Model Development and Comparative Study of Bayesian and ANFIS Inferences for Uncertain Variables of Production Line in Tile Industry

Amir Azizi, Amir Yazid b. Ali, and Loh Wei Ping
School of Mechanical Engineering, Universiti Sains Malaysia (USM), Malaysia
E-mail: amirazizi@gmail.com, meamir@eng.usm.my, meloh@eng.usm.my

Abstract: - The life cycle of tile products are decreasing especially for customized products. The demand changes also fluctuate from time to time for each product type. This phenomena created crucial issue in meeting customers' demands within required due date. The occurrences of uncertain conditions caused the production line performance not able to meet the requirement because they faced uncertain changes in setup time, machinery breakdown time, lead time of manufacturing, and scraps. Hence, an accurate estimation on the production line in the presence of these uncertainties is required. Robust decision making on production line could be made when an accurate estimation of uncertain variables is modeled. Two approaches based on Bayesian inference and adaptive neuro-Fuzzy inference system (ANFIS) were utilized in this study for models development to estimate the effect of uncertain variables of production line in the tile industry. The models were validated and tested based on data obtained from a tile factory in Iran. The strength of our developed models is that the coefficients of decision variables are nonconstant. The best model was judged according to the mean absolute percentage error (MAPE) criterion. The results demonstrated that the ANFIS model generates the lower MAPE by 0.022 and higher correlation by 0.991 compared to the Bayesian model. Consequently, better decisions are generated due to easier identification of uncertainty data and the elaboration made the production planning process better understood.

Key-Words: - Adaptive neuro-Fuzzy inference system, Bayesian, Uncertainties, Production, Throughput.

1 Introduction

The first stage of uncertainty modeling is the definition of uncertainty, where the true values of the input uncertainties are unknown. The classical (frequenters) time series mathematical models are not suitable and inadequate for handling the uncertainty in dynamic production system, because of their inability to handle stochastic variables with random coefficients.

Two more robust approaches based on possibility and probability theories were proposed in this study for modeling the production uncertainties. Fuzzy inference under the possibility theory and Bayesian inference under the probability theory were considered.

In the literatures, several generalization of Fuzzy inference systems were proposed, for example, IF-inference system in [33] and interval type-2 Fuzzy logic system in [34]. The robustness of Fuzzy inference system (FIS) was improved by utilizing the artificial neural network (ANN) and it is called Adaptive neuro-Fuzzy inference system (ANFIS). For example, an ANFIS model was developed under uncertainties by [38] for production throughput to study the prediction capability of ANFIS compared to multiple linear regression. [37] developed a simple Sugeno neuro-Fuzzy predictive controller based on the synergism of a Sugeno neuro-Fuzzy controller and a Sugeno plant predictor for the control of a nonlinear plant under uncertainties. [35] compared the results of the neural networks and Fuzzy logic based on the prediction accuracy. Beside, [36] proposed the software agent paradigm to model the behaviour of complex systems under several scenario conditions. [39] proposed an autoregressive integrated moving average (ARIMA) based on multiple polynomial regression for throughput modeling under production uncertainties. Later, the performance of the hybrid model of ARIMA and Bayesian has been developed by [40].

According to GUM/ISO, the propagation of uncertainty is known as the propagation of probability distributions. The uncertain inputs are characterized by prior probability distributions and treated mathematically as random variables. The best type of probability distribution for defining uncertainty is introduced as normal distribution. Hence, the uncertainty framework characterized the output quantity by a Gaussian function.

This study presents the efficiency of ANFIS and Bayesian approaches, by modeling the actual production throughput of a tile factory using five input uncertainties: demand, breakdown time, scrap, setup time, and lead time. ANFIS and Bayesian approaches expressed the possibility and probability of occurrence of input uncertainties by defining the multimembership functions in Fuzzy environment and prior distributions in probability environment respectively. Estimations were presented with the lower and upper limits in both approaches.

2 Literature Review

Throughput is considered an important parameter of production line performance [1–3]. Mula et al (2006) reviewed models under uncertainty for production planning and highlighted that superior planning decisions were made by models for production planning that considered uncertainty compared to models that did not. Simulation method and approximation algorithm also analyze throughput under uncertainties, such as unreliable machine and random processing times [6, 7]. In a model consist of two workstations in a serial production line, [1] considered the same speed and buffer size for each workstation whereas [8] considered general case for workstations having unequal processing time, downtime, and buffer size, and provided an analytical equation.

Processing time and breakdown time affected the production throughput [3, 8] and [5] examined the effects of three uncertainties, namely, demand, manufacturing delay, and capacity scalability delay. A survey [9] performed on material shortage, labor shortage, machine shortage, and scrap showed the association of these uncertainties on the product tardy delivery.

Although against lean manufacturing principles, [10] proposed using buffer to manage uncertainty. Later, [11] reported on supply-demand mismatches. Lead time uncertainty was concluded to be the cause of increased delivery time to supplier. Methodology to manage lead time uncertainty was proposed, by assuming constant demand rate and not considering other production uncertainties. Approximate method was also used for forecasting throughput. Analytical algorithm presented by [12] analyzed and predicted the production throughput under unbalanced workstations. Linear regression models was used by [13] for formulating strategy, environmental uncertainty, and performance measurement. Bayesian approach was explicitly used by [14] for external evidence in the design, monitoring, analysis, interpretation, and reporting of scientific investigations. The most appropriate method in this context is Markov chain Monte Carlo (MCMC), and used in virtually all recently conducted Bayesian approaches [15]. The popular MCMC procedure is Gibbs sampling, which has also been widely used for sampling from the posterior distribution based on stochastic simulations for complex problems [18]. Gibbs Sampling (BUGS) was used by [19] to solve complex statistical problems. For moderate-sized datasets involving standard statistical models, a few thousand iterations should be sufficient [20].A complete statistical analysis always includes both descriptive statistics and statistical inference. Development moves gradually from description to inference. Bayesian probability can be applied in both stochastic and ignorance types of uncertainties. A probabilistic analysis requires that an analyst has information on the probability of all events. Whenever this information is unavailable, the uniform distribution function is often used, which is justified by Laplace's principle of insufficient reason [16]. Measurements of uncertainty almost exclusively investigated in terms of disjunctive variables. A disjunctive variable has a single value at any given time, but is often tentative because of limited evidence.

3 Methodology

Development model is divided into two section, ANFIS inference and Bayesian inference.

3.1 ANFIS inference

Work stages of uncertainty representation for modeling using ANFIS inference are illustrated in Fig. 1. Later, the details are explained in further subsections.

3.1.1 Load data

Data loading is about assigning the data set for training, testing, and checking. There are five sets of data for inputs and one set of data for output observed. Ten different datasets were randomly selected from 624 dataset to assign for training, testing, and checking, hence 384 dataset assigned for training, 120 for testing, and 120 for checking to make sure the majority of data sets are trained. The best result is to have the lowest training error.

3.1.2 Clustering

Clustering stage is the initial step of 'Fuzzification' in the FIS. The inputs were 'Fuzzified' after all numerical values of input uncertainties and output was loaded. The propagation of each uncertainty was broken into the different clusters of Fuzzy to see the behavior of uncertainties on the production throughput. Clustering includes selection of membership function and definition of Linguistic value. Fuzzy Logic Toolbox in MATLAB software was used for clustering Both Grid partitioning and subtractive clustering.

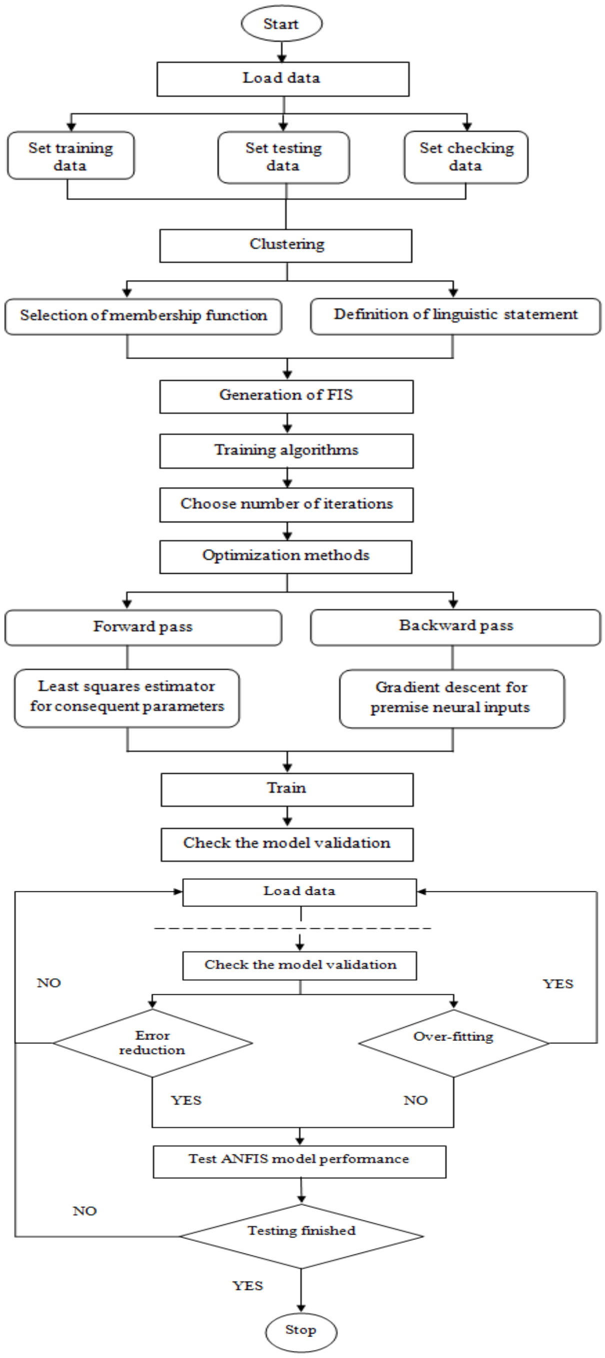

Fig. 1 Flow diagram of computations in ANFIS inference

3.1.3 Selection of membership function

For selecting the best type of membership function, different membership functions as nonlinear functions were considered for the input uncertainties and multi linear functions for the output in the Sugeno Fuzzy inference system (SFIS). SFIS is more accurate than Mamdani FIS [21]. Three of the most popular membership functions, which are widely used, namely, triangular, trapezoidal, and Gaussian were examined.

3.1.4 Definition of linguistic value

Two sets of Linguistic values were defined with respect to number of membership function to determine the quality and quantity of membership. Three Linguistic values were first examined for each uncertainty. The quality of three Linguistic values was defined as low, medium, and critical. The second set of Linguistic values was examined by five statements: very low, low, medium, high, and very high.

3.1.5 Generation of FIS

FIS was presented as black box diagram that has three parts: inputs uncertainties (defined as nonlinear), the output (production throughput prediction), and Fuzzy inference engine. Generating rules inferred the relationship between inputs and output. Subtractive clustering selected the optimal number of rules with the lower training error. The number of rules was found through the equation (1) until (4) [22].

$$Di = \sum_{j=1}^{N} \exp(-(\frac{2}{r_a})^2 \times \|xi - xj\|^2) \qquad (1)$$

where
D_i = centre of cluster i
N = data points
r_a = constant value

The first cluster was identified by the highest density measure (D_1*), which was at the centre of the cluster.

$$D_i = D_i - D_1^* \times \mu\,(x_i^*) \qquad (2)$$

$$\mu\,(x_i^*) = \exp\,(-\frac{\|x_i - x_i^*\|^2}{\left(\frac{r_b}{2}\right)^2}) \qquad (3)$$

Where r_b is a positive constant and it is greater than r_a according to [23]. A sufficient number of cluster centers were generated by repeating the same process for other clusters and revising the density measures. Gaussian membership functions determined weightage of each rule i for input variable j, as the polynomial function moves between 0 and 1. This approach presented accurate relationship between response and inputs by generating the optimal rules.

$$\mu_{ij}(x_i) = \exp\left(\frac{(x_i - x_i^*)}{\left(\frac{r_a}{2}\right)^2}\right) \quad (4)$$

The subtractive clustering parameters in ANFIS are the squash factor; accept ratio, reject ratio, and range of influence. Squash factor is used to multiply the radii values that determine the neighborhood of a cluster centre. The purpose is to squash the potential for outlying points to be considered as part of that cluster. The value of squash factor considered for clustering was 1.25 and the accept ratio was set at 0.5. The accept ratio is a fraction of the potential of the first cluster centre. Range of influence of cluster was set at 0.5. The reject ratio is a fraction of the potential of the first cluster centre, and was defined at 0.15

3.1.6 Training algorithms

The training adjusted the membership function parameters and displayed the error plots. ANN was utilized for training, testing, and checking for each uncertainty. Back propagation gradient descent and the least square of error are two optimization methods for training the generated FIS. Gradient error back propagation adjusts the Fuzzy sets coefficients while the least squares of error adjust the parameters of consequent polynomial function.

Hybrid learning algorithm includes both and was employed for identifying linear and nonlinear parameters.

3.1.7 Number of iteration

The number of iteration was selected to do the training process through the hybrid learning algorithm. Four different simulations, which are called epochs in ANFIS, were performed for each randomly assigned data set in order to achieve the lower training error. Training was started by 50 simulations then increased to 100, 150, and 200 to see if there was any possibility to more error reduction, and make sure the error not increasing and no overfitting. The training process was stopped when the maximum epoch number was reached.

3.1.8 Training

Training process was implemented in MATLAB software. The theory of the training process is described step by step with relevant equations.

Input node layer

In step 1, the output of five uncertainties is denoted by O.

$$O_i = \mu_i(D) \quad (5)$$
$$O_i = \mu_i(L) \quad (6)$$
$$O_i = \mu_i(Se) \quad (7)$$
$$O_i = \mu_i(S) \quad (8)$$
$$O_i = \mu_i(B) \quad (9)$$

$$\mu_i(U) = \frac{1}{1 + \left[\left(\frac{B - c_i}{a_i}\right)\right] b_i} \quad (10)$$

where
D = Demand,
L= lead time of manufacturing,
Se = Setup time,
S = Scrap,
B = Breakdown time
O_i = Output of cluster i,
i = 1,..,5,
μ = Membership function,
U = Uncertainty.

Rule nodes (inference layer or rule layer)

The weight of each cluster is found in step 2. The output of each input was obtained from step 1 and multiplies to other factors as shown in equation (11).

$$O_i = W_i = \mu_i(B) \times \mu_i(D) \times \mu_i(L) \times \mu_i(Se) \times \mu_i(S) \quad (11)$$

where
W_i = weight of cluster i.

Normalized layer (Average nodes layer)

Defuzzification method was done through the weighted average in step 3. The output i is the ratio of the weight of cluster i to the summation of all weights as shown in equation (12).

$$O_i = \overline{W}_i = \frac{W_i}{\sum W_i} \quad (12)$$

Consequent nodes layer (aggregation layer)

$\overline{W}_i$ is multiplied by the output of the cluster i in the step 4 as presented in equation (13).

$$O_i = \overline{W}_i \times F_i \quad (13)$$

where
F_i = the output of the cluster i.

Total output layer

In the step 5, the overall output as the summation of all incoming signals is computed by equation (14).

$$O_i = F = \frac{\sum_{i=1}^{5} W_i F_i}{\sum_{i=1}^{5} W_i} \quad (14)$$

3.1.9 Check the model validation
Model validation was done by overfitting and reducing the training error. The overfitting was determined by number of error plots during training. This was done by testing the trained FIS on the training data against the checking data. If the checking error is decreased the model does not have overfitting and it is valid.

3.1.10 Testing and checking datasets for validation
The test data and check data were plotted against the FIS output to validate the forecasted data was near to actual data.

3.2 Bayesian inference
Bayesian inference use distribution-based approach where the prior probabilities were utilized to quantify uncertainty regarding the occurrences of events. Stages of uncertainty are illustrated in Fig. 2.

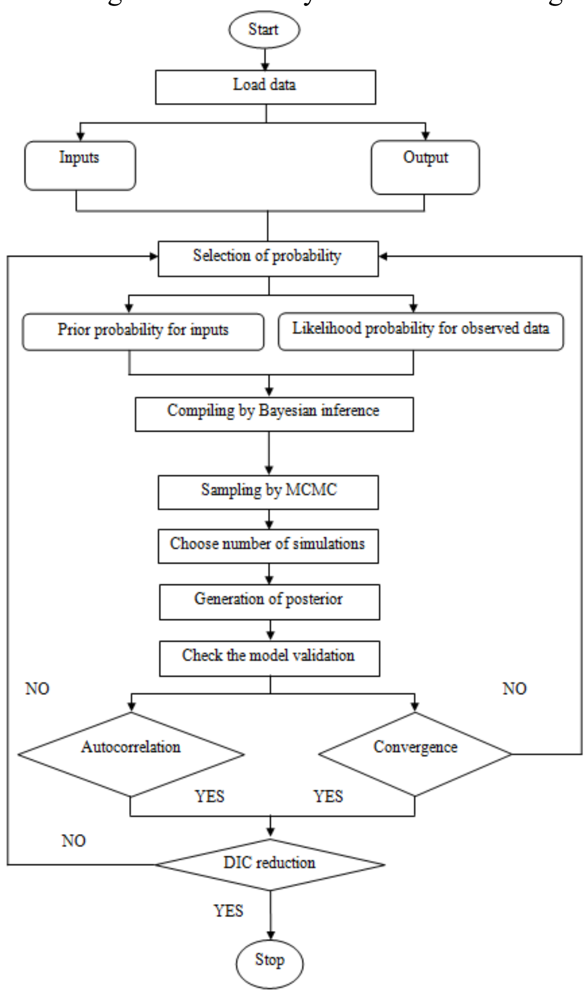

Fig. 2 Flow diagram of computations in Bayesian inference

3.2.1 Load data
The data observed for input uncertainties and throughput of production was translated to the BUGS language by inserting them into the R software. The translated data was loaded by importing them to the model programmed in BUGS. A list from a vector of output and a vector for each uncertain variable was developed by using a command for reading the data.

3.2.2 Selection of probability
Problem formulation with predefined probability levels explicitly considered the stochastic property of the uncertainties. The selection of probability was divided into prior distribution of inputs and likelihood probability for observed data. These two probability selections were two main input components of Bayesian inference.

3.2.3 Prior distribution
Prior distribution refers to the historical behavior of the inputs. Its selection for inputs is done before observing the data. This behavior can be elicited from the experts [14]. The distribution of prior usually is defined in question by the normal distribution with mean of zero and low variance. Unfortunately, as the propagation of uncertainty may change with time, the prior information on the inputs cannot assume true. Therefore, the determination of prior probability distribution is done by the trial and error method.

BUGS can modify the approximate prior by considering the sum of Gaussians cantered on each sample generated. The selection of prior probability distribution to express the uncertainty propagation of inputs can be examined with different distribution to see which one is more accurate based on lower error generated.

One way to compare the models with different probability distributions is to use a criterion based on trade-off between the fit of data to the model and the corresponding complexity of the model. A Bayesian model [24] was proposed to compare criterion based on deviance information criterion (DIC). For each uncertain variable, three popular probability distributions were examined: uniform, exponential and normal. The posterior probability distribution function of the model parameters was computed from the defined prior probability distribution function. The best prior probability distribution was based on lower DIC comparison.

3.2.4 Likelihood
The purpose of selecting likelihood probability distribution is to identify the best probability function which can fit the observed data. The likelihood function for production throughput was computed using the conditional distributions given the data observed in a tile industry. The probability distributions of normal, exponential, Weibull, and logistic function were tested. The procedure was to maximize the likelihood to fit the data better. Dependencies values between variables were also identified through the conditional probabilities. It was gained by integrating the unknown parameters through the equations (15) and (16).

$$p(\hat{y}|y) = \int p(\hat{y}|x)\, p(x|y)\, dx \quad (15)$$

$\Rightarrow$ for normal distribution = $p(\hat{y}|y)$ =

$$\int \frac{1}{\sigma\sqrt{2\pi}}\, e^{\left(-\frac{1}{2\sigma^2}(\hat{y}-x)^2\right)} \times \frac{1}{\acute{\sigma}\sqrt{2\pi}}\, e^{\left(-\frac{1}{2\acute{\sigma}^2}(x-\acute{\mu})^2\right)}\, dx \quad (16)$$

where
$\hat{y}$ = future observation,
y = observed at given x.

3.2.5 Compilation
The compilation process utilizes both prior and likelihood. It synchronizes the information about the uncertainty before observation and the behavior of data after observation. The compiling is to multiply the prior distribution and likelihood probability.

3.2.6 Sampling
Various samplings were computed from the joint posterior distribution. Markov chain method is used to obtain sample from full conditional distributions. A vector of unknown parameter was considered to consist of *n* subcomponents. Then the sampling started choosing the value of unknown parameters from the conditional distribution to find the best value of the beta for the posterior distribution, where the posterior distribution was maximized. Gibbs sampling algorithm was utilized because it is the robust procedure of MCMC. The Gibbs sampling algorithm approximated the posterior distribution function by making random draws from the probability distributions of the input uncertainties and evaluating the model at the resulting values.

3.2.7 Quantity of simulations
Five simulation runs of 1000, 5000, 8000, and 10000 for drawing samples were examined to test the model based on DIC. Simulation started from 1000 and was increased until it reached convergence. The amount optimal simulation run was determined by lower value of convergence and DIC.

3.2.8 Generation of posterior
The posterior is the product of observation probability (likelihood) and previous information (prior). Different samplings were performed to generate posterior of unknown parameters. Each kernel of the generated sample had weightage in term of closeness to the posterior. Kernel is a function of the sample variance. Closer kernels dominated the posterior. Final posterior was obtained by weight-normalizing of sum of kernel products, which had the best posterior mean and variance.

Fig 3 showed a construction of Bayesian black box diagram. A processor of Bayesian inference engine including rules of probabilities and Bayesian theory to derive the posterior mean and variance of the model is at the centre of the diagram.

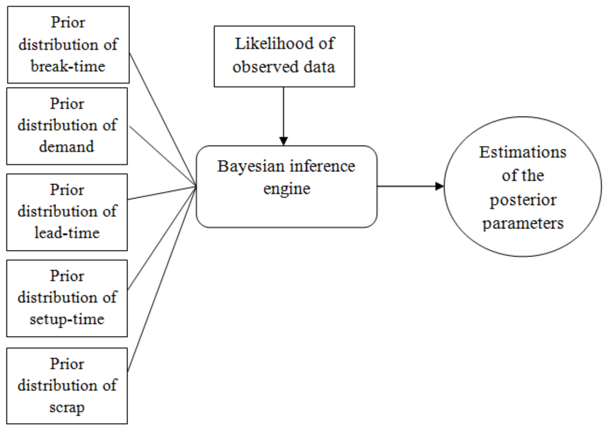

Fig. 3 The construction of Bayesian inference model

Bayesian inference engine used the Bayes factor (BF) to analyze the model proposed as shown in equation (19). Two different sets of prior uncertainty were assigned for each uncertain variable. Two competing models were generated into two chains denoted by M1 and M2 as in equation (17). The data observed for each uncertainty was denoted by X. The posterior was found through the equation (18).

$$M_1: f_1(x \mid \beta_i) \text{ and } M_2: f_2(x \mid \beta_i) \quad (17)$$

$$\frac{\pi(M_1 \mid x)}{\pi(M_2 \mid x)} = \frac{p(M_1)/p(x)}{p(M_2)/p(x)} \times \frac{\int f_1(x \mid \beta_1) p_1(\beta_1) d\beta_1}{\int f_2(x \mid \beta_2) p_2(\beta_2) d\beta_2} \quad (18)$$

$$BF(x) = \frac{\pi(M_1 \mid x)/\, p(M_1)}{\pi(M_2 \mid x)/\, p(M_2)} \quad (19)$$

When the M_1 is as the null model, the possibilities of BF results are as follows.

If BF(x) ≥ 1 => M_1 supported,
If $1 >$ BF(x) $\geq 10^{-1/2}$ => minimal evidence faced for M_1,
If $10^{-1/2} >$ BF(x) $\geq 10^{-1}$ => substantial evidence faced for t M_1,
If $10^{-1} >$ BF(x) $\geq 10^{-2}$ => strong evidence faced for M_1,
If $10^{-2} >$ BF(x) => decisive evidence faced for M_1.

The error of Monte Carlo (MC) for sampling procedures was calculated for each uncertain parameter by equation (20).

$$\text{MC error} = \frac{\text{SD}}{\sqrt[2]{\text{Number of iterations}}} \quad (20)$$

3.2.9 Check the model validation

The model validation was checked firstly through two ways of checking. First checking was by visual inspection of trace/history plots to see if the model is convergence. The model convergence was achieved when the chains were overlapping. The second way of checking was to check the autocorrelation. The convergence graphically presents the distribution of uncertainty. Gelman Rubin statistic (GRS) showed the convergence ratio [25]. The autocorrelation is defined between zero and one. A slow convergence shows the high autocorrelation, indicating validity of model.

3.3 Models comparison

Many authors [26-29] used Mean Absolute Percentage Error (MAPE) and Correlation to compare forecasting models. They measure the accuracy of fitted time series values. MAPE expresses error as a percentage, which is the average of the absolute of the difference between actual and forecasted divided over actual. It is used to measure within sample goodness-of-fit and out-of-sample forecast performance. The value of MAPE is computed by equation (21).

$$\text{MAPE} = \frac{1}{n}\left(\sum_{t=1}^{n}\left|\frac{(\text{Actual}-\text{Forecasted})}{\text{Actual}}\right|\right) \times 100 \quad (21)$$

Correlation criterion showed an association between the fitted value and the actual value. Absolute correlation value nearing to 1.0 implies high accuracy while absolute correlation value greater than 0.8 is considered as strong relationship [31]. The correlation value is calculated through the equation (22) [30]:

$$\text{Correlation} = \frac{\text{Covariance (Actual and Forecasted)}}{\sigma_{\text{actual}} \times \sigma_{\text{forecasted}}} \quad (22)$$

where
Covariance of actual and forecasted

$$= \sum_{t=1}^{n}\frac{(\text{Actual}-\overline{\text{Actual}})\,(\text{Forecasted}-\overline{\text{Forecasted}})}{n-1} \quad (23)$$

4 Results

We will comment on the experimental results.

4.1 Membership function

The most suitable and efficient membership functions for defining the propagation of uncertainties was found with the lowest training error, which was Gaussian membership function. Fig. 4 and fig 5 showed the Gaussian membership functions of each uncertainty in SFIS with five clusters for breakdown time and demand respectively. The Gaussian membership functions for lead time, setup time and scrap have similar trend as break down time and demand.

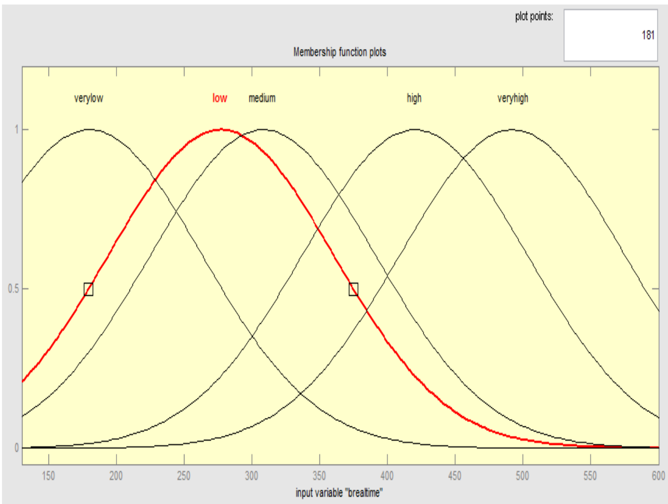

Fig. 4 Fuzzy membership function of breakdown time

The propagation of breakdown time was presented in five values of linguistic variables and corresponding membership functions as follows:

$B_{\text{very low}} \sim N\,(180, 83.08)$
$B_{\text{low}} \sim N\,(277, 83.09)$

$B_{\text{medium}} \sim N$ (308, 83.08)
$B_{\text{high}} \sim N$ (420, 83.08)
$B_{\text{very high}} \sim N$ (492, 83.08)

where
N represents Gaussian membership function

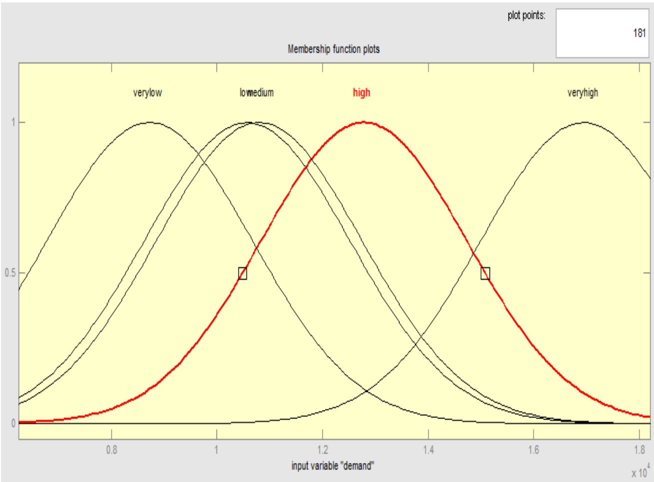

Fig. 5 Fuzzy membership function of demand

The propagation of demand was presented in five values of linguistic variables and corresponding membership functions as follows:

$D_{\text{very low}} \sim N$ (8725, 1953)
$D_{\text{low}} \sim N$ (10570, 1953)
$D_{\text{medium}} \sim N$ (10800, 1953)
$D_{\text{high}} \sim N$ (12780, 1953)
$D_{\text{very high}} \sim N$ (16950, 1953)

Similarly, the propagation of lead time was shown in five values of linguistic variables and corresponding membership functions as follows:

$L_{\text{very low}} \sim N$ (5200, 338.5)
$L_{\text{low}} \sim N$ (5718, 338.5)
$L_{\text{medium}} \sim N$ (5782, 338.5)
$L_{\text{high}} \sim N$ (6028, 338.5)
$L_{\text{very high}} \sim N$ (6720, 339)

Similarly set up time propagation was presented in five values of linguistic variables and corresponding membership functions as follows:

$Se_{\text{very low}} \sim N$ (190, 12.37)
$Se_{\text{low}} \sim N$ (218, 12.36)
$Se_{\text{medium}} \sim N$ (230, 12.39)
$Se_{\text{high}} \sim N$ (240, 12.38)
$Se_{\text{very high}} \sim N$ (242, 12.37)

Similarly scrap propagation was presented in five values of linguistic variables and corresponding membership functions as follows:

$S_{\text{very low}} \sim N$ (1800, 535.1)
$S_{\text{low}} \sim N$ (2650, 535.1)
$S_{\text{medium}} \sim N$ (3020, 535.1)
$S_{\text{high}} \sim N$ (3420, 535.1)
$S_{\text{very high}} \sim N$ (3800, 535.1)

4.2 ANFIS model structure

Figure 9 exhibited the generated FIS whereby a processor of SFIS to elaborate five Fuzzy rules is located at the centre of the Fuzzy black box diagram

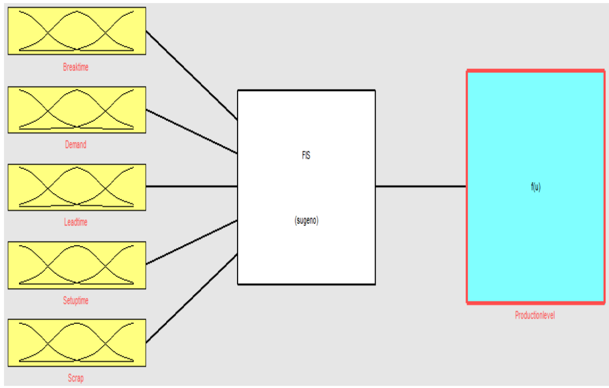

Fig. 6 Constructed Fuzzy model with five inputs and one output

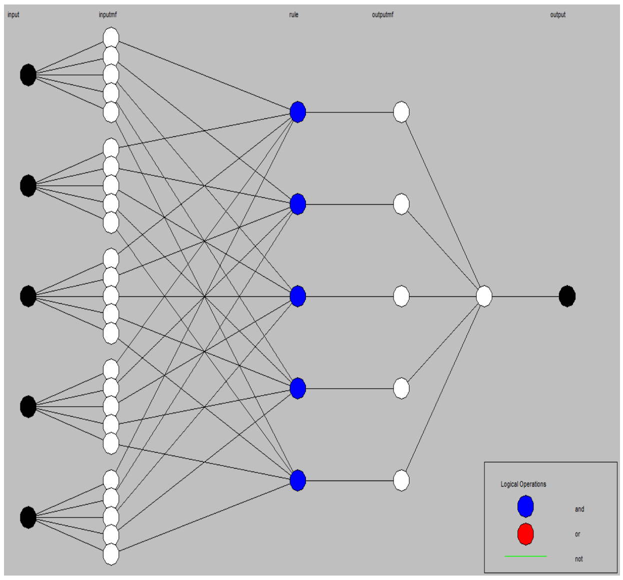

Fig. 7 ANFIS model structure with five rules

The ANFIS model was structured by five rules. The model divided the five uncertainties space into the Fuzzy subspaces and also structured the polynomial function of throughput response using five linear functions. The five uncertainties defined by the Gaussian membership functions were inserted to the ANFIS model. Fig. 7 showed the ANFIS model structure.

4.3 Rules

Five "if-then" rules were extracted to represent how to achieve to the different levels of production throughput. Rule 1: If breakdown time, setup time and scrap fall in low cluster while lead time falls in medium and demand is high; then production level will be high, Rule 2: If breakdown time, scrap, lead time and setup time fall in high cluster while demand is very low; then level of production is low, Rule 3: If lead time and demand are in low cluster while breakdown time, setup time and scrap fall in medium cluster; then level of production will be medium, Rule 4: If breakdown time, setup time, scrap and lead time fall in very low cluster while demand is very high; then production level will be very high, Rule 5: If breakdown time, scrap and setup time fall in very high cluster while lead time falls in high cluster and demand is medium; then production level will be very low.

Table 1 Estimated parameters of each uncertainty

Uncertainties	Clusters	σ	μ
Breakdown time	Very low	9.11	180
	Low	9.11	277
	Medium	9.11	308
	High	9.11	420
	Very high	9.11	492
Demand	Very low	44.19	8725
	Low	44.19	10570
	Medium	44.19	10800
	High	44.19	12780
	Very high	44.19	16950
Lead time	Very low	18.39	5200
	Low	18.39	5718
	Medium	18.39	5782
	High	18.39	6028
	Very high	18.41	6720
Setup time	Very low	3.51	190
	Low	3.51	218
	Medium	3.52	230
	High	3.51	240
	Very high	3.51	242
Scrap	Very low	23.13	1800
	Low	23.13	2650
	Medium	23.13	3020
	High	23.13	3420
	Very high	23.13	3800

4.4 Mean and standard deviation estimation of parameters

The parameters including the mean and standard deviation of each uncertainty were tabulated in Table 1 with respect to their clusters that were expressed in the rules section. The membership functions for all the parameters are Gaussian.

4.5 Training error

Trend of training error is shown in Fig. 8.The figure indicates no overfitting during the training process with testing trend and the error rate was reducing. This showed that the combination of the least squares method and back propagation gradient descent method used for training FIS membership function parameters generated lower training error. For example, the error trend of training performed for 200 iterations presented in Fig. 8 indicated low error in training.

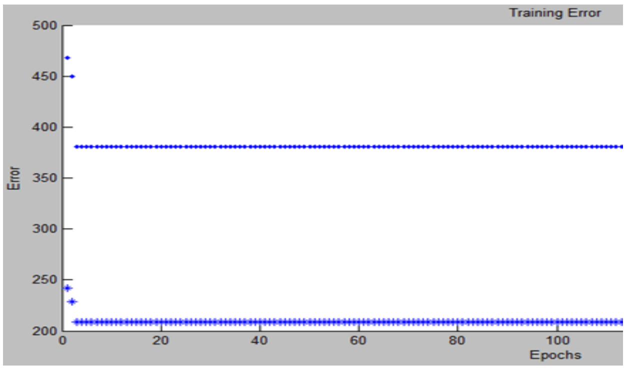

Fig. 8 Error trend for training

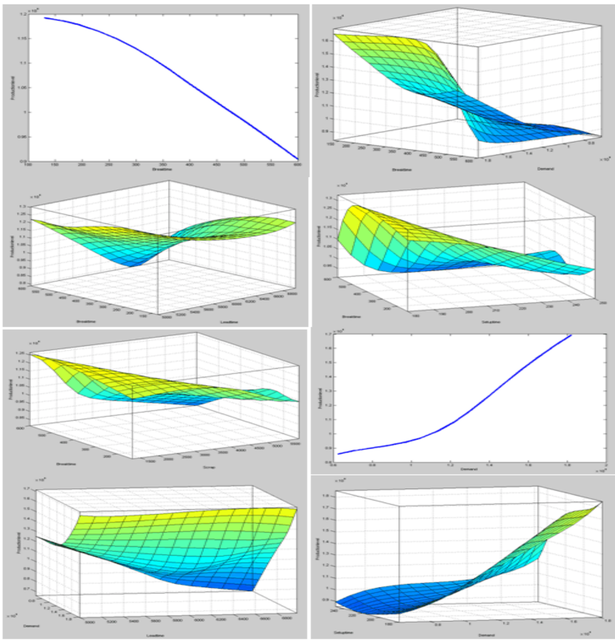

Fig. 9 Nonlinear relationship between uncertain variables and throughput in 2-D and 3-D diagram

4.6 Uncertainties and Throughput Relationships

Based on the extracted rules, the nonlinear relationship between the uncertainties and response were identified. Fig. 9 showed some of the effects of inputs on response.

4.7 Coefficients estimation of parameters

The coefficients of Sugeno Fuzzy inference linear functions (SFILF) after final output results were computed and shown in table 2.

Table 2 Estimated coefficients of SFILF

Clusters	Inputs' coefficients of SFILF					
	β_0	β_1	β_2	β_3	β_4	β_5
Very low	35170	-8.16	0.01785	-3.441	2.755	-0.4042
Low	7058	-5.091	0.8506	-1.05	10.99	0.2962
Medium	42140	-9.371	0.0434	-4.045	-19.79	-0.5418
High	1157	3.393	0.7026	0.7061	-13.47	0.2097
Very high	-12110	-2.319	1.174	0.9599	19.36	0.6033

The estimated coefficients of the five uncertainties were inserted into the model as presented in (24).

$$P(t) \sim \beta_0 + \beta_1 B(t) + \beta_2 D(t) + \beta_3 L(t) + \beta_4 Se(t) + \beta_5 S(t) \quad (24)$$

where
P(t) = Production throughput (level) over the time,
B(t) = Breakdown time,
D(t) = Demand volume over the time,
L(t) = Lead time of manufacturing,
Se(t) = Setup time,
S(t) = Scrap volume over the time,
β_0 = Intercept,
$\beta_1, \ldots, \beta_5$= Coefficient of inputs.

The five SFILF were formulated for all clusters as shown in (25) until (29).

$$P_1(t) \sim 35170 - 8.16\, B(t) + 0.018 D(t) - 3.441\, L(t) + 2.755\, Se(t) - 0.404\, S(t) \quad (25)$$

$$P_2(t) \sim 7058 - 5.091\, B(t) + 0.851\, D(t) - 1.05\, L(t) + 10.99\, Se(t) + 0.296\, S(t) \quad (26)$$

$$P_3(t) \sim 42140 - 9.371\, B(t) + 0.043 D(t) - 4.045\, L(t) - 19.79\, Se(t) - 0.542 S(t) \quad (27)$$

$$P_4(t) \sim 1157 + 3.393\, B(t) + 0.703\, D(t) + 0.7061\, L(t) - 13.47\, Se(t) + 0.210\, S(t) \quad (28)$$

$$P_5(t) \sim -12110 - 2.319\, B(t) + 1.174\, D(t) + 0.9599\, L(t) + 19.36\, Se(t) + 0.603\, S(t) \quad (29)$$

4.8 Rule viewer

The rule viewer was performed to expose all parts of the Fuzzy inference process from inputs to output. Each row of plots corresponds to one rule, and each column of plots corresponds to either an input variable or an output variable.

4.9 Model programmed in BUGS

Table 3 described Table 3 Description of the BUGS model expressions. The sign ~ indicates a stochastic relationship, where Tau =1/variance showed precision level. The c function combines objects into a vector, where the variable x was collected by different values that were measured in different period of time.

Table 3 Description of the BUGS model expressions

Expression	Type	Usage
dnorm	Normal distribution	x ~ dnorm (mu, tau)
c	Vector of data set	$x = c\,(x_1, x_2, \ldots, x_n)$

4.10 Probability distribution test

Four popular probability distributions including normal, Weibull, logistic, and exponential were tested. Fig.10 showed the normal distribution is the best fit for production throughput while Fig. 11 showed the summary of the normal distribution function.

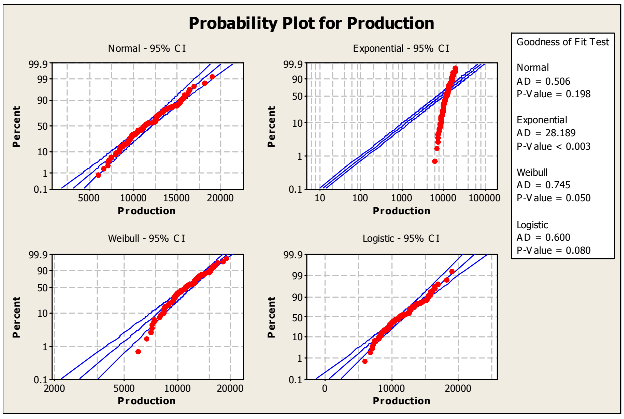

Fig.10 Testing four popular probability distributions

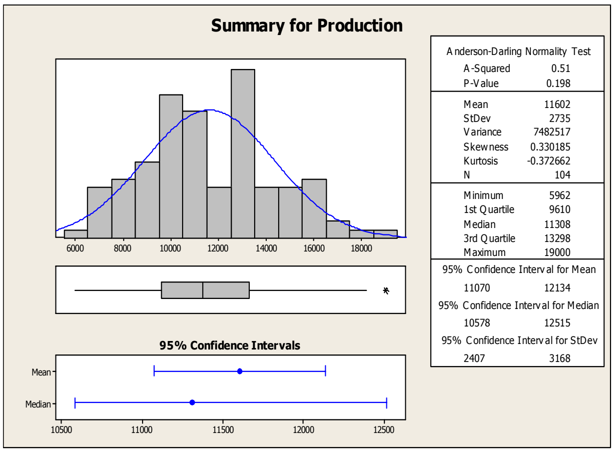

Fig. 11 Anderson-Darling normality test

4.11 Checking the programmed model

After programming, the model was checked for any completeness and consistency with the data. The initial values were generated by sampling from the prior. The model programmed was proven syntactically correct and compiled.

4.12 Convergence diagnostics test

Computational results of the lowest MAPE were selected in this section for the Bayesian model. The convergence diagnostics were checked through two chains results. The convergence was achieved because both chains overlapped each other, according to [25]. The dynamic race plots of the stochastic parameters with 10,000 iterations were done to check the convergence on 95% credible interval. Fig. 12 graphically showed the results.

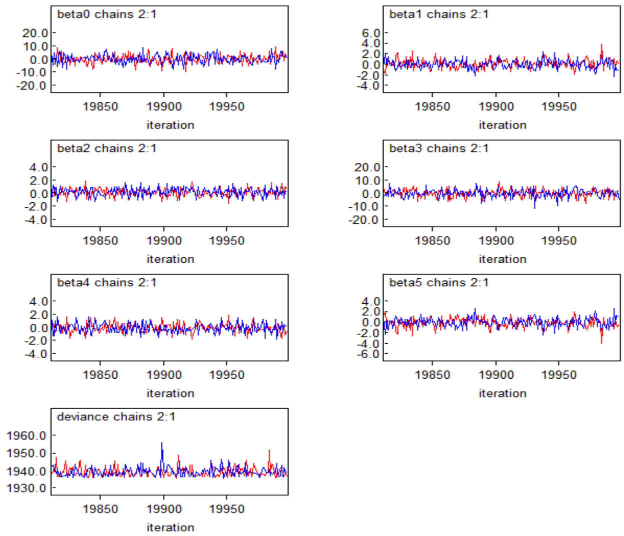

Fig. 12 Dynamic trace plots of uncertain parameters

DIC is the summation of goodness of fit and complexity. Deviance is the average of the log likelihoods calculated at the end of iteration in Gibbs Sampler. The definition of deviance is - 2 × log (likelihood). Likelihood is defined as p (y|theta), where y comprises all stochastic parameters given values and theta comprises the stochastic parents of y - 'stochastic parents' are the stochastic parameters upon which the distribution of y depends, when collapsing over all logical relationships.

4.13 Kernel density

Fig. 13 showed the value of Kernel density for each stochastic parameter was performed on 10000 samples. The diagrams indicated smoothed kernel density estimate. The trends indicated the posterior distribution of each stochastic parameter is normal like prior distribution, thus proving the estimations were robust and logical.

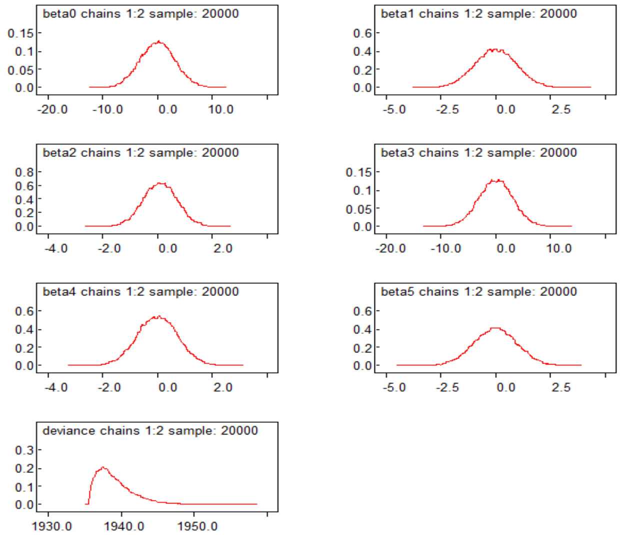

Fig. 13 Kernel density of the uncertain parameters

4.14 Running quartiles

Running quantiles plot out the running was done for mean with running 95% confidence intervals where 10000 iterations were used. Results are presented in Fig. 14.

4.15 Bivariate posterior

"Bivariate posterior scatter plots" present the correlation between two stochastic parameters. For example, the Fig. 15 shows correlation between (β_5) and (β_2).

4.16 Pair-wise correlations

Table 4 exhibited the calculated values of pair-wise correlations of all parameters. The highest correlation value was between beta2 and beta5 while its lowest value was between beta0 and beta3.

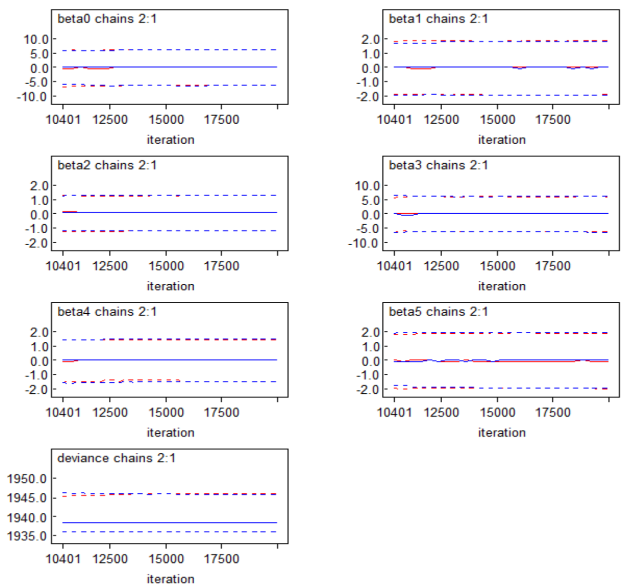

Fig. 14 Running mean of the uncertain parameters

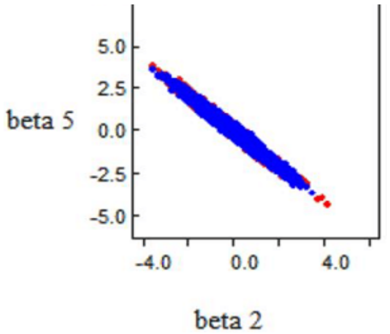

Fig. 15 Pairwise correlation of (β_5) and (β_2)

Table 4 Pairwise correlations of all inputs

Variables		Correlation values
beta0	beta1	-0.00705918
beta0	beta2	-1.65774E-4
beta0	beta3	2.87397E-5
beta0	beta4	-0.00629102
beta0	beta5	0.00675832
beta1	beta2	0.319391
beta1	beta3	-0.384504
beta1	beta4	-0.00271504
beta1	beta5	-0.0436661
beta2	beta3	-0.873166
beta2	beta4	-0.0208795
beta2	beta5	0.592831
beta3	beta4	-0.155095
beta3	beta5	-0.817384
beta4	beta5	-0.0179657

4.17 Autocorrelation function

The autocorrelation function for the chain of each parameter indicated the dimensions of the posterior distribution were mixing slowly before 20 lags in each case. Slow mixing is often associated with high posterior correlations between parameters.

4.18 Gelman Rubin statistics

Gelman Rubin statistic (GRS) was performed for all stochastic parameters, which were modified by [25] in equation (30). The idea was to generate the multiple chains starting at overdispersed initial values, and assesses the convergence by comparing within-chain and between-chain variability over the second half of those chains.

$$GRS = A / W \quad (30)$$

where
A= width of the empirical credible interval based on samples pooled together (2 chains × 10000 iterations).
W= width average of the intervals across the two chains

The GRS is to average the interval widths (shown in red color). It should be 1 if the starting values are suitably overdispersed and the convergence is approached. The blue and green interval lines should be approximately stabilized to constant value (not necessarily 1). It is proven and shown for all five stochastic parameters in Fig. 16.

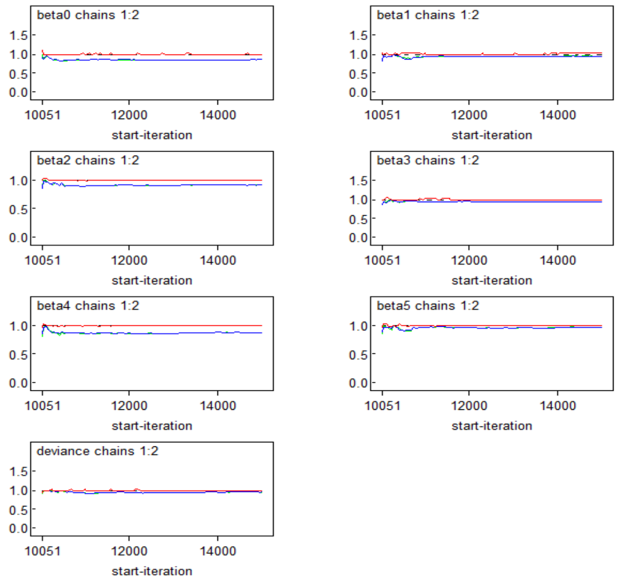

Fig. 16 Gelman Rubin statistic for the uncertain parameters

Where
Green = width of 80% intervals of pooled chains: should be stable
Blue = average width of 80% intervals for chains: should be stable
Red = ratio of pooled/within: should be near 1

4.19 Box plot of posterior

Box plot of posterior efficiency distributions were presented in Fig. 17. The calculated baseline value was 11595.7809089724.

4.20 Model fit

Fitted values were compared with actual values in 95% interval for production output, breakdown, demand lead time, setup time and scrap was calculated and plotted. The results showed production throughput and demand had similar upward trend while breakdown time, lead time, set up time and scrap were having similar downward trend. Fig. 18 showed comparison between fitted value to actual value for production throughput, while Figure 23 showed the similar comparison for breakdown time.

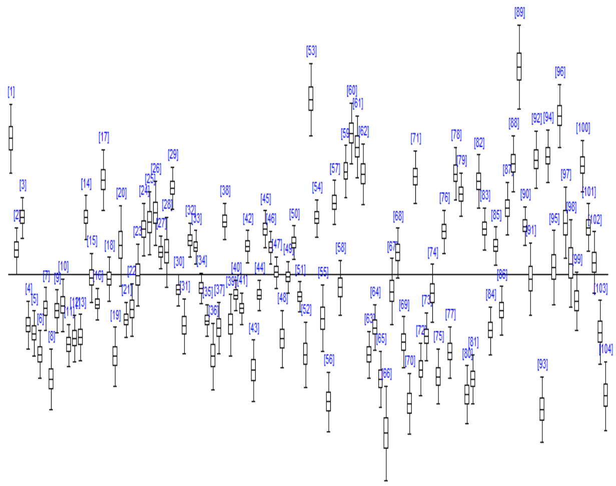

Fig. 17 Box plot of posterior efficiency distributions

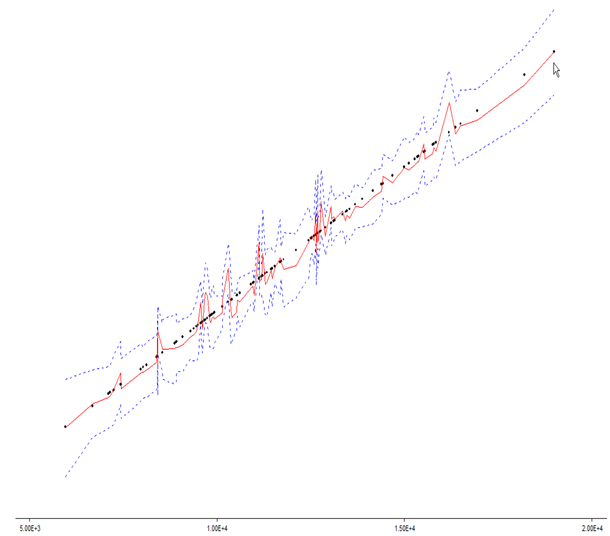

Fig. 18 Fitted value compare with actual values over production throughput observed with 95 % interval

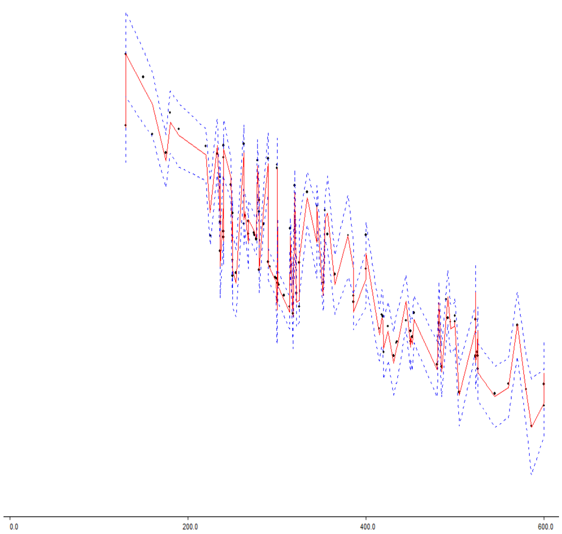

Fig. 19 Fitted value compare with actual values over breakdown time observed with 95 % interval

where
Red = posterior mean of μ_i,
Blue = 95% interval,
Black dot = observed data

4.21 Posterior estimates

The final set of posterior estimates using Gibbs sampling in 95% credible interval was summarized in Table 5. The percentiles of 2.5% and 97.5% of posterior estimates produce an interval, which the parameter lies with probability of 0.95.

Table 5 Summaries of the posterior distribution

Coefficient	mean	Std. Dev.	MC error	median
β_0	0.01343	3.179	0.0242	0.02376
β_1	-0.0849	2.896	0.01872	-0.1016
β_2	0.9585	0.1596	0.001056	0.958
β_3	0.1268	0.6618	0.004444	0.1246
β_4	-0.0458	3.156	0.02213	-0.0614
β_5	-0.1481	0.7179	0.005325	-0.1474
Deviance	1939.0	2.383	0.01624	1939.0

The value of MC error shows an estimate of (σ / $\sqrt[2]{N}$). The batch means method outlined by [32] was used to estimate σ.

Finally, the Bayesian model is formulated as in equation (31).

$$P(t) \sim 0.01343 - 0.0849\,B(t) + 0.9585\,D(t) + 0.1268\,L(t) - 0.04589\,Se(t) - 0.1481\,S(t) \quad (31)$$

4.22 Comparison

The forecasting accuracy was calculated using Pearson correlation and the MAPE for both Bayesian and ANFIS models. MAPE index used to compare the performance of ANFIS and Bayesian models. The values in Table 6 indicated that the ANFIS model significantly yield a better fit than the Bayesian Model for production level under the five uncertain variables

Table 6 Comparison of ANFIS and Bayesian models

Model	MAPE	Pearson correlation
Bayesian	0.0261403	0.989
ANFIS	0.0223005	0.991

To achieve higher production throughput level for the case study using the ANFIS model, the coefficients of the production uncertainties were indicated; where for breakdown time was -2.319, for demand = 1.174, for manufacturing lead time = 0.9599, for setup time = 19.36, and for scrap = 0.6033. The lower and upper limits for very high level of production throughput were identified as follows.

- Breakdown time should fall between 474.14 and 509.85,
- Demand between 16863.38 and 17036.61,
- Lead time between 6683.92 and 6756.08
- Setup time between 235.12 and 248.87
- Scrap between 3754.66 and 3845.33.

5 Conclusion

This study found that the application of the Bayesian and ANFIS inferences on detecting the production uncertainties and their impacts on the production throughput level as more viable and accurate than classical approach. ANFIS model was proven as more efficient and provides better production forecasting accuracy compared with Bayesian model. Hence, ANFIS model is recommended to be used for production estimation under random uncertainties.

Different combinations in terms of number of simulations, types of membership functions for the ANFIS model, and different prior distributions for stochastic variables in the Bayesian model were also examined and found to be viable.

200 epochs were found to be the best iterations number in the case study and the best membership function was the Gaussian in SFIS for the ANFIS model. The best simulations iterations of MCMC were 10000 and the best prior distributions for stochastic variables were normal distributions for the Bayesian model.

References:

[1] 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, 2005, pp. 293-308.

[2] J. Li, D. E. Blumenfeld, J. M. Alden, Comparisons of two-machine line models in throughput analysis, *International Journal of Production Research*, vol. 44, 2006, pp. 1375-1398.

[3] J. Li, D. E. Blumenfeld, N. Huang, J. M. Alden, Throughput analysis of production systems: recent advances and future topics, *International Journal of Production Research*, vol. 47, 2009, pp. 3823-3851.

[4] J. Mula, R. Poler, J. Garcia-Sabater, F. Lario, Models for production planning under uncertainty: A review, *International Journal of Production Economics*, vol. 103, 2006, pp. 271-285.

[5] A. M., Deif and H. A. ElMaraghy, Modelling and analysis of dynamic capacity complexity in multi-stage production, *Production Planning and Control*, vol. 20, 2009, pp. 737-749.

[6] H. Tempelmeier, Practical considerations in the optimization of flow production systems, *International Journal of Production Research*, vol. 41, 2003, pp. 149-170.

[7] M. S. Han and D. J. Park, Optimal buffer allocation of serial production lines with quality inspection machines, *Computers & Industrial Engineering*, vol. 42, 2002, pp. 75-89.

[8] 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.

[9] S. Koh, A. Gunasekaran, S. Saad, A business model for uncertainty management, *Benchmarking: An International Journal*, vol. 12, 2005, pp. 383-400.

[10] 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*, Gronegen, the Netherlands, 2008.

[11] P. Kouvelis and J. Li, Flexible Backup Supply and the Management of Lead Time Uncertainty, *Production and Operations Management*, vol. 17, 2008, pp. 184-199.

[12] K. R. Baker and S. G. Powell, A predictive model for the throughput of simple assembly systems, *European journal of operational research*, vol. 81, 1995, pp. 336-345.

[13] Z. Hoque, A contingency model of the association between strategy, environmental uncertainty and performance measurement: impact on organizational performance, *International Business Review*, vol. 13, 2004, pp. 485-502.

[14] D. J. Spiegelhalter, , K. R. Abrams, J. P. Myles, *Bayesian approaches to clinical trials and health-care evaluation,* vol. 13: Wiley, Chichester, 2004.

[15] G. Koop, M.F.J. Steel and J. Osiewalski, Posterior analysis of stochastic frontier models using Gibbs sampling, *Computational Statistics*, vol. 10, 1995, pp. 353-373.

[16] R. E. Kass and L. Wasserman, The selection of prior distributions by formal rules, *Journal of the American Statistical Association*, 1996, pp. 1343-1370.

[17] M. A. Tanner, *tools for statistical inference*, 2ed., New York, Springer-Verlag 1993.

[18] W. R. Gilks, S. Richardson, D. J. Spiegelhalter, *Markov chain Monte Carlo in practice*, New York, Chapman and Hall/CRC, 1996.

[19] D. Spiegelhalter, A. Thomas, N. Best, W. Gilks, *BUGS 0.5: Bayesian inference using Gibbs sampling manual (version ii)*, MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK, 1996.

[20] C. Sheu and S. L. O'Curry, Simulation-based Bayesian inference using BUGS, *Behavior Research Methods*, vol. 30, 1998, pp. 232-237.

[21] A. T. Azar, Adaptive Neuro-Fuzzy Systems, *Fuzzy Systems*, 2010, pp. 85–110.

[22] S. C. Jang JR, Mizutani E, *Neuro-Fuzzy and Soft Computing*, New Delhi, Prentice-Hall of India, 2006.

[23] S. L. Chiu, Fuzzy model identification based on cluster estimation, *Journal of intelligent and Fuzzy systems*, vol. 2, 1994, pp. 267-278.

[24] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, A. Van Der Linde, Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 2002, pp. 583-639.

[25] S. P. Brooks and A. Gelman, Alternative methods for monitoring convergence of iterative simulations, *Journal of Computational and Graphical Statistics*, vol. 7, 1998, pp. 434-455.

[26] G. B. Hua, Residential construction demand forecasting using economic indicators: a comparative study of artificial neural networks and multiple regression, *Construction Management and Economics*, vol. 14, 1996, pp. 25–34.

[27] L. Aburto and R. Weber, Improved supply chain management based on hybrid demand forecasts, *Applied Soft Computing*, vol. 7, 2007, pp. 136-144.

[28] F. Zheng and 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, 2011, pp. 1471-1475.

[29] C. F. Chien, C. Y. Hsu, C. W. Hsiao, Manufacturing intelligence to forecast and reduce semiconductor cycle time, *Journal of Intelligent Manufacturing*, 2011 pp. 1-14.

[30] S. F. Arnold, *Mathematical Statistics,* Prentice-Hall, 1990.

[31] R. E. Walpole, Mayers, R.H., Mayers, S.L, *Probability and statistics for engineers and scienticts*, 6 ed., New Jersey, Prentice Hall Int. , 1998.

[32] G. O. Roberts, Markov chain concepts related to sampling algorithms, *Markov chain Monte Carlo in practice*, vol. 57, 1996.

[33] V. Olej, P. Hajek, IF-inference systems design for prediction of ozone time series: the case of Pardubice micro-region, *in Proceedings of the 20th International Conference on Artificial Neural Networks*, Thessaloniki, Greece, 2010, pp. 1-11.

[34] J. M. Mendel, R. I. John and F. Liu, Interval type-2 Fuzzy logic systems made simple, *IEEE Trans. on Fuzzy Systems*, vol. 14, 2006, pp. 808-821.

[35] C.D Căleanu – Fuzzy versus Neural Techniques for Prediction, *Proceedings of the International Conference communications 2002, Military Technical Academy*, Politehnica University of Bucharest and IEEE Romanian Section, ISBN 973-8290-67-8, 5 – 7 December, Bucharest, 2002, pp. 288-293.

[36] F. Neri. Software agents as a versatile simulation tool to model complex systems. *WSEAS Transactions on Information Science and Applications*, WSEAS Press (Sofia Bulgaria), issue 5, vol. 7, 2010, pp. 609-618.

[37] Yordanova, S., Petrova, R., Mastorakis, N. E., & Mladenov, V. (2006). Sugeno predictive neuro-fuzzy controller for control of nonlinear plant under uncertainties. *WSEAS Transactions on Systems*, issue 5, vol. 8, pp. 1814-1821.

[38] A. Azizi, A. b. Ali Yazid, L. W. Ping. Prediction of the Production Throughput under Uncertain Conditions Using ANFIS: A Case Study. *International Journal for Advances in Computer Science*, ISSN 2218-6638, Volume 2, Issue 4, 2011, pp 27-32.

[39] A. Azizi, A. b. Ali Yazid, L. W. Ping, and M. Mohammadzadeh. A Hybrid model of ARIMA and Multiple Polynomial Regression for Uncertainties Modeling of a Serial Production Line. *Proceedings of the ICETM 2012 : International Conference on Engineering and Technology Management*, P-ISSN 2010-376X and E-ISSN 2010-3778, Kuala Lumpur, Malaysia, 2012.

[40] A. Azizi, A. b. Ali Yazid, L. W. Ping, and M. Mohammadzadeh, A Bayesian

Autoregressive Integrated Moving Average Model for Estimating the Production Throughput under Uncertain Conditions: A Case Study, *International Journal for Advances in Computer Science*, ISSN 2218-6638, Volume 2, Issue 4, 2011, pp. 5-10.