

A Hybrid Model of ARIMA and Multiple Polynomial Regression for Uncertainties Modeling of a Serial Production Line

Amir Azizi, Amir Yazid b. Ali, Loh Wei Ping, and Mohsen Mohammadzadeh

Abstract—Uncertainties of a serial production line affect on the production throughput. The uncertainties cannot be prevented in a real production line. However the uncertain conditions can be controlled by a robust prediction model. Thus, a hybrid model including autoregressive integrated moving average (ARIMA) and multiple polynomial regression, is proposed to model the nonlinear relationship of production uncertainties with throughput. The uncertainties under consideration of this study are demand, break-time, scrap, and lead-time. The nonlinear relationship of production uncertainties with throughput are examined in the form of quadratic and cubic regression models, where the adjusted R-squared for quadratic and cubic regressions was 98.3% and 98.2%. We optimized the multiple quadratic regression (MQR) by considering the time series trend of the uncertainties using ARIMA model. Finally the hybrid model of ARIMA and MQR is formulated by better adjusted R-squared, which is 98.9%.

Keywords—ARIMA, multiple polynomial regression, production throughput, uncertainties

I. INTRODUCTION

THOUGHT analysis is an important and efficient way to control and match the production output with the ordered demands. Mostly the throughput of production line does not meet the required demand on the shop floor of production especially in presence of product mix and multi stages of production line. Many variables can affect on the throughput degradation of each stage for example break down of machine, lead time of manufacturing, and scrap, which caused maybe by error of machines, material, and workers. Changes in demand in terms of type and volume also affect the throughput because of changing of customer needs and interests. On the other hand, the company requires having innovation on design of new products in order to survive in today's competitive manufacturing world.

Amir Azizi is with the Universiti Sains Malaysia, School of Mechanical Engineering, Penang, Malaysia (phone: +604-5937788; fax: +604-5941025; (e-mail: amirazizi@gmail.com).

Amir Yazid b. Ali is with the Universiti Sains Malaysia, School of Mechanical Engineering, Penang, Malaysia (phone: +604-5996331; fax: +604-5941025; (e-mail: meamir@eng.usm.my).

Loh Wei Ping is with the Universiti Sains Malaysia, School of Mechanical Engineering, Penang, Malaysia (phone: +604-5996397; fax: +604-5941025; (e-mail: meloh@eng.usm.my).

Mohsen Mohammadzadeh is with the Tarbiat Modares University Tehran, Iran (phone: +9821-82883483; fax: +9821-82883483; (e-mail: mohsen_m@modares.ac.ir).

[11] emphasized to have the right demand quantity estimating for surviving in a constantly fluctuating business environment. Managing random variables of production by making a robust estimating maximize the profitability. Production line uncertainty is taken to attention recently because of needs to handle the uncertainty using development of new methodology and computational approaches. Study on production uncertainty is an opportunity for new research and development [6]. Analyzing and estimating throughput is being crucial because the efficiency of the production system is usually measured using throughput [15]. Managing production in terms of supply and demand requires forecasting of both time delivery and quantity [11], [12]. Different strategies and approaches are proposed to overcome the production uncertainties. However there is not a formal strategy or standard approach [11]. It is still under development and optimization. The empirical study on the real production line and actual data from industry are required overcoming to the production uncertainties.

Typically estimating and forecasting methods are divided into four main groups [11], [10]: (1) Qualitative methods are primarily subjective like Delphi technique and scenario methods; they rely on human judgment and opinion to make a forecast. (2) Time-series methods use historical data to make a forecast. (3) Causal methods involve assuming that the demand forecast is highly correlated with certain factors in the environment (e.g., the state of the economy, interest rate). (4) Simulation methods imitate the consumer choices that give rise to demand to arrive at a forecast. The combination of several approaches may results a superior accuracy and become a robust problem solving methodology. Simulation can be applied for any production system to estimate the throughput, however it is not robust and this becomes a computational chore when the number of alternatives to be examined is large [15]. This paper offers a simple, robust procedure for determining the throughput.

II. LITERATURE REVIEW

Throughput is considered for analysis and modeling as an important measure of production line performance [16]-[18]. [20] provided a review paper of models under uncertainty for production planning. He highlighted that the models for production planning, which consider the uncertainty can make superior planning decisions compared to those models that do

not present for the uncertainty. On the other hand, [23] have shown using simulation that ignoring uncertainty sources lead to wrong decisions. [3] categorized uncertainties into two groups: (1) environmental uncertainty and (2) system Uncertainty. Before 1990 focusing on uncertainty was more on environmental uncertainty [2]. Investigating about uncertainties on a production line is launched by [1]. [22] compiled all the uncertain factors through different sources, which are system uncertainty, lead time uncertainty, environmental uncertainty, supply uncertainty, operation yield uncertainty, interrelationship between levels, demand uncertainty, probabilistic market demand and product sales price, capacity, breakdown, changing product mix situation, labour hiring and lay-offs, quantity uncertainty, cost parameters, and quality. Many papers worked on throughput analysis using conventional approaches such as simulation and analytical methods [16].

Simulation method and approximation algorithm are applied for analyzing throughput under uncertainty such as unreliable machine and random processing times, for example studied by [24] & [25]. [19] provided an analytical equation for the general case where there are two workstations in a serial production line. In his model, the workstations have unequal processing time, downtime, and buffer size, while [16] considered a serial production line including two workstations with same speed and buffer size. [19] and [18] demonstrated that the processing time and down time affect the throughput or production volume. [23] examined the effects of three uncertainties namely demand, manufacturing delay, and capacity scalability delay. They found that manufacturing delay has highest impact. A recent survey have been performed on material shortage, labor shortage, machine shortage, and scrap to show the association of these uncertainties on the product tardy delivery through analysis of variance, correlation analysis and cluster analysis [21].

[4] proposed to use buffer to manage uncertainty in production system. However they did not make a robust decision by forecasting based on relationship of uncertainties and throughput. Later, [5] studied on supply-demand mismatches. They believe that the long delivery time of throughput to supplier caused because of lead time uncertainty in production system, which leads to lost sales. However in their proposed methodology to manage lead time uncertainty, they did not consider other production uncertainties. And also the rate of demand is assumed to be constant in their work. Approximate method also is used for forecasting throughput, [15] presented an analytical algorithm to analyze and predict the production throughput under unbalanced workstations, where operation times of stations are random. A hybrid combination of autoregressive integrated moving average models and neural network for demand forecasting in supply chain management is presented by [12], [13]. They developed a replenishment system for a Chilean supermarket. The linear regression models for strategy, environmental uncertainty and performance measurement in New Zealand manufacturing firms are formulated by [7].

A data mining approach is utilized for cycle time prediction by [14]. A panel or longitudinal data sets for uncertain demand and price have been considered to evaluate the alternative capacity strategies using simulation [26]. Recently [27] proposed an autoregressive moving average model for throughput bottleneck prediction of a serial production line under production blockage and starvation times. Other new attempts have been carried out using ARIMA or other methods combined with RAIMA to develop the forecasting method in manufacturing area [28], [29].

Stochastic variables of production lines are studied separately for example on breakdown by [17] and on processing time by [15]. This study is considering more variability into the consideration. Variability can be measured by the coefficient of variation [15]. Therefore the economic uncertainty needs the mathematical Models [11]. To do the accurate estimating and modeling on production uncertainties, this paper is organized in three sections. Section III presents the multiple polynomial regressions (MPR). And section IV stated the procedure of modeling using autoregressive integrated moving average (ARIMA). A real case study from tile industry, located in Iran, is considered for modeling and estimating the production uncertainties. Section V gives the results of the hybrid model. Section VI shows the conclusion of this paper and it provides a future direction of this study.

III. MULTIPLE POLYNOMIAL REGRESSION MODELING

MPR is formulated in terms of quadratic form and cubic form according to [8]. The accuracy of both MPR models is examined by their adjusted R-squared. The fitted multiple quadratic regression (MQR) is formulated in (1).

$$\hat{P} \sim 24123626 + 33061 B - 202 B^2 - 2389 D + 1.22 D^2 - 23867 L + 7.21 L^2 - 1820 S + 0.369 S^2. \quad (1)$$

R-squared = 98.4% and adjusted R-squared = 98.3%.

Where

B = Breakdown time over the 104 weeks,

D = Demand volume over the 104 weeks,

L = Lead time of manufacturing over the 104 weeks,

S = Scrap volume over the 104 weeks,

\hat{P} = estimated throughput over the 104 weeks.

The fitted multiple cubic regression (MCR) is formulated in (2).

$$\hat{P} \sim 7.71E+11 + 4.27E+08 B - 4014894 B^2 + 11044 B^3 - 28634140 D + 7645 D^2 + 0.207 D^3 - 1.40E+09 L + 888066 L^2 - 187 L^3 - 24403538 S + 22425 S^2 - 8.1 S^3. \quad (2)$$

R-squared = 98.4% and adjusted R-squared = 98.2%.

IV. METHODOLOGY OF ARIMA MODELING

The ARIMA uses the correlation techniques. ARIMA has three parameters; p, d, and q. Where p is autoregressive, d is integrated, and q is moving average. The ARIMA parameters should be integer and positive [9]. The minimum value is zero and the maximum value allowed to apply is 5 for

p, d, q. Note that at least one of the p, d, q should not be zero. The flow chart of ARIMA modeling is illustrated in Fig 1.

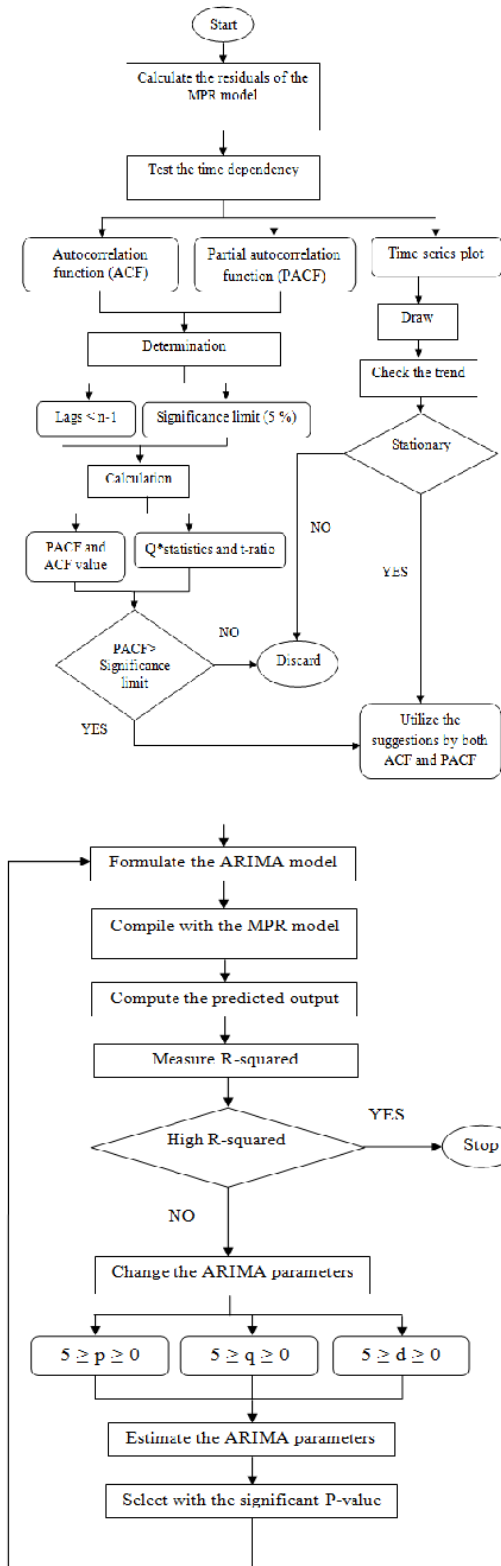


Fig. 1 Flow chart

A. Calculation of the residuals of the MPR model

To examine the time dependency of the random variables affected the production throughput, the residuals of the MQR model are calculated to show the error and gap of the MQR model output with the actual production throughput.

B. Test the time dependency

The time dependency is tested through three ways, which are partial autocorrelation function (PACF), autocorrelation function (ACF), and time series plot. These tests demonstrate that whether the ARIMA model is suitable or not.

C. Autocorrelation function (ACF)

ACF is applied to identify autoregressive and moving average processes, shown in Fig. 2. It breaks off with q in moving average process and it does not break off for the autoregressive process [30].

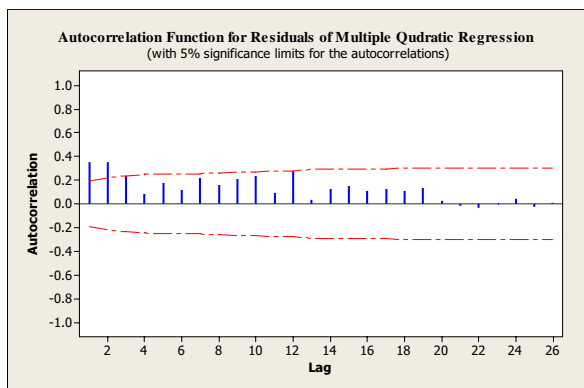


Fig. 2 ACF of residuals of MQR

D. Partial autocorrelation function (PACF)

PACF is another method used correlation technique to present autoregressive and moving average, shown in Fig. 3. The difference of PACF with ACF is that PACF breaks off with p in the autoregressive process and it does not break off for the moving average process [30].

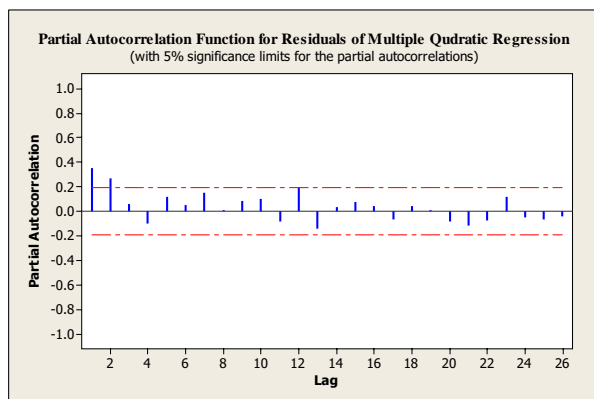


Fig. 3 PACF of residuals of MQR

E. Determination

In this action, the lag and significance limit should be determined for calculation of ACF and PACF. The maximum

lag must be less than 103 for our data set. The significance limit usually considered 5%. It implies that the parameters of ARIMA are significance if the absolute value of t-ratio is large than 1.96. From the ACF and PACF analysis if there is correlation over a time period it will show up by amount of ACF and PACF, which are bigger than significance limit. To determine whether the time series is white noise or not, the Ljung-Box-Q* (LBQ) statistic is applied to compare with the chi-square distribution with (L-n) degrees of freedom, where L is the number of lags and n is the number of parameters. It is obtained through (3). If it is lower it is better to determine white noise. LBQ statistic is computed as [10]:

$$LBQ = T(T+2) \sum_{i=1}^L \frac{r_i^2}{(T-i)} \quad (3)$$

Where

r = Autocorrelation,

T = Number of observation,

L = Lag.

F. Time series plot

Time series plot is another test to expose the trend of data over a period of time. It depicts if the output has any pattern over a period of time. It shows how values of the series are correlated with past values of the series. If the data is not stationary around the average, the moving average method is not adequate. Time series plot of residuals of MQR is presented in Fig. 4.

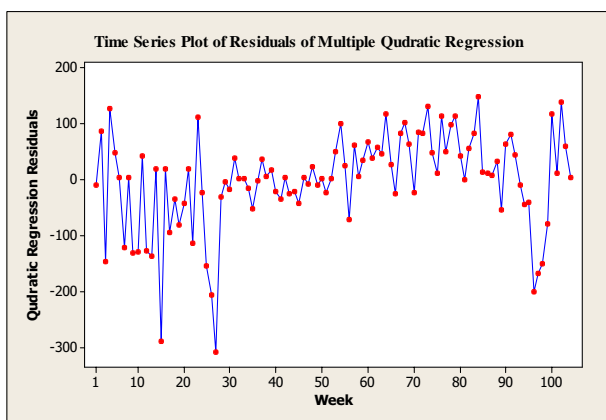


Fig. 4 Time series plot of residuals of MQR

G. Formulate the ARIMA model

Three stages are performed for ARIMA in order to find the ARIMA parameters (p,d,q); (1) the identification, (2) estimation, and (3) forecast, which is summarized as follows.

In the identification stage, the response is specified and a candidate ARIMA model is identified. The autocorrelations and partial autocorrelations are computed. The white noise test should be determined, which is an approximate statistical test of the hypothesis.

In the estimation stage, the parameters of the model are estimated and the ARIMA model is specified to fit to the determined variable in the identification stage.

In the forecasting stage, the future values of the time series are forecasted and confidence intervals for these forecasts can be generated.

After analyzing and testing by ACF, PACF and time series plot, both seasonal and non seasonal are examined to determine the parameters. We started by a seasonal autoregressive (p) = 0, difference (d) = 0, and seasonal moving average (q) = 1 until the random noise remains and lowest forecasting error obtained. Different iterations and simulations have been performed by Minitab software to generate the best value for p, d, q. The results demonstrated that the best value of ARIMA parameters are when p=2, d=0, and q=2. The coefficients of the ARIMA model are tabulated in Table I.

TABLE I
FINAL ESTIMATES OF ARIMA (2,0,2) PARAMETERS

Type	Coefficient	SE Coefficient	T	P
SAR12	1.3362	0.0443	30.19	0.000
SAR24	-0.9644	0.0432	-22.32	0.000
SMA12	1.0731	0.120	8.89	0.000
SMA24	-0.7148	0.1285	-5.56	0.000
Constant	1.502	0.942	2.32	0.037

Where

SAR= seasonal autoregressive,

SMA= seasonal moving average,

SE= Standard error,

T= t statistic,

P=P-value,

Number of observations = 104,

Degree of freedom = 99.

The ARIMA (2,0,2) model is formulated by 95% probability according to [10] in (4).

$$ARIMA(2,0,2) \sim 1.502 + 1.3362 \varepsilon_{t-1} - 0.9644 \varepsilon_{t-2} + a_t + 1.0731 a_{t-1} - 0.7148 a_{t-2} \quad (4)$$

V. RESULTS COMPARISON AND DISCUSSION

Final result is obtained from combination of the (4) and (1), which is presented in (5) by adjusted R-squared 98.9 % accuracy.

$$\hat{P} \sim 24123626 + 33061 B - 202 B^2 - 2389 D + 1.22 D^2 - 23867 L + 7.21 L^2 - 1820 S + 0.369 S^2 + 1.502 + 1.3362 \varepsilon_{t-1} - 0.9644 \varepsilon_{t-2} + a_t + 1.0731 a_{t-1} - 0.7148 a_{t-2} \quad (5)$$

Adjusted R-Square is calculated through the following equations.

$$\text{Adjusted R-Square} = 1 - \frac{\text{MSE}}{\text{SST} / \text{DF Total}} \quad (6)$$

Where

$$\text{MSE} = \frac{1}{n} (\sum_{t=1}^n (\text{Observed data} - \text{Forecasted data})^2) \quad (7)$$

$$\text{SST} = \text{SSR} + \text{SSE} \quad (8)$$

$$\text{SSE} = \sum (y - \hat{y})^2 \quad (9)$$

$$\text{SSR} = \sum (\hat{y} - \bar{y})^2 \quad (10)$$

DF= degree of freedom.

And

y = observed response,

\hat{y} = fitted response,

\bar{y} = mean response.

The output of (5) is compared with the actual throughput data to see the gap as shown in Fig. 5.

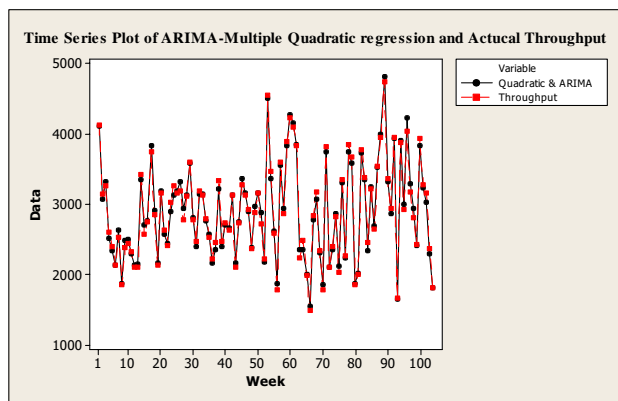


Fig. 5 Time series plot of the ARIMA-MPR model and actual throughput

VI. CONCLUSION

This paper have modeled the random variables of a serial tile production line consist of demand, break-time, scrap, and lead-time. The contribution of this paper was to consider more uncertainties and propose a hybrid method of ARIMA and MQR to estimate the nonlinear relationship of the random variables with the production throughput. The proposed model can be used to predict the production throughput efficiently, and presents the mathematical relationship between the key production uncertainties and throughput. It provides quick "what-if" comparisons. Other types of production systems and industries are recommended for future studies.

ACKNOWLEDGMENT

This study was supported by the Universiti Sains Malaysia (1001/PMEKANIK/804242).

REFERENCES

- [1] C. S. Tang, "The impact of uncertainty on a production line," *Management Science*, pp. 1518-1531, 1990.
- [2] P. M. Swamidass and W. T. Newell, "Manufacturing strategy, environmental uncertainty and performance: a path analytic model," *Management Science*, pp. 509-524, 1987.
- [3] C. J. Ho, "Evaluating the impact of operating environments on MRP system nervousness," *International Journal of Production Research*, vol. 27, pp. 1115-1135, 1989.
- [4] R. Stratton, Robey, D., and Allison, I., "Utilising buffer management to manage uncertainty and focus improvement," in *Proceedings of the International Annual Conference of EurOMA*, Groningen, the Netherlands, 2008.
- [5] P. Kouvelis and J. Li, "Flexible Backup Supply and the Management of Lead Time Uncertainty," *Production and Operations Management*, vol. 17, pp. 184-199, 2008.
- [6] S. C. Graves, "Uncertainty and Production Planning," *Planning Production and Inventories in the Extended Enterprise*, pp. 83-101, 2011.
- [7] Z. Hoque, "A contingency model of the association between strategy, environmental uncertainty and performance measurement: impact on organizational performance," *International Business Review*, vol. 13, pp. 485-502, 2004.
- [8] X. Yan and X. Su, *Linear regression analysis: theory and computing*: World Scientific Pub Co Inc, 2009.
- [9] G. Kirchgässner and J. Wolters, *Introduction to modern time series analysis*: Springer Verlag, 2007.
- [10] T. h. Boon, *Business & economic forecasting techniques and applications*: University Putra Malaysia press, serdang, Selangor darul ehsan, 2006.
- [11] T. Efendigil, Önüt, S., Kahraman, C., "A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis," *Expert Systems With Applications*, vol. 36, pp. 6697-6707, 2009.
- [12] L. Aburto and R. Weber, "Improved supply chain management based on hybrid demand forecasts," *Applied Soft Computing*, vol. 7, pp. 136-144, 2007.
- [13] M. Khashei and M. Bijari, "A New Hybrid Methodology for Nonlinear Time Series Forecasting," *Modelling and Simulation in Engineering*, 2011.
- [14] C. F. Chien, Hsiao, C.W., Meng, C., Hong, K.T. and S. T. Wang, "Cycle time prediction and control based on production line status and manufacturing data mining," 2005, pp. 327-330.
- [15] K. R. Baker and S. G. Powell, "A predictive model for the throughput of simple assembly systems," *European journal of operational research*, vol. 81, pp. 336-345, 1995.
- [16] D. E. Blumenfeld and J. Li, "An analytical formula for throughput of a production line with identical stations and random failures," *Mathematical Problems in Engineering*, vol. 3, pp. 293-308, 2005.
- [17] J. Li, et al., "Comparisons of two-machine line models in throughput analysis," *International Journal of Production Research*, vol. 44, pp. 1375-1398, 2006.
- [18] J. Li, et al., "Throughput analysis of production systems: recent advances and future topics," *International Journal of Production Research*, vol. 47, pp. 3823-3851, 2009.
- [19] J. Alden, "Estimating performance of two workstations in series with downtime and unequal speeds," *General Motors Research & Development Center, Report R&D-9434*, Warren, MI, 2002.
- [20] J. Mula, et al., "Models for production planning under uncertainty: A review," *International Journal of Production Economics*, vol. 103, pp. 271-285, 2006.
- [21] S. Koh, et al., "A business model for uncertainty management," *Benchmarking: An International Journal*, vol. 12, pp. 383-400, 2005.
- [22] M. A. Wazed, et al., "Uncertainty factors in real manufacturing environment," *Australian Journal of Basic and Applied Sciences*, vol. 3, pp. 342-351, 2009.
- [23] A. M. Deif and H. A. ElMaraghy, "Modelling and analysis of dynamic capacity complexity in multi-stage production," *Production Planning and Control*, vol. 20, pp. 737-749, 2009.
- [24] H. Tempelmeier, "Practical considerations in the optimization of flow production systems," *International Journal of Production Research*, vol. 41, pp. 149-170, 2003.
- [25] M. S. Han and D. J. Park, "Optimal buffer allocation of serial production lines with quality inspection machines," *Computers & Industrial Engineering*, vol. 42, pp. 75-89, 2002.
- [26] Y. C. Chou, et al., "Evaluating alternative capacity strategies in semiconductor manufacturing under uncertain demand and price scenarios," *International Journal of Production Economics*, vol. 105, pp. 591-606, 2007.
- [27] L. Li, et al., "Throughput Bottleneck Prediction of Manufacturing Systems Using Time Series Analysis," *Journal of Manufacturing Science and Engineering*, vol. 133, p. 021015, 2011.
- [28] C. M. Lee and C. N. Ko, "Short-term load forecasting using lifting scheme and ARIMA models," *Expert Systems With Applications*, 2011.
- [29] F. Z. a. S. Zhong, "Time series forecasting using a hybrid RBF neural network and AR model based on binomial smoothing," *World Academy of Science, Engineering and Technology*, vol. 75, pp. 1471-1475, 2011.
- [30] R. I. D. Harris and R. Sollis, *Applied time series modeling and forecasting*: J. Wiley, 2003.



Amir Azizi was born in Qazvin, Qazvin Province, Iran in 1982. He received the BSc. degree from Qazvin Islamic Azad University, Iran in 2005, and M.Sc. from University Putra Malaysia (UPM), Malaysia in 2007. He is currently pursuing the Ph.D degree in industrial engineering from Universiti Sains Malaysia (USM), Malaysia. His research interests are decision making, modelling, and estimating under uncertainty in production systems management. He gained the ASME grant at the International Student Conference, Yeditepe University, Department of Mechanical Engineering, Istanbul, Turkey, May 26-28, 2006, and the USM Fellowship for Doctoral program in 2008. He was the lecturer at the Kuala Lumpur Infrastructure University College, Kajang, Selangor, Malaysia in 2007 and at the Paramount College of Technology (partner of International University College of Technology Twintech), Prai, Penang, Malaysia in 2009. He is a member of American Society of Mechanical Engineering (ASME) and Institute of Industrial Engineers (IIE).

of Statistical Sciences and associate editor for the Iranian Journal of Mathematical Sciences and Informatics. He has supervised 33 MSc and 8 PhD Students. He was Academic Visitor in Department of Statistics, Leeds University, 2009. He is interested in spatial and tempo-spatial data analysis, Bayesian statistical inference and official statistics. Prof. Mohammadzadeh is a member of IRSS, IAMG, ASA and IMS scientific societies.



Amir Yazid b. Ali is currently a senior lecturer in Universiti Sains Malaysia (USM). He received his Ph.D degree from Universiti Malaya, Malaysia. His expertise is in Industrial Engineering.



Loh Wei Ping is currently a senior lecturer in Universiti Sains Malaysia (USM). She received her Ph.D degree from Universiti Sains Malaysia. Her expertise is in Mathematical modeling.



Mohsen Mohammadzadeh was born in Mashhad, Iran at April 5, 1957. He received his B.S. (1980) in Statistics from Ferdowsi University of Mashhad, Iran, M.Sc. (1986) in Tarbiat Modares University, Tehran, Iran, and PhD (1996) in Leeds University, Leeds, UK, in the field of Spatial Statistics. He is an associate professor of Statistics at Tarbiat Modares University Tehran, Iran. Since 2004, he has joined to the Ordered and Spatial Data Centre of Excellence at Ferdowsi University of Mashhad, Iran. His main research interests are Spatial Statistics, Space-time data analysis, Statistical Inference and official Statistics. Since 2005, he has been the chief editor of the Journal