Journal of Scientific Research and Development 2 (12): 81-87, 2015
Available online at www.jsrad.org
ISSN 1115-7569

A multiple regression analysis approach for mathematical model development in dynamic manufacturing system: a case study

Kok Hoe Ho *, Joshua Prakash, Kok Seng Ong

Department Faculty of Engineering and Green Technology, University Tunku Abdul Rahman, 9, Jalan Bersatu 13/4, 46200 Petaling Jaya, Selangor, Malaysia

Abstract: We proposed an approach using multiple regression analysis to develop a mathematical model that represents a dynamic manufacturing system. Simulation data are specifically analyzed using this multiple regression analysis approach to obtain a data pattern. The aim of the approach is to reduce the gap between theory and real-time data of the system. To evaluate the effectiveness of the proposed mathematical mode, simulation model was first validated using real-time data. The applicability of the proposed mathematical model was evaluated by testing with real-time data. The outcome positively demonstrated that the proposed mathematical model based on multiple regression analysis approach can be used to make predictions in the dynamic manufacturing environment with an acceptable error percentage range. The mathematical development in this field will enhance the future establishment of a decision making model using a spreadsheet in the management field.

Key words: Multiple regression analysis; Dynamic manufacturing system; Mathematical model; Decision making model

1. Introduction

Mathematical modeling in dynamic manufacturing system requires an effective approach to narrow the gap between theory and practice of the model's application (Mula et al., 2006). A survey of related literature has indicated the possibility of using a data pattern, such as fault pattern variable, to formulate a mathematical equation to represent a dynamic manufacturing environment system (Li, 2007). In general, the main objective of this approach is to identify a feasible method to link data pattern to a mathematical equation for practical application. A number of methodologies for developing mathematical model in the system are available, and one of these methodologies is the application of multiple regression analysis in the model. An example of such work is available in (Li, 2007), and this previous work showed that it is possible to develop mathematical model using functional regression approach. The challenge of creating a mathematical model with a small error comparison using simulation and real-time data has been addressed (Chincholkar et al., 2004; Muhammad Marsudi et al., 2009). Mathematical model development and simulation are used simultaneously in research papers. In contrast to simulation, mathematical model still plays an important role in a dynamic manufacturing system. Mathematical model applications in numerical analysis and spreadsheet models are a common practice in the field (Baudin et al., 1992; Jennifer Robinson et al., 2003; Susan Cole and Jennifer Rowley, 1996).

In the work of (Al-Zuheri et al., 2012), the area of dynamic manufacturing environment study is on walking worker assembly line. The authors worked on improving the decision-making process by using a mathematical model and simulation. The authors (Al-Zuheri et al., 2012) compared the predicted results from simulation model and mathematical model including the error value between the models using root-mean square (RMS).

Other researchers (Cao, Z et al., 2012; Catay et al., 2003; Zhang et al., 2014) have stated that semiconductor process is the most complicated manufacturing system involved in a dynamic system. In a previous study (Cao et al., 2012), a bottleneck prediction mathematical model that is based on an improved Adaptive network-based fuzzy inference system (ANFIS) has been proposed. The prediction accuracies from simulation and mathematical models were analyzed (Cao et al., 2012). (Catay, B et. al., 2003) presented a mathematical model for multi-period tool capacity planning in semiconductor manufacturing. The proposed model is validated using computational experiments using Lagrangean-based heuristic solution procedure, which is coded in C-programming language (Catay et. al., 2003). (Jacomino et al., 2014) presented two production planning approaches using mixed-integer programming and heuristics methods to realize a step toward the development of capacity planning at a finite capacity in semiconductor manufacturing. No

* Corresponding Author.

validation activity has been presented for the proposed models in the study (Jacomino et al., 2014). (Jacomino et al., 2014) used CAPACE and ILOG CPLEX solver software to provide and analyze the result from the developed models.

Mathematical model formulation using data pattern analysis has been conducted previously (Li, 2007). (Li, 2007) presented that a model-based signal detection and estimation approach can be applied, and the process can be represented by a suitable dynamics model (either in the form of state-space or transfer function). The author (Li, 2007) studied the data pattern and developed the mathematical model as a regression model that can be used to make predictions by using a functional regression method. (Faraway, 1997) noted that the functional regression approach is suitable when the response variable for prediction is functional. The author (Li, 2007) did not perform validation on the regression model and demonstrated the proposed model using an example from the resistance spot welding process mentioned in previous studies. A previously published review (Zhang et al., 2014) indicates that the multivariate regression approach tends to be more convenient when input data are noise or when the complex relationship among the input variables is not fully understood. (Zhang et al., 2014) cited the work of (Wang and Malakooti, 1992) and (Malhotra et al., 1999) because this approach is widely used in research strategy and is proven to be effective and adaptive.

In this paper, we further extended the work of another researcher (Li, 2007) on different dynamic manufacturing environments. The approach we adopted was to use a data pattern from a simulation model to formulate a mathematical model via regression method. Our definition of dynamic manufacturing system is product movement from the first process to the last process in an assembly semiconductor. A search in a literature database indicated the lack of information related to the development of mathematical model in this defined dynamic system. Hence, we are the first to work in this area to provide additional information by extending the results of a similar study (Li, 2007). The organization of this paper is described as follows. Section 2 discusses an approach that uses multiple regression analysis to formulate the mathematical equation of the system. Section 3 develops the paper's problem formulation. The details of the proposed model and its validation are presented in Section 4. In Section 5, real-time case study from the data collection time frame is tested in the proposed model, and results are discussed and analyzed. Conclusion and suggestions for further work are presented in Section 6.

2. An approach using multiple regression analysis

Fig. 1 illustrates the process flow chart of the approach using multiple regression analysis to develop a mathematical model.

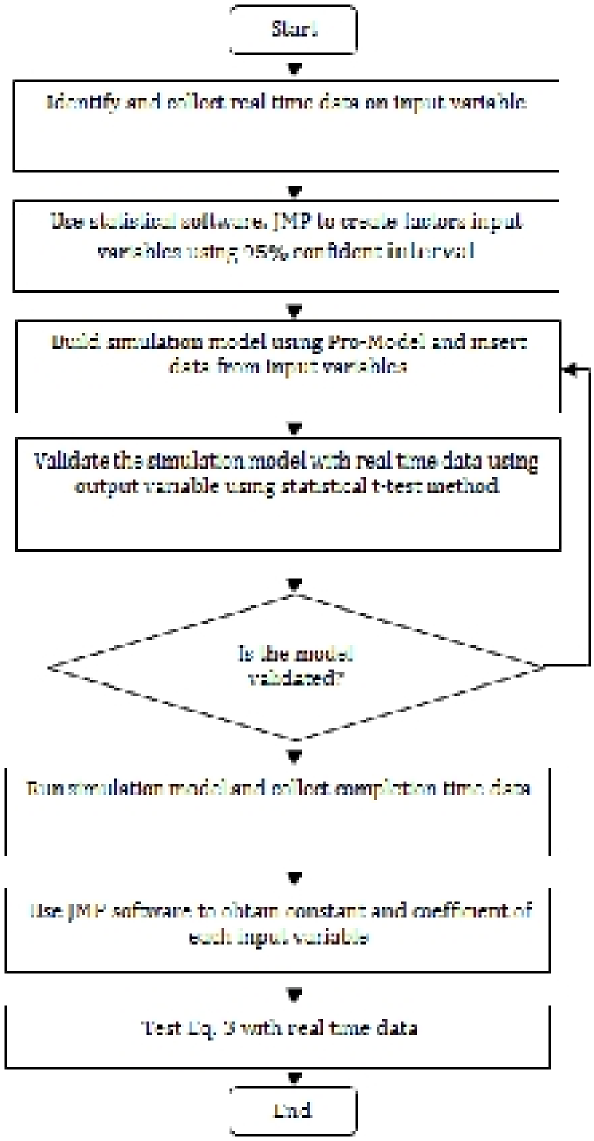

Fig. 1: Approach using multiple regression analysis to develop a mathematical model

The paper began with real-time data collection for input variables. The data were analyzed using JMP software to fit into most suitable data distribution on 2 level factors for 9 input variables based on 95% confidence interval and one 3 level factor for batch quantity. A simulation model was built using Pro-Model, and input variable data are inserted into the model. Model validation activity had to be addressed to ensure its accuracy as a representation of a real system (Martens, J et al., 2006). This paper's simulation model validation was performed by using historical data technique. (Sargent, 2011) described this technique to test whether simulation model behaves as the system does. A statistical t-test was used to validate the model with real-time data on respective processes. When the simulation model was validated, a sample of completion time of 30 batches for each run was collected. There is total of 1536 runs in the design of experiment table using full factorial with 1 replication run. 1 replication run for 1 full factorial study is justified as each factor of each level are covered in the design of experiment table for the data collection of completion time. The full factorial design is applied in the table as this is the first study to develop mathematical model using multiple regression approach on a manufacturing system and test it with real-time data using this approach. For each run, an average of total 30 batches is calculated and inserted into JMP software. Then, the constant and coefficient of each input variables were obtained using JMP analysis. When the equation was established, the equation was tested by performing a comparison between prediction analysis and real-time data (case study).

3. Problem formulation

We examined an example from semiconductor manufacturing as a case study in **Fig. 2**. The configuration of the process is as follows: 3 die attach machines, 1 continuous available oven cure machine, 9 wire bond machines, and 3 pre-cap inspection machines.

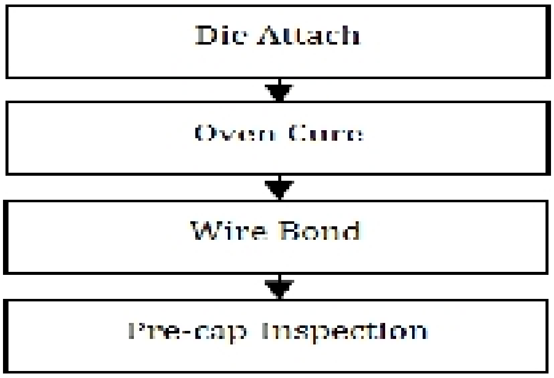

Fig. 2: Semiconductor manufacturing process flow

The response variable is completion time with 11 input variables as follows: cycle time per unit (Die Attach), cycle time per unit (Wire Bond), cycle time per unit (Pre Cap Inspection), machine downtime duration (Die Attach), machine downtime duration (Wire Bond), machine downtime frequency (Die Attach), machine downtime frequency (Wire Bond), setup time duration (Die Attach), setup time duration (Wire Bond), batch quantity, and cycle time per batch (oven cure). The task was to perform regression analysis on the data pattern of completion by using a combination of 2 and 3 level factors of input variables. A total of 1536 runs from a full factorial design of an experiment table using (29 x 31) were performed using Pro Model software to collect completion time data. Three level factors for batch quantity consisted of 2200, 3080, and 11264 units per batch, which were obtained from the assembly. In regression analysis, the completion time could be represented as a function of input variables as per Eq. 1, and db could be represented as the difference between T_b and T_{b-1} in Eq. 2.

The annotation used to explain the regression model is described below:

T_b = average completion time per batch, b
d_b = different between completion time per batch current minus completion time per batch previous (flow time of batch b)
CT_{DA} = cycle time per unit (Die Attach) / unit = seconds
CT_{WB} = cycle time per unit (Wire Bond) / unit = seconds
CT_{PC} = cycle time per unit (Pre Cap Inspection) / unit = seconds
DD_{DA} = machine downtime duration (Die Attach) / unit = seconds
DD_{WB} = machine downtime duration (Wire Bond) / unit = seconds
DF_{DA} = machine downtime frequency (Die Attach) / unit = minutes
DF_{WB} = machine downtime frequency (Wire Bond) / unit = minutes
ST_{DA} = setup time duration (Die Attach) / unit = seconds
ST_{WB} = setup time duration (Wire Bond) / unit = seconds
BQ = batch quantity / unit = quantity
CT_{OC} = cycle time per batch (Oven Cure) / unit = seconds
a_1 = coefficient of CT_{DA}
a_2 = coefficient of CT_{WB}
a_3 = coefficient of CT_{PC}
a_4 = coefficient of DD_{DA}
a_5 = coefficient of DD_{WB}
a_6 = coefficient of DF_{DA}
a_7 = coefficient of DF_{WB}
a_8 = coefficient of ST_{DA}
a_9 = coefficient of ST_{WB}
a_{10} = coefficient of BQ
a_{11} = coefficient of CT_{OC}
k = constant in the Eq. 1.

$$T_b = k + a_1CT_{DA} + a_2CT_{WB} + a_3CT_{PC} + a_4DD_{DA} + a_5DD_{WB} + a_6DF_{DA} + a_7DF_{WB} + a_8ST_{DA} + a_9ST_{WB} + a_{10}BQ + a_{11}CToc \quad (1)$$

$$d_b = T_b - T_{b-1} \quad (2)$$

4. Establishment of proposed model and its validation

Table 1 summarizes the method to analyze raw data into 2 level factors for 9 input variables and into 3 level factors for 1 input variable.

Table 1: Method analysis for 9 input variables using real-time data

Input variables	Data	Distribution
CT_{DA}, CT_{WB}, CT_{PC}	95% confident internal (lower, upper)	Normal distribution
DD_{DA}, DD_{WB}, DF_{DA}, DF_{WB}	95% confident internal (lower, upper)	Exponential distribution, MTBF = mean time between failure
ST_{DA}, ST_{WB}	95% confident internal (lower, upper)	Exponential distribution

Table 2: Input variable data for simulation model

Input variable	Data		
	Low	Middle	High
CT_{DA} (s)	2.8072	-	2.9460
CT_{WB} (s)	6.0902	-	6.0902
CT_{PC} (s)	0.9882	-	1.0498
DD_{DA} (s)	2141	-	4391
DD_{WB} (s)	1364	-	2797
DF_{DA} (min)	1066	-	2187
DF_{WB} (min)	1421	-	2193
ST_{DA} (s)	2957	-	6063
ST_{WB} (s)	1324	-	2714
BQ (unit)	2200	3080	11264
CT_{OC} (s)	7200 (fixed)		

Note: s = seconds; min = minutes

Table 3: Validation low and high setting data of Die Attach process from Table 2 using statistical t-test analysis in JMP software environment between simulation model and real time data on output variable

	Low	High		Low	High
Difference	-1441.5	2404.0	t ratio	-0.94408	1.69163
Std Err Dif	1526.9	1421.1	DF	57.99863	57.98362
Upper CL Dif	1614.9	5248.7	Prob > |t|	0.3490	0.0961
Lower CL Dif	-4498.0	-440.7	Prob > t	0.8255	0.0480
Confidence	0.95	0.95	Prob < t	0.1745	0.9520

Note: Assuming unequal variances

Table 4: Data collection using simulation model in design of experiment table format

Run	CT_{DA}	CT_{WB}	CT_{PC}	DD_{DA}	DD_{WB}	DF_{DA}	DF_{WB}	ST_{DA}	ST_{WB}	BQ	CT_{OC}	T_b
1	2.9460	6.0902	1.0498	4391	2797	1066	1421	6063	1324	3080	7200	38267
2	2.9460	6.0902	1.0498	2141	2797	1066	1421	2957	1324	11264	7200	120811
3	2.8072	6.4609	1.0498	4391	1364	1066	2193	6063	1324	3080	7200	38982
4	2.8072	6.4609	0.9882	4391	2797	2187	2193	2957	2714	3080	7200	38792
5	2.9460	6.4609	0.9882	2141	1364	2187	1421	2957	1324	2200	7200	30072
6	2.9460	6.4609	1.0498	2141	1364	1066	2193	6063	1324	3080	7200	39409
7	2.8072	6.0902	1.0498	4391	1364	2187	2193	2957	2714	2200	7200	29086
8	2.9460	6.0902	1.0498	2141	1364	2187	2193	6063	2714	11264	7200	120811
9	2.8072	6.4609	1.0498	4391	1364	2187	2193	2957	1324	2200	7200	29902
⋮												
1526	2.9460	6.4609	1.0498	4391	1364	2187	2193	6063	1324	3080	7200	39409
1527	2.9460	6.4609	1.0498	2141	1364	2187	2193	6063	1324	2200	7200	30207
1528	2.8072	6.4609	1.0498	2141	2797	1066	1421	6063	1324	11264	7200	123423
1529	2.9460	6.0902	1.0498	2141	2797	1066	1421	2957	2714	2200	7200	29392
1530	2.8072	6.0902	1.0498	4391	1364	2187	2193	6063	1324	3080	7200	37840
1531	2.9460	6.0902	0.9882	4391	2797	1066	2193	2957	2714	2200	7200	29256
1532	2.8072	6.4609	1.0498	4391	2797	2187	2193	6063	2714	2200	7200	29902
1535	2.8072	6.4609	1.0498	2141	2797	2187	2193	6063	2714	3080	7200	38982
1536	2.8072	6.4609	0.9882	2141	2797	1066	2193	2957	2714	3080	7200	38792

Table 2 represents the raw data from a real-time system for input variables to a simulation model. The simulation model from Pro-Model is validated with real-time data of output variable. Table 3 shows the model validation for Die Attach process.

Table 3 illustrates the validation high and low setting data of Die Attach process from Table 2 using statistical t-test analysis in JMP software environment between simulation model and real time data on output variable. In this paper validation, α = 0.05. The p-value in high setting is 0.0961. Since p value was larger than α, it accepts null hypothesis indicating no significant between the model and real-time data. The p-value in low setting is 0.3490. Since p-value is larger than α, the null hypothesis is accepted and the model has no significant different with real-time system.

After validation process was completed, a total of 1536 runs with 1 replication were set using full factorial design (2^9x3^1=1536 when x_{11} is fixed input variable and x10 is 3 level factors). For each run, a 30 sample of completion time are collected to determine the flow time (d_b) and average completion time (T_b). Table 4 indicates the design of experiment table for data collection during simulation run using model from Pro-Model software.

The collected data for Table 4 were analyzed in JMP environment to determine the value of constant and coefficient of each input variables in the regression analysis model. Two interaction input variables were omitted because the result of two interactions had no significant different with one interaction result. The result from JMP is shown below:

The equation T_b in Eq. 2 is stated as per below:

$T_b = k + a_1CT_{DA} + a_2CT_{WB} + a_3CT_{PC} + a_4DD_{DA} + a_5DD_{WB} + a_6DF_{DA} + a_7DF_{WB} + a_8ST_{DA} + a_9ST_{WB} + a_{10}BQ + a_{11}CToc$

$$= 7187.5725 + 387.1239\frac{[\quad -2.8766]}{0.0694} + 1008.0011\frac{[\quad -6.2756]}{0.1854} + 157.9690\frac{[\quad -1.019]}{0.0308} + 9.5167\frac{[\quad -3266]}{1125} + 3.9325\frac{[\quad -2080.5]}{716.5} + 2.4477\frac{[\quad -1626.5]}{560.5} - 0.1807\frac{[\quad -1807]}{386} - 2.6148\frac{[\quad -4510]}{1553} + 24.3713\frac{[\quad -2019]}{695} + 10.1729BQ \quad (3)$$

Note: When a_{11} is zero as CT_{OC} is fixed value (1 level factor)

$$d_b = t_b - t_{b-1} = 0 \quad (4)$$

Based on the regression mode in **Fig. 3**, the r2 value was 0.9995. Since 0.9995 value was near to 100%, it was noted that the model explained all the variability of the response data around its mean and fitted the study's data.

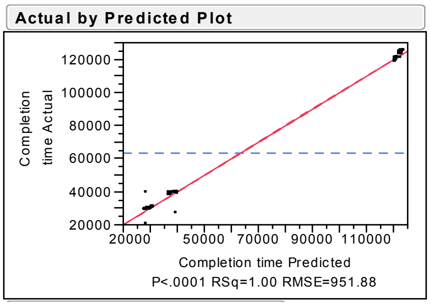

Fig. 3: R_2 value on the graph between completion time predicted and simulation data result

5. Proposed model testing and its validation

Table 5 indicates the comparison of 9 sample results between prediction mathematical and the real-time data on completion time response variable. The 9 samples data are consist of 3 batches which each batch contains 3 samples data. T_{b3} shows result data from prediction mathematical model using Eq. 3. T_{b4} shows the result data from the real time system on completion time. These real-time data of 9 samples were collected within the period along with the real-time data collection of the input variables for the simulation model. The relative error was in the range 11.62%–22.01%, with an average value of 18.85%.

Table 5: Relative error analysis between T_{b3} and T_{b4}

T_{b3}	T_{b4}	Relative error	% error
31371	25380	5991	19.10
33396	28380	5016	15.02
29299	33960	4661	15.91
30086	25260	4826	16.04
27295	22440	4855	17.79
30416	26880	3536	11.62
84660	68040	16620	19.63
83090	64800	18290	22.01
88197	69480	18717	21.22
Average % error			18.85

The error between the mathematical model and sampling real-time data in the study was in the range 10%–25%, which was satisfactorily acceptable in a previous study (Eickemeyer et al., 2014) this previous study cited the work of (Eickemeyer et al., 2013).

According to (Gottlich et al., 2014), such validation is important to show the capability of the model to predict the behavior of the study. In contrast to the work of (Eickemeyer et al., 2014), (Gottlich et al., 2014) used graphical method for the validation and did not state the acceptable level between the model and real time data. Thus, the present paper's proposed mathematical model is tested with real-time data for the validation, citing (Eickemeyer et al., 2014) as reference. Based on the results, we concluded that the model had satisfactory quality estimate and was acceptable for performing prediction in the study with an error percentage in the range 10%–25%.

The Eq. 3 indicates the flow time of a manufacturing system. Simulation activity results showed that the d_b value is zero in the long run because t_b and t_{b-1} were nearly constant. This pattern of the simulation result was probably due to the configuration setting of the process in this study. The regression analysis approach was found in the literature, and this paper used additional variables, such as setup and machine downtime with validation process using different manufacturing environments. The extension work using this approach includes the development of a methodology that uses simulation, design of experiment concept, and statistical tool analysis to derive a mathematical model. This approach is an alternative feasible way of finding a mathematical model in the setting of a configuration process. Although the approach is not new, the methodology is not commonly applied to manufacturing practice. The novelty of this paper is to signify the approach that uses multiple regression analysis to develop a mathematical model to represent a dynamic manufacturing environment. Our work contributes towards introduction of a feasible methodology to formulate a mathematical equation in dynamic manufacturing environment. Our results will benefit the future model work by other researchers, and we list advantages in the following statements. The first advantage is to enable the model to perform prediction in a real-time system with an acceptable error margin. The second advantage is the development of a feasible approach to build a model using multiple regression analysis technique without affecting configuration.

6. Conclusion and future work

In this paper, an approach that involved the use of multiple regression analysis was proposed for the mathematical development in a dynamic manufacturing system. The main goal of this mathematical model was to introduce a method for model development to narrow the gap between theory and practice in a dynamic manufacturing environment.

The effectiveness of the proposed mathematical model was evaluated using real-time data. The result between proposed mathematical model and real-time data showed positive effects when considering data from previous literature. Results showed that the model is satisfactory and acceptable and can be used to make predictions.

Future studies need to focus on improving the error percentage between the mathematical model and real-time data and on enhancing the methodology for mathematical development in this field. Our results can be used to continuously enhance the proposed model and to apply the model to real-time dynamic manufacturing systems in the future.

Acknowledgment

The authors wish to thank MyBrain15 program from Ministry of Education in Malaysia and University Tunku Abdul Rahman (UTAR) under UTARRF 6200/J09 who provided financial support for this research. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, who helped improve our presentation.

References

Al-Zuheri, a., Luong, L., & Xing, K. (2012). The role of randomness of a manual assembly line with walking workers on model validation. *Procedia CIRP*, *3*, 233–238. doi:10.1016/j.procir.2012.07.041

Baudin, Michel Mehrotra, V., Tullis, B. Yeaman, D. Hughes, R.A. (1992). From spreadsheet to simulations: a comparison of analysis methods for IC manufacturing performance. Semiconductor Manufacturing Science Symposium, IEEE / SEMI International

Cao, Z., Deng, J., Liu, M., & Wang, Y. (2012). Bottleneck prediction method based on improved adaptive network-based fuzzy inference system (ANFIS) in semiconductor manufacturing system. *Chinese Journal of Chemical Engineering*, *20*(6), 1081–1088. doi:10.1016/S1004-9541(12)60590-4

Çatay, B., Erengüç, S.¸S., & Vakharia, A. J. (2003). Tool capacity planning in semiconductor manufacturing. *Computers and Operations Research*, *30*, 1349–1366. doi:10.1016/S0305-0548(02)00075-8

Chincholkar, M.M, Barroughs, T., Herrmann, J.W. (2004). Estimating manufacturing cycle time and throughput in flow shops with process drift and inspection. Institutes of System Research and Department of Mechanical Engineering, University of Maryland

Eickemeyer, S. C., Steinkamp, S., Schuster, B., Bodenhage, F., & Nyhuis, P. (2014). Reliable Capacity Planning Despite Uncertain Disassembly, Regeneration and Reassembly Workloads by Using Statistical and Mathematical Approaches – Validation in Subsidiaries of a Global MRO Company with Operations in Asia, Europe and North America. *Procedia CIRP*, Volume *23*,pages 252–257. doi:10.1016/j.procir.2014.10.097

Göttlich, S., Hoher, S., Schindler, P., Schleper, V., & Verl, A. (2014). Modeling, simulation and validation of material flow on conveyor belts. *Applied Mathematical Modelling*, *38*, 3295–3313. doi:10.1016/j.apm.2013.11.039

Jacomino, M. E. M. F. M. P. V. G. L. (2014). A Step Toward Capacity Planning at Finite Capacity in Semiconductor Manufacturing, 2239–2250.

Jennifer Robinson, John Fowler, Eileen Neacy. (2003). Capacity loss factors in semiconductor manufactuiring. Available from: www.fabtime.com/files/CapPlan.pdf [Accessed 28 December 2013]

J. Faraway. (1997). Regression analysis for functional response. Technometrics, 254-261. Available:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.27.5397&rep=rep1&type=pdf [Accessed 29 December 2014]

Li, W. (2007). Manufacturing process diagnosis using functional regression. *Journal of Materials Processing Technology*, *186*(April 2004), 323–330. doi:10.1016/j.jmatprotec.2006.12.052

Martens, J., Put, F., & Kerre, E. (2006). A fuzzy set theoretic approach to validate simulation models. *ACM Transactions on Modeling and Computer Simulation*, *16*(4), 375–398. doi:10.1145/1176249.1176253

Muhammad Marsudi, Dzuraidah Abdula Wahab, Che Hassan Che Haron. (2009). Application of spreadsheet and queueing network model to capacity optimization in product development. World Academy of Science, Engineering and Technology. Available from: http://waset.org/publications/8394/application-of-spreadsheet-and-queuing-network-model-to-capacity-optimization-in-product-development [Accessed 18 December 2013]

Mula, J., Poler, R., García-Sabater, G. S., & Lario, F. C. (2006). Models for production planning under uncertainty: A review. *International Journal of Production Economics*, *103*(1), 271–285. doi:10.1016/j.ijpe.2005.09.001

Sargent, R. G. (2011). Verification and validation of simulation models. *Proceedings of the 2011 Winter Simulation Conference S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M.*, 183–198.

Susan Coles, Jennifer Rowley. (1996). Spreadsheet modeling for management decision making. IndustrialManagement and Data System, Emerald

Zhang, J., Qin, W., Wu, L. H., & Zhai, W. B. (2014). Fuzzy neural network-based rescheduling decision mechanism for semiconductor manufacturing. *Computers in Industry*, *65*, 1115–1125. doi:10.1016/j.compind.2014.06.002