SUPPLY CHAIN ANALYSIS USING SIMULATION, GAUSSIAN PROCESS MODELLING AND OPTIMISATION

Smew, W.; Young, P. & Geraghty, J.

Enterprise Process Research Centre, School of Mechanical and Manufacturing Engineering,
Dublin City University, Dublin 9, Ireland
E-Mail: walid.smew@dcu.ie, paul.young@dcu.ie, john.geraghty@dcu.ie

Abstract

This paper presents a simulation study on production and inventory control at the supply chain (SC) level and examines through a series of experiments the potential impact of the Hybrid Kanban-CONWIP production control strategy on the trade-off between the conflicting objectives of maximising customer service level and minimising Work-In-Process (WIP). Simulation based optimisation is computationally expensive approach to determining an optimal configuration of operational parameters for any stochastic system. In this paper we demonstrate an optimisation framework that will yield solutions with an accuracy that is reasonable for decision makers and computationally less expensive than simulation based optimisation. A simulation model for a centralised serial SC adopting Hybrid Kanban-CONWIP to process a single product was developed in order to explore the impact of some essential input factors on customer service level and average WIP through Design of Experiments (DOE), Gaussian Process Modelling (GP) and Metamodel-Based Optimisation using the Desirability Function. The precision of the results from this approach was determined by comparison to results from Simulation-Based Optimisation by means of Genetic Algorithms (GA). It has been shown that this framework will address the trade-off between accuracy and computational efficiency requirements of the decision maker.
(Extended paper from the International Conference on Industrial Logistics – ICIL 2012, Zadar, Croatia. Received in August 2012, accepted in February 2013. This paper was with the authors 3 months for 1 revision.)

Key Words: Supply Chain Management, Hybrid Kanban-CONWIP, Discrete Event Simulation, Gaussian Process Modelling, Optimisation

1. INTRODUCTION

Supply chain management focuses on the integration of suppliers, manufacturers, warehouses, and stores. The challenge in supply chain (SC) integration is to co-ordinate activities across the entire supply chain so that the enterprise can improve performance, reduce system inventory levels and potential inventory cost, increase customer service level, better utilise resources, and effectively respond to changes in the market. Designing and implementing a globally optimal SC is quiet difficult because of its dynamics and the conflicting objectives employed by the different facilities and partners in the SC. Sharing information and implementing a proper SC control strategy are the keys to successfully achieve these goals.

SCs are often categorised as push-based, pull-based or push-pull SCs. This is probably stemmed from the manufacturing revolution of the 1980s, in which manufacturing systems were divided into these categories. In push-based SCs, such as material requirements planning (MRP) systems, production and distribution decisions are based on long term demand forecast and products are pushed as quickly as possible through the channel, from the production side upstream to retailers downstream and hence control throughput and observe WIP from time to time. This characteristic may enable the system to reduce delivery lead time since many semi-finished or finished products are available but also, it will lead to the inability to meet changing demand patterns and to the bullwhip effect and all its inefficiencies [1-3].

In pull-based SCs, such as Kanban systems, production and distribution decisions are based on true customer demand rather than forecasted demand and SC members do not hold any excess inventory and only respond to specific orders. In pull-based supply chains usually result in significant reductions in system inventory levels, reduced costs, and better response to market changes. However, this system may not work well in multi-product environments and in environments with large demand variations. This in turn may result in significant backorders, longer delivery lead times, and higher late penalty costs [1-4].

In push-pull SCs, some stages of the SC, typically the initial stages are operated in a push-based manner while the remaining stages are operated in pull-based manner. This hybrid control often compromises the conflicting performance characteristics of the push and the pull environments so that a better system performance can be anticipated. For advantages and disadvantages of the push and pull systems and other details see the literature [2, 5-8]. Of all the existing production control strategies, only the Hybrid Kanban-CONWIP will be considered in this paper.

2. LITERATURE REVIEW

Hybrid Kanban-CONWIP is a pull production control strategy that was firstly proposed by Bonvik et al. in 1997 [9] in an attempt to combine the control advantages of CONWIP (high throughput and limited WIP levels in the system at all times) with those of Kanban to limit excessive build-up of WIP in front of bottleneck stages (control of inventory levels at the individual stages). The mechanism of the Hybrid Kanban-CONWIP system is shown in Fig. 1 where a limit known as work-in-process capacity (*WIPCAP*) is placed on the amount of inventory that may be in the whole system at any time by using CONWIP cards and a limit is placed on the amount of inventory in each stage (excluding the last stage) at any time using Kanban cards. When orders arrive at the final node and there are enough finished goods it is assumed that they will be immediately shipped to final customers. Production orders and required materials are then released to the first node (considering its on hand Kanban cards and the *WIPCAP*) where they will be pushed through the different stages in the SC till they are processed completely and leave the final node.
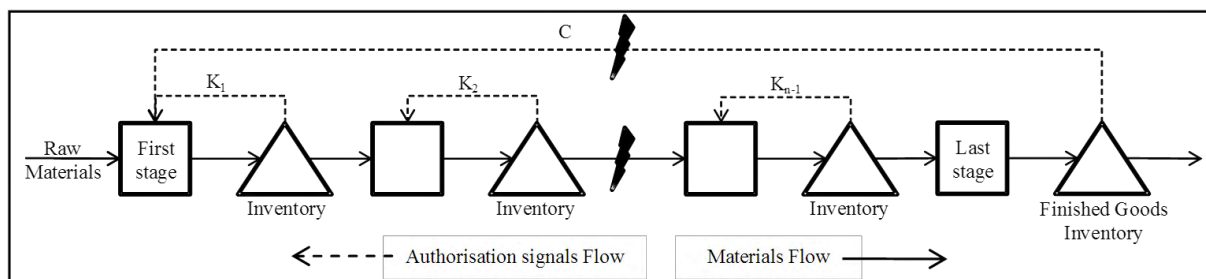

Figure 1: Hybrid Kanban-CONWIP control mechanism [10].

A comparison of Kanban Control Strategy (KCS), minimal blocking KCS, Basestock Control Strategy (BSCS), CONWIP, and Hybrid Kanban-CONWIP was presented in [9]. These different pull production control strategies (PCS) were compared in a four-stage tandem production line using simulation. Each of the control policies was compared using constant demand and demand that had a stepped increase/decrease. It was found that the Hybrid Kanban-CONWIP strategy decreased inventories by 10 % to 20 % over KCS while maintaining the same service levels. The performance of BSCS and CONWIP strategies fell between those of KCS and Hybrid Kanban-CONWIP. Two papers proposed a generic pull model encapsulating the three basic pull control strategies, KCS, CONWIP and BSCS and

allows customised pull control strategies such as Hybrid Kanban-CONWIP to be developed [10, 11]. Simulation and an evolutionary algorithm were used to study the generic model. Details of the evolutionary algorithm were given in [10] while results on extensive experimentation on the effect of factors (i.e., line imbalance, machine reliability) on the proposed generic pull model were given in [11].

Kleijnen and Gaury [12] presented a methodology that was a stage wise combination of four techniques: (i) simulation, (ii) optimisation, (iii) risk or uncertainty analysis, and (iv) bootstrapping. The methodology was illustrated through a production-control study for the four-stage, single product production line utilised by [9]. They defined Robustness as the capability to maintain short-term service in a variety of environments i.e. the probability of the short-term fill-rate (service level) remaining within a pre-specified range. In their research work Kleijnen and Gaury compared four systems, namely KCS, CONWIP, Hybrid Kanban-CONWIP, and Generic. The optimal parameters found in [9] were used for KCS, CONWIP and Hybrid Kanban-CONWIP and they used a Genetic Algorithm to determine the optimal parameters for the Generic pull system. For the risk analysis step, seventeen inputs were considered; the mean and variance of the processing time for each of the four production stages, mean time between failures and mean time to repair per production stage, and the demand rate. The inputs were varied over a range of ±5 % around their base values. They concluded that in this particular example, Hybrid Kanban-CONWIP was best when risk was not ignored; otherwise Generic was best and therefore, risk considerations can influence the selection of a PCS.

Geraghty and Heavey [13] showed that under certain conditions the Horizontally Integrated Hybrid Production Control Strategy, HIHPS, favoured by Hodgson and Wang [14] is equivalent to the Hybrid Kanban-CONWIP production control strategy introduced by Bonvik et al. [9]. Lavoie et al. [8] conducted a comparison study of pull control mechanisms for unreliable tandem transfer lines producing a single product using a combined discrete/continuous simulation and considering "inventory & storage" holding and shortage costs. They observed that the Hybrid Kanban-CONWIP mechanism always outperforms CONWIP and Kanban when storage space and inventory costs are considered explicitly. However, Hybrid Kanban-CONWIP was equivalent to CONWIP and both outperform Kanban when storage space costs are not considered explicitly but rather aggregated with the inventory costs in the holding costs as usually done.

In this research work, mathematical models applicable to a serial SC controlled by Hybrid Kanban-CONWIP are developed. These mathematical models are then translated into a discrete event simulation model using ExtendSim. Through a series of designed experiments; the devolved simulation model will be used to investigate the potential impact of the Hybrid Kanban-CONWIP control on the trade-off between the conflicting objectives of maximising customer service level and minimising average WIP. A software package, JMP, was used to construct the experimental design matrices, analyse the data, and develop the GP meta-models. A multi-objective optimisation using the developed meta-models will be conducted by means of the Desirability Approach (DA) method, also the Pareto-Optimal Genetic Algorithm (POGA) will be used to conduct a simulation-based optimisation as well. The aim of this work is to provide a framework that could be used by SC managers to assess the suitability of a production control strategy to their system with high accuracy and computational efficiency.

3. MODELLING A HYBRID KANBAN-CONWIP SUPPLY CHAIN

The centralised Hybrid Kanban-CONWIP SC in this paper is a production–distribution system, in which the production line of each firm has a similarity to a "work center" being a

part of a "global line" of supply and also in which a virtual center of control governs the SC and manages the information and parts flow and the inventories along the chain. The proposed SC consists of four nodes in series representing 4 different firms (e.g.; a supplier, a manufacturer, a distributer, and a retailer). A *WIPCAP* is assigned to the entire SC in order to limit the inventory that may be in the system at any time by using CONWIP cards and also, a limit is allocated on the amount of inventory at each stage in the SC (excluding the last stage since the amount of parts in the entire SC can never exceed the inventory allowed in this stage) using Kanban cards. When orders arrive at the final node and there are enough finished goods they will be immediately shipped to customers and production orders and required materials are then released to the first node considering its production capacity constraints, its on hand Kanban cards, and the *WIPCAP*. The first node requires two authorisation cards: one from the second node (Kanban pattern) and another one from the last node (CONWIP pattern). Both cards are attached to the part and when at node 2 only the Kanban card is sent back to node 1 and the CONWIP card remains attached to the part until it reaches the finished good inventory of the last node and is delivered. If the supply of finished goods from the last node to customers is limited, a portion of the orders will not be fulfilled and will be considered as lost sales.

3.1 Notations and Definitions

i- Material flow variables

P_t^i: Pipeline (WIP) in node i in period t, where $i = 1, 2, …, n$
Y_t^i: Finished Goods Inventory (FGI) of node i in period t
S_t^i: Shipments from node i to node $i+1$ in period t
O_t^i: Output from the pipeline of node i in period t
I_t^i: Input to the pipeline of the node i in period t

ii- Information flow variables

D_t: Incoming orders to the SC at final node in period t
OP_t: Orders placed by the SC to the first node in period t
B_t: Backlog of incoming orders in the SC in period t
APC_t: Available number of CONWIP cards in the SC in period t
AKC_t^i: Available number of Kanban cards in node i in period t
TY_t^i: Available total FGI in node i, in period t

iii- Model parameters

L^i: Cycle (lead) time for a unit in the pipeline to arrive to the FGI of the node i
MLP^i: Maximum number of units to be processed in the node i (node capacity)
$WIPCap$: Total number of CONWIP cards in the SC
K^i: Total number of Kanban cards of node i

3.2 Model Formulation

i- Shipments

Incoming orders at the final node (D_t), will be shipped to customers but the inventory constraints may effect these shipments. Therefore, S_t^n, is given by:

$$S_t^n = \min[TY_t^n, D_t] \qquad (1)$$

Otherwise, S_t^i depends on TY_t^i, the maximum number of parts to be processed in node $i + 1$ during its lead time (MLP^{i+1}/L^{i+1}), and the available number of Kanban cards in node $i + 1$:

$$S_t^i = \begin{cases} \min\left[TY_t^i, \left(\frac{MLP^{i+1}}{L^{i+1}}\right), \left(K^{i+1} - [P_{t-1}^{i+1} + Y_{t-1}^{i+1}]\right)\right] & i = 1 \text{ to } n - 2 \\ \min\left[TY_t^i, \frac{MLP^{i+1}}{L^{i+1}}\right] & i = n - 1 \end{cases} \tag{2}$$

ii- Materials flow, WIP, and inventory

Assuming that the initial conditions are known, then P_t^i and Y_t^i are given by:

$$P_t^i = P_{t-1}^i + I_t^i - O_t^i \tag{3}$$
$$Y_t^i = Y_{t-1}^i + O_t^i - S_t^i \tag{4}$$

The available total FGI (TY_t^i) for node i in period t can be calculated as follows:

$$TY_t^i = \begin{cases} Y_{t-1}^i + O_t^i & \text{for nodes } i = 1 \text{ to } i = n \text{ - } 1 \\ \max[(Y_{t-1}^n + O_t^n - B_{t-1}), 0] & \text{for node } i = n \end{cases} \tag{5}$$

where:

$$O_t^i = I_{t-L^i}^i \tag{6}$$

$$I_t^i = \begin{cases} OP_t & \text{for } i = 1 \\ S_t^{i-1} & \text{for } i \neq 1 \end{cases} \tag{7}$$

iii- Orders

OP_t depends on the maximum number of parts to be processed in the first node during its cycle time (MLP^1/L^i), and the available number of Kanban and CONWIP cards in node 1 and is given by:

$$OP_t = \max\left[\min\left(\frac{MLP^1}{L^1}, [K^1 - (P_{t-1}^1 + Y_{t-1}^1)], APC_t\right), 0\right] \tag{8}$$

APC_t is the available number of CONWIP cards in the SC in period t and can be calculated as follows:

$$APC_t = WIPCap - \left[\left(\sum_{i=1}^n P_{t-1}^i + \sum_{i=1}^n Y_{t-1}^i\right) - S_t^n\right] \tag{9}$$

iv- Performance Measures

The performance measures for the Hybrid Kanban-CONWIP SC are the Average Inventory in the system ($AWIP$) and the Average Service Level (ASL) achieved by the system after N periods of time:

$$AWIP(N) = \frac{\sum_{n=1}^N [TWIP(n)]}{N} \tag{10}$$

$$ASL(N) = \frac{\sum_{n=1}^N [SL(n)]}{N} \tag{11}$$

4. SIMULATION MODELLING

Simulation modelling enables powerful "what-if" analyses to test several strategies and scenarios; on the other hand it permits the comparison of various operational alternatives leading to better future decision. The modelling phase starts after the supply chain is analysed and the key components and performance measures are identified [15]. The conceptual, mathematical model described earlier has been implemented in a periodic review simulation model using ExtendSim. The conventional modelling approach employed by ExtendSim is object oriented so; the translation of the conceptual model in a computer simulation model could be performed using library objects. The flow of items and information could be modelled by means of dynamic entities. In this work an advanced modelling approach based on programming code, tables and events generators is proposed. ExtendSim provides the user with the compiled simulation language Mod-L that can be used for writing codes. The SC

flow of dynamic entities, representing items and information, is substituted by information recorded in tables. Without dynamic entities flow all the simulation events are generated using event generator objects (provided by the library) and, in correspondence of such events, the code elaborate and update the information stored in tables. This modelling approach yielded a flexible, parametric and time efficient simulation model.

5. META-MODEL DEVELOPMENT

Gaussian Process (GP) models are increasingly used as surrogate statistical models for predicting output of computer experiments. Generally, GP models are both interpolators and smoothers of data and are effective predictors when the response surface of interest is a smooth function of the parameter space. GP prediction is simply a weighted linear combination of all output values already observed. The weights depend on the distance between the new input to be predicted and *all* the old inputs. A good introduction on GP emulators (metamodels), key steps in building them and how to design the experiments are given in O'Hagan [16]. GP models, also known as *Kriging* and *Spatial Correlation*, are global rather than local (i.e., fitted to data that are obtained form larger experimental space than the small spaces used in low order polynomial regression models) [17]. In 1989, Sacks et al. [18] developed the spatial correlation parametric regression modelling approach that predicts unknown values of deterministic simulation models. In stochastic simulations, this property disappears and averages of replicated observations are obtained for each scenario. GP interpolates these averages which are still random however, GP still attractive because it may decrease the prediction bias and hence the Mean Square Errors (MSE) at scenarios close together [19]. Mitchell and Morris [20] described the spatial correlation model that is appropriate for stochastic simulation responses as follows:

$$g(\theta) = \mu(\theta) + \sum_{i=1}^{n} \gamma_i \times exp\left(-\sum_{j=1}^{p} \omega_j \left|\theta_j - \theta_j^i\right|^2\right) \tag{12}$$

where μ, γ_i, and ω_j is a set of coefficients to be estimated.

The standard design for computer experiments is space filling designs. These designs aim to fill the space with number of points (runs) so that the complete input parameter space is sampled. The usual space filling design for computer experiments is the Latin Hypercube Design (LHD) [21]. A LHD in n runs for k input variables is a $(n \times k)$ matrix where each column is a random permutation of levels $1, 2, ..., n$. In LHD, the possible number of runs of the simulator has to be selected first; a common practice for an effective initial number is $n = 10 \times k$ [22], and the range for each input variable is then divided into n equal sections or levels. The maxi-min criterion that maximises the minimum Euclidean distance between points in the design is used to produce a LHD with good properties [23].

A total of five independent variables ($k = 5$) will be considered for Hybrid Kanban-CONWIP SC: node capacity, WIPCAP, total number of Kanban cards of node 1, node 2, and node 3. Each of these inputs will be varied over a selected working range as per the selected experimental design to investigate their impact on the conflicting SC performance measures (responses), customer service level and average WIP. In this study, the incoming orders to the SC (customer demand) is assumed to be a random variable follows a log-normal distribution with a constant mean demand of 4 items per period and a variable demand standard deviation, SD which will be considered at three different levels: a lower level of 1 item corresponding to a demand coefficient of variation (CV) of 25 %, a medium level of 4.5 (CV of 112.5 %), and a higher level of 8 items (CV of 200 %). Any node capacity will be from a minimum level of capacity of 8 units corresponding to a mean demand of 4 units per period and a cycle (lead) time of 2 periods for an item in the pipeline to arrive to the FGI of any node and a higher level of capacity of 24 items corresponding to three times the minimum available capacity.

Simulation runs were performed by changing the selected input factors one at a time in order to determine the best working range of the WIPCAP and the total number of Kanban cards of nodes 1-3 that will enable us to extract as much information as possible from system. The selected ranges were decided to be from a minimum of 12 Kanban cards for nodes 1-3 and 20 WIPCAP cards to a maximum of 20 Kanban cards for nodes 1-3 and 135 WIPCAP cards.

JMP, statistical and data analysis software developed by SAS Institute Inc. (www.jmp.com), was used to apply GP modelling to the output simulation data; construct the different LHD matrices in accordance with $n = 10k$ and provided that all the possible levels of each input factor is considered, estimate the different terms of model (12) using the method of maximum likelihood estimation in order to fit, build, and validate the SC performance measures meta-models under $SD = 1, 4.5,$ and 8. ExtendSim was also used to conduct the simulation experiments according to the constructed LHDs. For all experiments conducted: (i) simulation run-length was 11000 periods with a 1000 period warm-up run excluded which was sufficient for deleting the influence of nodes initial conditions, (ii) the average output of 50 replications was taken for each response variable, (iii) the lead time is fixed, the same for all nodes, and has a value of 2 periods, and (iv) all nodes have the same node capacity value and will be varied according the specified working range. As an example, Table I provides a summary of the experimental design for $SD = 8$.

Table I: Experimental Design Summary at Demand $SD = 8$.

Factor	Name	Units	Low	High	
A	Node Capacity	Units	8	24	
B	WIP-Cap	Cards	50	136	
C	Node1Kanbans	Cards	12	20	
D	Node2Kanbans	Cards	12	20	
E	Node3Kanbans	Cards	12	20	
Response	Name	Units	Min	Max	Ratio
Y_1	Service Level	%	83.3175	98.4118	1.1812
Y_2	Average WIP	Items	44.1443	127.5539	2.8895

To estimate the different terms of the GP metamodels (μ, σ^2, γ_i, ω_j known as Theta in JMP) and the Nugget term which stands for the noise or randomness in the simulation model, Eq. (12) was fitted to the simulation output data of the different generated LHDs by using the method of maximum likelihood estimation. JMP generates a model report summarising the Functional ANOVA and the estimated terms of the GP metamodel and also generates its actual by predicted plot. It is always necessary to examine the fitted metamodels to ensure that they provide an adequate approximation to the true system. The jackknife or leave one out procedure is used to generate Actual by Predicted plots as a measure of goodness-of-fit. As an example, the Hybrid Kanban-CONWIP SC GP metamodels at $SD = 8$ reports generated by JMP are presented in Table II, Table III, and Fig. 2.

Table II: Hybrid Kanban-CONWIP service level GP metamodel report at $SD = 8$.

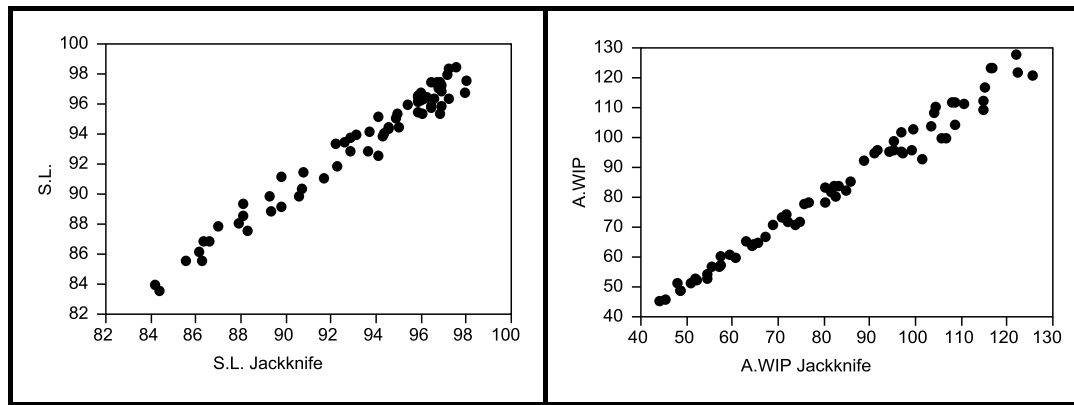| Column	Theta	Total Sensitivity	Main Effect	Capacity Interaction	WIP-Cap Interaction	K1 Interaction	K2 Interaction	K3 Interaction
Capacity	3.66E-06	2.62E-03	2.59E-03	-	3.21E-05	4.55E-11	3.16E-09	2.34E-10
WIP-Cap	9.09E-05	9.87E-01	9.86E-01	3.21E-05	-	6.72E-05	6.04E-04	1.92E-06
K1	2.58E-05	5.71E-03	5.65E-03	4.55E-11	6.72E-05	-	3.24E-09	1.60E-10
K2	3.55E-05	4.05E-03	3.45E-03	3.16E-09	6.04E-04	3.24E-09	-	2.28E-10
K3	8.07E-06	1.26E-03	1.26E-03	2.34E-10	1.92E-06	1.60E-10	2.28E-10	-
μ	83.81	σ^2	239.41	Nugget	0.0017624			

Figure 2: Hybrid Kanban-CONWIP SC GP metamodels actual by predicted plot at $SD = 8$.

Table III: Hybrid Kanban-CONWIP Average WIP GP metamodel report at $SD = 8$.

Column	Theta	Total Sensitivity	Main Effect	Capacity Interaction	WIP-Cap Interaction	K1 Interaction	K2 Interaction	K3 Interaction
Capacity	2.64E-03	4.33E-03	4.29E-03	-	3.69E-05	1.69E-06	3.59E-07	0.00E+00
WIP-Cap	2.21E-05	9.88E-01	9.87E-01	3.69E-05	-	3.74E-04	4.62E-07	0.00E+00
K1	1.38E-04	7.68E-03	7.30E-03	1.69E-06	3.74E-04	-	9.38E-10	0.00E+00
K2	9.90E-06	6.92E-04	6.91E-04	3.59E-07	4.62E-07	9.38E-10	-	0.00E+00
K3	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	-
μ	44.80	σ^2	4144.93	Nugget	0.0020133			

The different metamodels reports show how the considered performance measures are affected by each of the input factors as a main and total sensitivity effect under a specific demand SD. This main and total sensitivity effect percentage (%), if exists, reveals the amount of influence and importance an input factor have on the considered performance measure. By looking at the different actual by predicted plots it can be seen that the different points are scattered close the 45 degree diagonal line which reflects how accurately the models predict simulation outputs. In addition, examination of the Service Level and Average WIP trade-off curves of the developed meta-models and the simulation model using new different optimum scenarios for the Hybrid Kanban-CONWIP SC policy (as illustrated in Fig. 3 in the optimisation section), reveals how accurately the fitted models predict simulation outputs. According to these results the developed models are statistically significant and will be considered valid and can be used for further analysis.

6. OPTIMISATION

A multi-objective optimisation problem with conflicting objectives implies that there is no unique optimal solution to the problem. Instead, a set of solutions can be obtained that are indifferent to each other such that no improvement can be found in terms of any of the objectives without resulting in deterioration of the performance of at least one of the objectives so, decision makers must make a judgment and find good compromises (trade-offs) among them to arrive at a particular decision. A set of such optimal solutions is known as the Pareto-frontier. These optimal solutions are non-inferior or non-dominated in the sense that there is no other solution in the search space superior than them when all the objectives are taken into consideration [24-26]. To generate a set of such non-dominated solutions, two different multi-objective optimisation techniques are considered in this research work; the DA to perform a metamodel-based optimisation and the POGA to carry out a simulation-based optimisation.

6.1 Solution using the Desirability Approach

The Desirability Function Approach (DA) [27] was applied to the developed GP metamodels and the optimum input factors combinations that minimise the average WIP for targeted service levels were obtained. To build the trade-off curves for the Hybrid Kanban-CONWIP, the service level (at several possible targets) and the average WIP were optimised simultaneously for each selected demand *SD*. A two-sided desirability function was defined for each targeted service level as illustrated in the following example:

$$d_1 = \begin{cases} 0 & \text{service level} < 83.5 \\ (0,1) & 83.5 \leq \text{service level} \leq 84 \\ (0,1) & 84 \leq \text{service level} \leq 84.5 \\ 0 & \text{service level} > 84.5 \end{cases} \tag{13}$$

For the average WIP meta-models the goal is always to minimise them and their one-sided desirability function can be defined as in the following example (average WIP meta-model under *SD* = 8:

$$d_2 = \begin{cases} 1 & \text{average } WIP < 44.14 \\ (0,1) & 44.14 \leq \text{average } WIP \leq 127.55 \\ 0 & \text{average } WIP > 127.55 \end{cases} \tag{14}$$

The desirability approach is implemented JMP statistical software package and can be applied by selecting a goal for each response and then by setting the lower and upper limits for each goal, the software will search the design factors space, using the fitted models, and will generate a list of potential optimum factor settings (GP-DA solutions) that meet the defined criteria.

6.2 Solution using Genetic Algorithm

GA is a search algorithm for optimisation that begins from a set of randomly generated potential solutions (chromosomes). GA makes this initial population evolve through successive iterations (generations) towards a population that is expected to contain the optimum solution [10]. The Pareto-Optimal Genetic Algorithm (POGA) code for ExtendSim developed by [25] was applied directly to the simulation model of this work to find the set of non-dominated solutions (SIM-POGA) in a discrete space comprised of 5 decision variables (e.g. for SD8: *WIPCAP* with a range of 50 to 136, nodes 1, 2 and 3 Kanbans with a range of 12 to 20, and node capacity with a range of 8 to 24) for an objective vector comprised of 2 components; average service level to be *maximised* and average WIP to be *minimised.* Therefore, the aim is to find the set of pareto-optimal solutions so that:

$$\forall N_i, N_j \in X \neg \{S(N_i) \geq S(N_j) \wedge W(N_i) \leq W(N_j) \wedge$$
$$\lfloor S(N_i) > S(N_j) \vee W(N_i) < W(N_j) \rfloor \} \tag{15}$$

where: N_i = {Node Capacity, *WIPCAP*, Node 1 Kanbans, Node 2 Kanbans, Node 3 Kanbans},
S = Service Level, and W = Average *WIP*.

6.3 Optimisation results

The actual simulation outputs (SIM-GP-DA) as estimate refinements of the resulted optimal solutions (GP-DA); which is improving the accuracy of the estimated optimal solutions through experimentation with the simulation model under the parameters found to be "optimal" for given Service Level targets, also, the resulted optimal solutions SIM-POGA are illustrated in the trade-off curve shown in Fig. 3 as an example for *SD* = 8. The percentage deviations (errors) between each of the metamodels based optimisation approaches (SIM-GP-

DA) and the simulation based approach (SIM-POGA; as the solutions from this method are the most accurate and generated from applying the POGA directly to the SC simulation model) for the Hybrid Kanban-CONWIP SC are presented in Table IV and Fig. 4 as an example for Demand $SD = 8$.
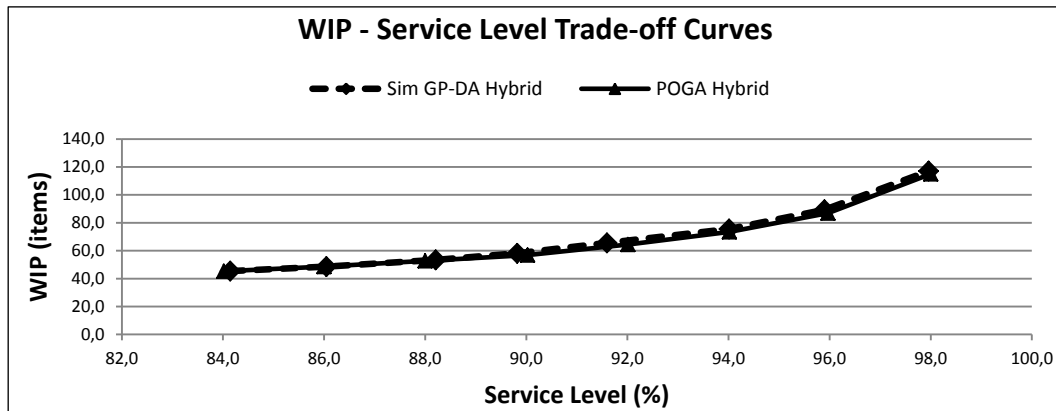

Figure 3: Hybrid Kanban-CONWIP SC GP trade-off curve for $SD = 8$.

Table IV: Percentage deviations from SIM-POGA of Hybrid Kanban-CONWIP SC for $SD=8$.

Scenario Number	SL Target	SIM-GP-DA	
		Error SL	Error WIP
1	84±0.5	-0.152 %	-0.344 %
2	86±0.5	0.053 %	-0.050 %
3	88±0.5	-0.236 %	-1.166 %
4	90±0.5	0.228 %	-2.142 %
5	92±0.5	0.448 %	-1.773 %
6	94±0.5	0.008 %	-2.854 %
7	96±0.5	0.074 %	-2.778 %
8	98±0.5	0.041 %	-1.807 %
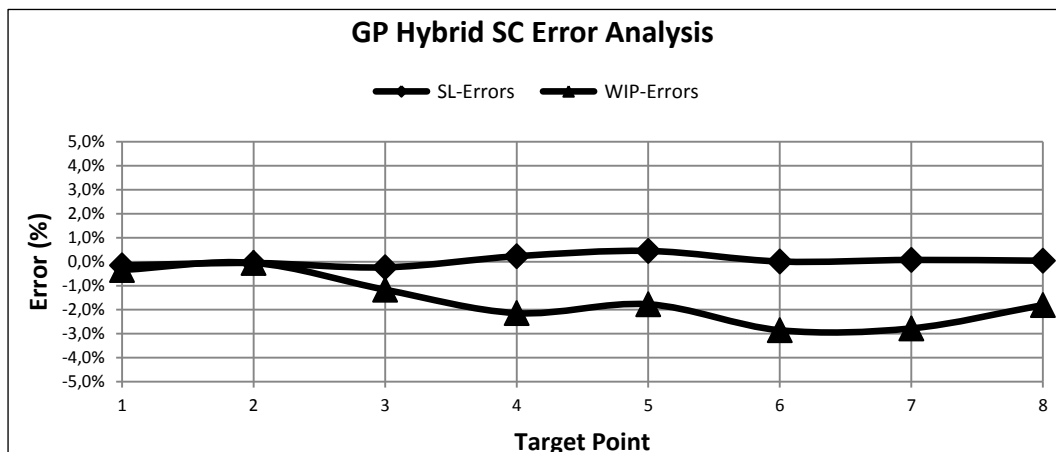
Figure 4: Errors analysis of the Hybrid Kanban-CONWIP SC at $SD = 8$.

It can be seen that the errors result from estimates of average WIP are higher as opposed to service levels However, the GP metamodels can be considered very accurate and gave superior approximation to the SC simulation model under different demand SD. The level of inventory is the main warning sign in any SC therefore, how much inventory is being held to

achieve targeted customer service levels at an acceptable price is very important issue to understand. The Hybrid Kanban-CONWIP SC trade-off curves shown in Fig. 3 can effectively be used to attain a sense of balance between the conflicting objectives as there is gain along one dimension and loss along the other one. Furthermore, the relationship between this two conflicting objectives is not a straight line; increasing service level gives an exponential curve relationship in the extra stock to be carried. For example, it can be seen that there is a huge increase of an average of 65 % in WIP for an increase of 5 % in service level (when moving from 93 % towards a high 98 % service level). It is clear that the increase in average service levels is coming at the cost of increase in average inventory levels. The decision of whether such a trade-off is acceptable depends on the cost of the inventory, cost of lost sales and several other aspects that need to be considered before making a decision by the supply chain managers.

7. CONCLUSIONS

The aim of this research was to provide a framework that could be used by SC managers to assess the suitability of various production control strategies to their system with reasonably high accuracy while being computationally efficient. Simulation optimisation is possible by embedding an optimisation algorithm such as the POGA within the developed simulation model; however the technique can take substantial computer time and may be impractical for relatively small problems. If the initial effort for building metamodels is undertaken, then the process of finding an optimum, or at least, a set of good solutions, can be made easier (as the use of metamodels can significantly speed-up the search for an optimum, given the initial overhead of developing these metamodels). In this paper, this methodology has been used to generate trade-off curves for a Hybrid Kanban-CONWIP controlled serial supply chain subjected to high demand variability. A trade-off curve can be used to assess the cost (in terms of inventory levels) of providing higher service levels and to determine the operational parameters of the production control strategy. This eliminates the need for *a priori* knowledge of the solution domain as is required for single objective optimisation approaches that require the setting of fitness values or weights and, thereby, cause the solution to be partly a result of decisions taken to determine such weights.

REFERENCES

[1] Simchi-Levi, D.; Kaminsky, P.; Simchi-Levi, E. (2003). *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies*, McGraw-Hill, New York

[2] Ghrayeb, O.; Phojanamongkolkij, N.; Tan, B. A. (2009). A hybrid push/pull system in assemble-to-order manufacturing environment, *Journal of Intelligent Manufacturing*, Vol. 20, No. 4, 379-387, doi:10.1007/s10845-008-0112-6

[3] Takahashi, K.; Nakamura, N. (2004). Push, pull, or hybrid control in supply chain management, *International Journal of Computer Integrated Manufacturing.*, Vol. 17, No. 2, 126-140, doi:10.1080/09511920310001593083

[4] Krishnamurthy, A.; Suri, R.; Vernon, M. (2004). Re-examining the performance of MRP and Kanban material control strategies for multi-product flexible manufacturing systems, *International Journal of Flexible Manufacturing Systems*, Vol. 16, No. 2, 123-150, doi:10.1023/B:FLEX.0000044837.86194.19

[5] Spearman, M. L.; Woodruff, D. L.; Hopp, W. J. (1990). CONWIP: a pull alternative to kanban, *International Journal of Production Research*, Vol. 28, No. 5, 879-894, doi:10.1080/00207549008942761

[6] Deleersnyder, J.-L.; Hodgson, T. J.; King, R. E.; O'Grady, P. J.; Savva, A. (1992). Integrating Kanban type pull systems and MRP type push systems: Insights from a Markovian model, *IIE Transactions*, Vol. 24, No. 3, 43-56, doi:10.1080/07408179208964223

[7]   Spearman, M. L.; Zazanis, M. A. (1992). Push and Pull Production Systems: Issues and Comparisons, *Operations Research.*, Vol. 40, No. 3, 521-532, doi:10.1287/opre.40.3.521

[8]   Lavoie, P.; Gharbi, A.; Kenne, J.-P. (2010). A comparative study of pull control mechanisms for unreliable homogenous transfer lines, *International Journal of Production Economics*, Vol. 124, No. 1, 241-251, doi:10.1016/j.ijpe.2009.11.022

[9]   Bonvik, A. M.; Couch, C. E; Gershwin, S. B. (1997). A comparison of production-line control mechanisms, *International Journal of Production Research*, Vol. 35, No. 3, 789-804, doi:10.1080/002075497195713

[10]   Gaury, E. G. A.; Pierreval, H; Kleijnen, J. P. C. (2000). An evolutionary approach to select a pull system among Kanban, Conwip and Hybrid, *Journal of Intelligent Manufacturing.*, Vol. 11, No. 2, 157-167, doi:10.1023/A:1008938816257

[11]   Gaury, E. G. A.; Kleijnen, J. P. C.; Pierreval, H. (2001). A Methodology to Customize Pull Control Systems, *Journal of The Operational Research Society*, Vol. 52, No. 7, 789-799, doi:10.1057/palgrave.jors.2601153

[12]   Kleijnen, J. P. C.; Gaury, E. (2003). Short-term robustness of production management systems: A case study, *European Journal of Operational Research*, Vol. 148, No. 2, 452-465, doi:10.1016/S0377-2217(02)00437-X

[13]   Geraghty, J.; Heavey, C. (2004). A comparison of Hybrid Push/Pull and CONWIP/Pull production inventory control policies, *International Journal of Production Economics*, Vol. 91, No. 1, 75-90, doi:10.1016/S0925-5273(03)00210-X

[14]   Hodgson, T. J.; Wang, D. (1991). Optimal hybrid push/pull control strategies for a parallel multistage system: Part II, *International Journal of Production Research*, Vol. 29, No. 7, 1453-1460, doi:10.1080/00207549108948022

[15]   Jain, S.; Workman, R. W.; Collins, L. M.; Ervin, E. C.; Lathrop, A. P. (2001). Development of a high-level supply chain simulation model, *Proceedings of the 33rd Winter Simulation Conference*, 1129-1137

[16]   O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial, *Reliability Engineering and System Safety*, Vol. 91, No. 10-11, 1290-1300, doi:10.1016/j.ress.2005.11.025

[17]   Barton, R. R.; Meckesheimer, M. (2006). Metamodel-based simulation optimization, Henderson, S. G.; Nelson, B. L. (Eds), *Handbooks in Operations Research and Management Science: Simulation*, Elsevier, New York, Vol. 13, 535-574, doi:10.1016/S0927-0507(06)13018-2

[18]   Sacks, J.; Welch, W. J.; Mitchell, T. J.; Wynn, H. P. (1989). Design and analysis of computer experiments, *Statistical science*, Vol. 4, No. 4, 409-423, doi:10.1214/ss/1177012413

[19]   Van Beers, W. C. M.; Kleijnen, J. P. C. (2008). Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping, *European Journal of Operational Research*, Vol. 186, No. 3, 1099-1113, doi:10.1016/j.ejor.2007.02.035

[20]   Mitchell, T. J.; Morris, M. D. (1992). The spatial correlation function approach to response surface estimation, *Proceedings of the 24th Winter Simulation Conference*, 565-571

[21]   McKay, M. D.; Beckman, R. J.; Conover, W. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, Vol. 21, No. 2, 239-245, doi:10.1080/00401706.1979.10489755

[22]   Loeppky, J. L.; Sacks, J.; Welch, W. J. (2009). Choosing the Sample Size of a Computer Experiment: A Practical Guide, *Technometrics*, Vol. 51, No. 4, 366-376, doi:10.1198/TECH.2009.08040

[23]   Johnson, M. E.; Moore, L. M.; Ylvisaker, D. (1990). Minimax and maximin distance designs, *Journal of Statistical Planning and Inference*, Vol. 26, No. 2, 131-148, doi:10.1016/0378-3758(90)90122-B

[24]   Keeney, R. L.; Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Cambridge University Press, Cambridge

[25]   Kernan, B.; Geraghty, J. (2004). A Multi-Objective Genetic Algorithm for Extend, *Proceedings of the First Irish Workshop on Simulation in Manufacturing*, Services and Logistics, 83-92

[26]   Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison-Wesley, Reading, MA

[27]   Derringer, G.; Suich, R. (1980). Simultaneous optimization of several response variables, *Journal of Quality Technology*, Vol. 12, No. 4, 214-219