsdrc.org/go/emerging-issues#crisis, Accessed: 26 Feb 2016

[5] Zimpapers, (2008), "University of Zimbabwe Forges Strategic Alliances with Industry", *Business Herald*

[6] Yeh, C. W., Li, D. C. and Zhang, Y. R., (2012), "Estimation of a data-collection maturity model to detect manufacturing change", *Expert Systems with Applications*, 39(2012), 7093–7101, Elsevier.

[7] van Antwerpen, C. and Curtis, N. J., (2016), "A data collection and presentation methodology for decision support: A case study of hand-held mine detection device", *European Journal of Operational Research*, 251(2016), 237–251

[8] Arsham, H. (2002), Statistical Data Analysis: Prove it with Statistics, University of Baltimore, Course Material, Available: http://home.ubalt.edu/ntsbarsh/stat-data/Topics.htm, Accessed: 4 March 2016.

[9] Mendenhall W. and Sincich T., (1988), "Statistics for the Engineering and Computer Sciences, Second Edition", *Macmillan Publishing Company*, ISBN 10: 0023804602.

[10] Lin, J. T. and Chen, C. M., (2015), Simulation optimization approach for hybrid flow shop scheduling problem in semiconductor back-end manufacturing, *Simulation Modelling Practice and Theory*, 51 (2015), 100–114, Elsevier

[11] Harish, V.S.K.V. and Kumar, A., (2016), A review on modeling and simulation of building energy systems, Renewable and Sustainable Energy Reviews, 56(2016), 1272–1292, Elsevier.

[12] Sudha, L., Dillibabu, R., Srinivas, S. S. and Annamalai, A., (2016), Optimization of process parameters in feed manufacturing using artificial neural network, *Computers and Electronics in Agriculture*, 120(2016), 1–6, Elsevier.