Deif, Hoda. Published in the International Journal of Flexible Manufacturing Systems, 19(3): 128-150. September 2007.

Assessing capacity scalability policies in RMS using system dynamics

Ahmed M. Deif · Hoda A. ElMaraghy

Abstract This paper presents a model for assessing different capacity scalability policies in Reconfigurable Manufacturing System (RMS) for different changing demand scenarios. The novelty of this approach is two fold: (1) it is the first attempt to explore different capacity scalability policies in RMS based on multiple performance measures, mainly scaling rate, Work In Process level, inventory level and backlog level; and (2) the dynamic scalability process in RMS is modeled for the first time using System Dynamics. Different policies for capacity scalability for various demand scenarios were assessed. Numerical simulation results obtained using the developed capacity scalability model showed that the best capacity scalability policy to be adopted for RMS is dependent on the anticipated demand pattern as well as the various manufacturing objectives. The presented assessment results will help the capacity scalability planners better decide the different tradeoffs between the competing strategic and operational objectives of the manufacturing enterprise, before setting the suitable capacity scalability plan parameters.

Keywords Capacity · Scalability · Dynamics · Reconfigurable manufacturing systems

A. M. Deif · H. A. ElMaraghy (✉)
Intelligent Manufacturing Centre, Department of Industrial and Manufacturing Systems Engineering, University of Windsor, Room 204, Odette Building, 401 Sunset Avenue, Windsor, ON, Canada N9B 3P4
e-mail: hae@uwindsor.ca

A. M. Deif
e-mail: deif@uwindsor.ca

1 Introduction

Today's manufacturing plants are facing an increasingly turbulent environment and rising customer requirements including, among others, mass customization of products, highly volatile demand patterns and the need for high delivery performance. In order to meet these challenges, the importance of responsive and cost effective manufacturing systems is growing. Reconfigurable Manufacturing Systems (RMS) have been characterized as having the capability to react to unpredictable market changes in a cost effective manner by adjusting their capacity and functionality. In other words, RMS aims to enhance manufacturing responsiveness in the production of low-cost and high-quality products. The key characteristics of RMS, which enable these systems to achieve their goals, are modularity, integrabilty, convertibility, customization and diagnosability (Mehrabi et al. 2000). Other enablers include reconfigurable process planning and changeable production planning and control systems (Wiendahl et al. 2007).

Manufacturers as well as researchers agree that the ability of a company to fulfill market demands is primarily determined by its capacity. Thus, in order to adequately respond to fluctuations in the level of market demand, the need for volume flexibility, as defined in the Flexible Manufacturing Systems (FMS) paradigm (Sethi and Sethi 1990) or capacity scalability as defined in RMS paradigm, is highly recognized.

ElMaraghy (2005) explains the dimensions of capacity scalability in RMS through classifying the scalability characteristics into physical scalability and logical scalability attributes. Examples of physical scalability include the adding or removing of material handling equipment, machines, and machine modules, such as axes of motions or heads, as well as tools or other components. Examples of logical scalability include increasing or decreasing the number of shifts or the number of workers as well as outsourcing workers. Modular components' design and interfaces as well as open control architecture are basic enabling technologies required for achieving physical capacity scalability in RMS.

The challenges facing the capacity scalability in RMS are not only the availability of required enabling technologies, but also the management process. Therefore, the question of "which is the best scalability strategy or policy to be adopted?" needs to be answered. The capacity scalability policies in RMS cannot be designed in isolation from the adopted marketing and operation strategies. The capacity scalability process functions best when its policies are consistent with the recognized priorities of market strategy as well as with the operational objectives. These priorities and objectives are usually translated into different manufacturing performance measures, mainly production rate, Work In Process (WIP) level, inventory level and demand fulfillment (or backlog level).

This paper presents an approach to fulfill the need to decide on the best capacity scalability policies in RMS that capture the competing objectives of market strategies and operational objectives. The developed system dynamics model assesses the performance of different scalability policies against different demand scenarios to assist the capacity scalability planners in deciding on the different tradeoffs involved in this process.

2 Literature review

The capacity scalability problem is classically addressed from a static view as the problem of capacity expansion to meet increasing demand at a minimum cost. The first study of the capacity expansion problem was conducted by Manne (1967). Extensive review of the classical capacity expansion problem can be found in Luss (1982). However, in today's market, manufacturing systems are typically faced with a rapidly changing and uncertain demand together with continuous advancement of technology, and thus the need to address the capacity scalability problem from a dynamic view point is becoming more obvious.

A dynamic model developed by Duffie and Falu (2002) for closed loop Production Planning and Control (PPC) was proposed to control WIP and capacity. They investigated the effect of choosing different capacity scalability controller gains as well as the WIP controller gains on system performance and how this can be used to achieve required system responses. This work was extended by Kim and Duffie (2004) to study the effect of capacity disturbances and capacity delays on system performance in single work stations. This was further applied to multiple workstations in Kim and Duffie (2005). Their results highlighted the fact that, if capacity can be adjusted more often with less delay, the system's performance would be significantly improved in changing demand environments.

Another dynamic model that manipulates feedback control with the help of logistics operating curves, developed by Nyhuis (1994) to control work in process WIP and capacity of manufacturing systems, was presented in Wiendahl and Breithaupt (1999, 2000). In this approach, the required capacity scalability was found using flexibility curves, which indicate the time delay of each capacity scaling step. The capacity scalability controller chooses the best capacity scaling decision based on the acceptable backlog value and delay.

In RMS literature, Asl and Ulsoy (2002) presented a dynamic approach to capacity scalability modeling based on the use of feedback control. Suboptimal solutions that are robust against demand variations and partially minimize the cost of capacity scalability were presented.

Deif and ElMaraghy (2006) developed a dynamic model for capacity scalability in RMS and analyzed the model based on control theoretic approaches to indicate the best design for the scalability controller. Results highlighted the importance of accounting for the different physical and logical delays together with the trade-off decisions between responsiveness and cost when designing the capacity scalability controllers. They further introduced an optimization unit to the capacity scalability model to optimally decide on the exact value of the scalability controller gain in Deif and ElMaraghy (2007).

The previous dynamic approaches to model and analyze the capacity scalability problem were based on the application of control theory as a dynamic tool and utilizing its inherent feedback mechanisms. Although they offered good solutions for controlling capacity under conditions of fluctuating demand, they did not offer any comparative assessment of different scalability policies or management strategies. In addition, the performance measures considered during the capacity scalability modeling were either the backlog level or the backlog and WIP levels.

Other measures such as the inventory level and production rate were not considered in the previous approaches.

Another candidate approach to dynamically model and analyze manufacturing systems, and especially their different planning and control policies, is System Dynamics (SD) introduced by Forrester (1961). Baines and Harrison (1999) argues that SD has distinctive performance when considering strategic issues in manufacturing companies. Furthermore, SD models have proven their applicability to analyzing strategic scenarios as well as simulation of policies and operations in manufacturing systems (Helo 2000).

Application of SD in manufacturing systems to date focused mainly on pure inventory supply chain where the objective was to study how the system can be designed and analyzed to respond to unanticipated demand with maximum stability and minimum cost. Examples include: Sterman (2000), Fowler (1999), Towill and Del Vechho (1994), Towill (1993), and Wikner et al. (1991).

The capacity scalability problem has rarely been tackled using SD models. An early attempt by Evans and Naim (1994) aimed at developing an SD model for supply chains with capacity constraints and studying the effect of capacity constraints on a system's performance. Helo (2000) suggested a capacity-based supply chain model that includes a mechanism for handling the trade-off between lead time and capacity utilization. It was shown that this capacity analysis, including the surge effect, in supply chains would improve their responsiveness. Goncalves et al. (2005) highlighted the issue of capacity variation in their push–pull manufacturing SD model through the effect of capacity utilization on the production start rate. Anderson et al. (2005) considered logical capacity scalability in supply chains for service and custom manufacturing. They showed the effect of reducing lead-time and sharing the demand information on improving system performance.

The previous work paved the road for capacity consideration in SD models. However, to the authors' knowledge, there is no reported work that has modeled RMS or capacity scalability management in RMS using the SD approach.

The presented modeling approach differs from previous dynamic capacity scalability models as it considers more performance measures to determine the best capacity scalability management policy. This takes into account the scaling rate, WIP level, inventory level and backlog level. The objective is to explore the best scalability policy to be adopted. From a dynamic perspective, this is also a new approach to model RMS using system dynamics.

3 Model description

The development of an appropriate model for capacity scalability in RMS, which incorporates different parameters involved in that process, is an essential step. Figure 1 shows a system dynamic model for capacity scalability in RMS. The model expresses the capacity as a stock level controlled by a scaling rate. This dynamic representation of the scaling process is suitable for capturing the ability of RMS to adjust their capacity and, hence, makes the model a valid representation for these systems. In addition, the model incorporates the WIP, inventory and backlog

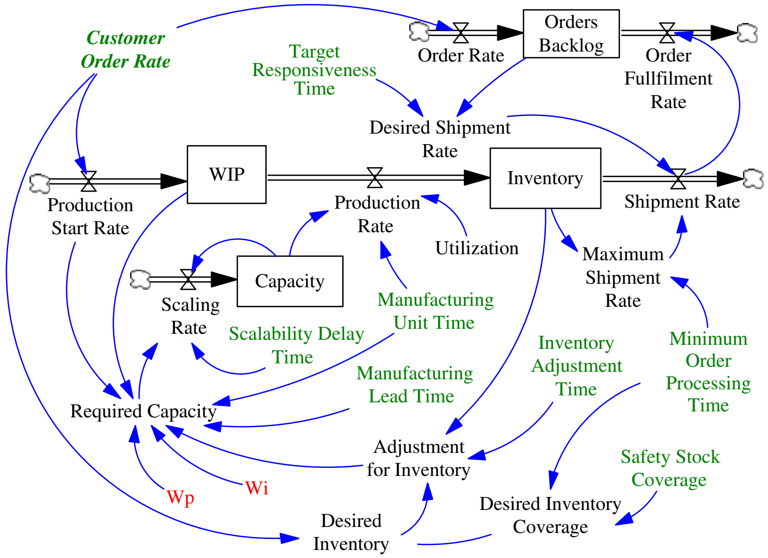

Fig. 1 Model structure for capacity scalability in RMS

levels of the system as additional manufacturing parameters that are involved in the scalability process as well as being used to evaluate the overall system performance. It is important to note that the developed model is suitable (or designed) for make-to-order industries.

In this paper, a continuous-time model is used because it provides an acceptable approximation of the continuous capacity scalability process in RMS at that level of abstraction and aggregation. Both the operations management and system dynamics literature support the use of continuous models for capacity planning (e.g., Anderson et al. 2005; Sethi and Thompson 2000; Holt et al. 1960). Finally, similar dynamic characteristics can be obtained using discrete-time models (John et al. 1994). Deterministic data is used in the analysis to provide a simple yet effective comparison between the various scenarios.

3.1 Model nomenclature

Let $C(t)$ = the capacity level at time t.
$B(t)$ = the backlog level at time t.
$I(t)$ = the inventory level at time t.
$WIP(t)$ = the WIP level at time t.
$PR(t)$ = the production rate at time t.
$PSR(t)$ = the production start rate at time t. Note that $PSR(t)$ equals customer order at time t.
$CO(t)$ = the customer order rate at time t.
$OR(t)$ = the order rate at time t. Note that $OR(t)$ equals customer order at time t.

$OFR(t)$ = the order fulfillment rate at time t. This rate presents the information flow of the products leaving the system.
TRT = the target responsiveness time. It represents the manufacturer's goal for the interval between placement and receipt of orders.
$DSR(t)$ = the desired shipment rate at time t.
$ShR(t)$ = the shipment rate at time t. It is the rate of physical product leaving the system.
$MSR(t)$ = the maximum shipment rate at time t. It depends on the system's current inventory.
$MOPT$ = the minimum order processing time. It represents the minimum time required to process and ship an order.
SSC = the safety stock coverage time. It is the time required to cover unexpected variation in demand (the higher this value, the greater the service level).
DIC = the desired inventory coverage time. It is the time required to cover shipments during the expected rate.
IAT = the inventory adjustment time. It is the time required to react for inventory discrepancy between the current inventory level and the desired level.
$DI(t)$ = the desired inventory level at time t. It is based on customer demand.
$AI(t)$ = the adjustment for inventory rate at time t.
U = the utilization level of the available capacity.
MLT = the manufacturing lead time. It is the time required to process products.
W_i = the relative weight of inventory consideration in capacity scalability decision.
W_p = the relative weight of demand consideration in capacity scalability decision.
$RC(t)$ = the required capacity at time t.
SDT = the scalability delay time. Time require to scale the system's capacity.
$SR(t)$ = the scalability rate at time t. This is the major decision variable in the capacity scalability process in RMS.
MUT = the manufacturing unit time (used to switch from stock to rate to maintain dimensional balance).

3.2 Model logic

3.2.1 Capacity scalability planning and control

Capacity scalability decisions are controlled through the scaling rate (Eq. 1).

$$\dot{C}(t) = SR(t) \tag{1}$$

The equation for the scaling rate is determined by the required capacity together with the scalability delay (Eq. 2).

$$SR(t) = \frac{C(t) - RC(t)}{SDT} \tag{2}$$

The required capacity (Eq. 3) has three components and each component reflects a planning and control policy.

$$RC(t) = \left[(W_p * PSR(t)) + (W_i * AI(t)) + \left((1 - W_p - W_i) * \frac{WIP(t)}{MLT} \right) \right] * MUT$$
$$\text{where } 0 \leq W_p \leq 1 \text{ and } W_p + W_i \leq 1 \tag{3}$$

The $RC(t)$ is defined in this manner to be able to change and to adapt to the capacity scalability policy based on various marketing and operational objectives. The first policy is based on chasing the demand. This is achieved by setting the production start rate equal to the customer order so that production is dedicated only to chase the demand. The second policy is inventory-based where the required capacity is controlled by inventory adjustments. Inventory adjustments refers to the filling rate compensating for the discrepancy between the current inventory level and the required inventory level (the later is usually set based on the service level set by the marketing strategy). The third policy is WIP-based where the capacity is changed to keep WIP at a constant level. The change in the WIP level is based on Little's law (WIP = Production Rate × MLT) where RC replaces the production rate. Integrating the three main parameters (production, inventory and WIP levels) and manipulating their interaction through the values of the different weights involved in this equation, captures the dynamics of capacity scalability of RMS in a make-to-order environment. Details of these dynamics are discussed in "Numerical simulation results and analysis".

3.2.2 Inventory control

The inventory control mechanism in the developed model follows the same one introduced by Sterman (2000). The inventory adjustment is controlled by the inventory gap between desired and current inventory levels (Eq. 4).

$$AI(t) = \frac{DI(t) - I(t)}{IAT} \tag{4}$$

The desired inventory level is calculated using Eqs. 5 and 6 to ensure enough coverage of products for the anticipated demand.

$$DI(t) = CO(t) * DIC \tag{5}$$

$$DIC = MOPT + SSC \tag{6}$$

The desired inventory coverage includes two components. First, the manufacturer should maintain enough coverage to ship at the expected rate requiring a base coverage level equal to MOPT. Second, to ensure an adequate level of service, the manufacturer adds safety stock coverage (SSC).

The current inventory level is controlled by Eq. 7.

$$\dot{I}(t) = PR(t) - ShR(t) \tag{7}$$

3.2.3 Production control

The WIP level is determined by the difference between the production start rate and the actual production rate (Eq. 8)

$$\dot{WIP}(t) = PSR(t) - PR(t) \tag{8}$$

The production start rate is set to be equal to the customer order (Eq. 9). The production rate is controlled by the capacity scalability level, as this is the typical case in RMS where recent technological solutions allow frequent capacity changes. Such a characteristic was the reason behind modeling the logic of the production control to be dependent on capacity scaling and then directly relating that scaled level of capacity to the production level. However, for practical consideration, the capacity is factored by the real system utilization level (Eq. 10) to account for variations between the two levels. It should be noted that the calculation of the utilization level is beyond the scope of this paper and is taken as an input.

$$PSR(t) = CO(t) \tag{9}$$

$$PR(t) = \frac{C(t) * U}{MUT} \tag{10}$$

3.2.4 Customer orders fulfillment

The customer orders are fulfilled by the order fulfillment rate, which is controlled by the shipment rate (Eq. 11). The shipment rate is given by the minimum of either the desired shipment rate or the maximum shipment rate (Eq. 12). This is the case for make-to-order industries considered in this work. However, make-to-stock industries can adopt the same model by maximizing rather than minimizing Eq. 12

$$OFR(t) = ShR(t) \tag{11}$$

$$ShR(t) = Min(DSR(t), MSR(t)). \tag{12}$$

The desired shipment rate is calculated as a function of the current backlog and the target responsiveness time (Eq. 13). In the RMS paradigm, the responsiveness time is a major performance measure of these responsive systems and tends to be low.

$$DSR(t) = \frac{B(t)}{TRT}. \tag{13}$$

The backlog level is calculated as the difference between the order rate (which is exactly equal to the customer orders as in Eq. 14) and the order fulfillment rate (Eq. 14). In RMS systems, backlog is supposed to be at a low level; practically, however, it cannot be zero.

$$OR(t) = CO(t) \tag{14}$$

$$\dot{B}(t) = OR(t) - OFR(t) \tag{15}$$

The maximum shipment rate is determined by the available inventory level and the minimum order processing time (Eq. 16)

$$MSR(t) = \frac{I(t)}{MOPT} \tag{16}$$

4 Numerical simulation results and analysis

In order to illustrate the dynamic behavior and performance of the different capacity scalability policies, two dynamic demand patterns are considered. The first pattern demonstrates a sudden step change in demand to give a dramatic shock to the system. If the system responds well to such change, then it bodes well for other inputs to which the system may be subjected. The other pattern represents cyclic demand to demonstrate the fluctuating scenarios for which RMSs are designed.

Four capacity scalability policies are selected for assessment. The first policy is based on chasing the demand, which is achieved by setting the production start rate equal to the customer order and setting W_p to be 1. In this case, the capacity scalability mechanism (or capacity stock correction mechanism using SD terminology) will change based on demand only. The second policy is inventory-based, where the capacity scalability level is changed to adjust production rate to meet the target inventory level. This is achieved in the model by setting $W_i = 1$. The third policy is WIP-based where the capacity scalability mechanism would strive to keep the WIP level constant at the target level, which is calculated based on Little's law (i.e., $WIP(0) = CO(0) * MLT$). This is similar to the PPC policy known as CONWIP. In this policy, W_i and W_p are both set to zero. The fourth policy considered for assessment is what we have called the hybrid policy, where the three previous parameters (demand, inventory and WIP) are considered equally when adjusting the capacity scalability level. This is achieved by setting W_i and W_p to be equal to 1/3. The best values of these weights may be specified based on experience or by conducting sensitivity analysis. A sensitivity analysis would be useful in providing insights into the impact of the different considered performance measures over the system performance in the case of using a hybrid policy. However, this analysis is not within the scope of this paper.

The performance measures used for the assessment are: (1) the capacity level since this level with its filling rate reflects the scaling effort and cost, (2) WIP level to reflect manufacturing stability, (3) inventory level to reflect part of the cost, and (4) backlog level to reflect responsiveness of the system.

The chosen parameters' values for the base case are shown in Table 1. The selected values for the different time parameters are based mainly on the practical experience of one of the authors in make-to-order computer monitors manufacturing. Altering the values of these parameters and examining the impact of each one of them can lead to some insights; however, such analysis is beyond the scope of this paper. The model is initialized at equilibrium (i.e., the initial values of the WIP, capacity, inventory and backlog levels are used as the target values for each policy, Sterman 2000) with the demand constant and simulated for 50 weeks. For

Table 1 Values of the base case parameters

Parameter	Value	Unit
Target responsiveness time (TRT)	2	Weeks
Manufacturing lead time (MLT)	4	Weeks
Scalability delay time (SDT)	2	Weeks
Inventory adjustment time (IAT)	4	Weeks
Minimum order processing time (MOPT)	1	Weeks
Safety stock coverage (SSC)	1	Weeks
Utilization level (U)	90%	N/A
Manufacturing unit time (MUT)	1	Weeks

simplicity, the scalability delay time is assumed to be constant reflecting cases where the times required for stopping the line to scale the capacity and the ramp up time are relatively higher than the time required to install the capacity unit itself. In practical cases this assumption may be relaxed.

4.1 Sudden change demand scenario

Figure 2 shows the step change in the demand pattern. In this scenario, the demand suddenly increases by 20% (from 10 K of products/week to 12 K of products/week) at week 5. The behavior of the SD systems is mainly analyzed through the model stock levels and rates. Thus, for the assessment purpose of the results, the different stocks and rates in the capacity scalability model of RMS will be plotted in the following figures for different scalability policies as performance measures.

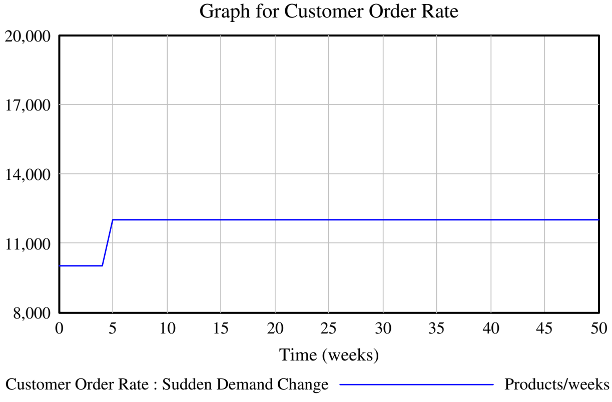

Fig. 2 Demand for sudden change scenario

Figure 3a–d shows the fluctuation of capacity, inventory, WIP and backlog levels under this sudden demand change for different policies. Figure 4a–d plots the same dynamic behavior for the scalability, inventory adjustment, production, and shipment rates. These results will be analyzed next for each capacity scalability policy.

4.1.1 *Chasing demand scalability policy*

In the legends to Figs. 3 and 4, this policy is referred to as number 4 and is expected to result in the best responsiveness level since the objective is to simply satisfy the demand. However, the numerical simulation results provide different insights.

The capacity scalability system immediately responds to the demand shock by increasing the capacity stock level by 20% (Fig. 3a) and in turn the production rate is increased to match the demand increase (Fig. 4c). However, the production rate

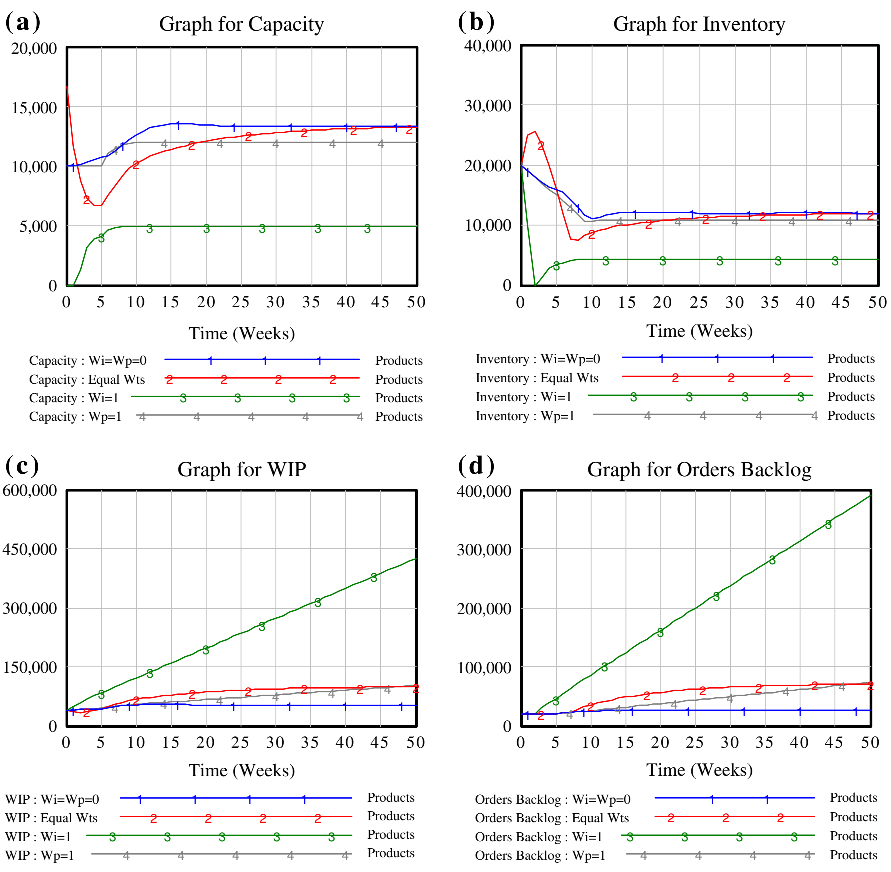

Fig. 3 Dynamic response for different stocks in the developed RMS capacity scalability model in a sudden demand scenario. (**a**) Capacity Stock, (**b**) Inventory Stock, (**c**) WIP Stock, (**d**) Backlog Stock

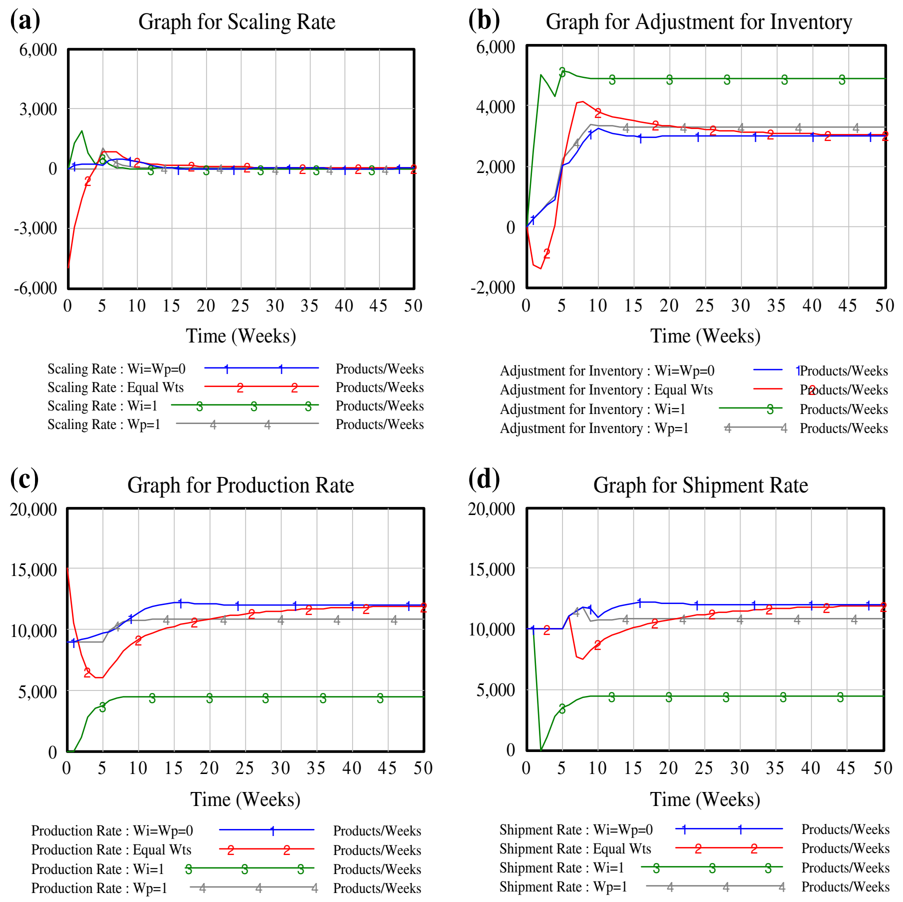

Fig. 4 Dynamic response for different rates in the developed RMS capacity scalability model in a sudden demand scenario. (**a**) Capacity Scaling Rate, (**b**) Inventory Adjustment Rate, (**c**) Production Rate, (**d**) Shipment Rate

does not exactly equal the demand since the system studied has a 90% utilization level. In addition, the scaling rate (Fig. 4a) increases for a short time to reach the required increased level showing a smooth scalability performance. The cost efficiency of the capacity scalability process is directly proportional to the amount of the capacity to be added/removed, and this policy reflects the objective of RMS in supplying the exact capacity required for the demand change. Thus, from a scalability cost perspective, this policy showed a good performance under this demand scenario. It should be noted that, in an RMS paradigm, the economy of scope is considered rather than the economy of scale (Koren et al. 1999; ElMaraghy 2005).

The backlog order level increases gradually until it reaches the level of 75 K of products, which is more than the required level of 24 K (Fig. 3d). The shipment rate controlling the backlog level in this policy overshoots the demand increase before it settles down to the required demand rate (Fig. 4d). This occurs to compensate for the sudden drop in the inventory level resulting from the demand increase. The high level of backlog orders indicates a low responsiveness performance of this scalability policy.

On the inventory side, the inventory stock level drops gradually from its initial value of 20 K products to 11 K products to satisfy the demand increase (Fig. 3b). The adjustment of inventory rate increases gradually to fill the gap between the desired level and the current level of inventory (Fig. 4b). However, this rate fails to get the inventory stock to the desired level since the capacity scalability policy depends only on the demand and does not account for the inventory. The low inventory level of this policy indicates a good cost performance, but unfortunately at the expense of the customer service level (i.e., higher probability of stock outs and delays).

The WIP stock level increases until it reaches 100 K of products, which is almost double the required level of 48 K of products (Fig. 3c). This high level of WIP is expected since the capacity scalability stock mechanism in this policy does not consider WIP level correction. Although WIP level plays a major role in the stability of manufacturing systems, this high WIP level (with its associated cost) is considered a disadvantage of this policy.

The general assessment of this policy is that it achieves an acceptable level of capacity scalability cost performance. However, the responsiveness performance is not satisfactory. This observation sounds surprising for a policy where the objective of which is chasing the demand as its only priority. Thus, a conclusion that can be drawn here would be that in make-to-order RMS and in a sudden demand change scenario, if inventory, WIP and backlog levels are considered together with the scalability level in system's performance assessment, the demand chasing policy will not achieve the best responsiveness level.

4.1.2 Inventory based scalability policy

This policy (referred to as number 3 in Figs. 3 and 4) scales the capacity, and in turn the production rate, to keep the finished inventory at a certain level based on the demand and desired inventory coverage. It is a typical policy used for make-to-stock (MTS) industries where marketing depends on the offered service level. This policy shows the minimum level of capacity stock (Fig. 3a) since production is adjusted to satisfy only the difference between the required and the actual inventory levels (Fig. 4c). However, the highest capacity rate overshoot is realized in this policy (Fig. 4a) at the beginning of the observed period. This undesirable reaction is due to the drop in the inventory level in response to the sudden change in demand and the delay in compensating for this drop by increasing production due to the capacity scalability delay time. The drop in inventory level contrasts sharply with the objective of this policy, and thus production has to exceed the shipment rate long enough to restore inventory back to its desired level.

In addition to the low capacity scalability cost of this policy at the steady state, it also has another profitable advantage by having the lowest inventory level (Fig. 3b).

This is because the inventory level is targeted only to satisfy the demand during the considered demand period. However, this is at the expense of the undesirable dynamic pattern of the inventory adjustment rate (Fig. 4b) where, in addition to having the highest value among all polices, it experienced two instances of overshooting to bring the inventory to the desired level. Furthermore, going below the desired inventory level is at the expense of responsiveness. The unexpected performance of this policy concerning inventory levels is mainly a result of accounting only for inventory adjustment to scale the capacity without taking into consideration the expected loss rate (or demand during the inventory adjustment time) as discussed in Sterman (2000).

The inventory based policy shows a continuous increase in both the backlog and the WIP levels (Fig. 3c, d). The low value of the shipment rate as a result of low inventory level (Fig. 4c) explains the backlog accumulation. The low value of the production rate also explains the WIP build-up as both the production and shipment rates target only maintaining the finished inventory target level.

In general, although this policy shows some profitable effects such as minimum capacity scalability requirements and low inventory level, the unsatisfactory dynamic performance with its associated instability will negatively affect both the profit as well as the performance of the manufacturing system.

4.1.3 WIP based scalability policy

In the legends to Figs. 3 and 4, this policy is referred to as number 1 and is based on Little's law, where the production (controlled by the capacity scalability) is based on both the WIP level and manufacturing lead-time. The highest level for capacity stock, and thus production rate together with a small overshoot, are witnessed in this policy (13.5 K products) as it has to account for both demand and WIP levels (Figs. 3a and 4c). As for the scaling rate (Fig. 4a), the policy has the lowest overshoot. This shows that this policy has a low performance in terms of production cost but has a good dynamic stability.

The same tradeoff is experienced in this policy concerning the inventory level as it shows no overshooting, which is a desirable dynamic behavior, but at the same time it has the highest value among other polices over the observed period (Fig. 3b). In addition, the inventory adjustment rate has the lowest overshoot and steady state values (Fig. 4b). The high production rate observed in this policy explains both the high inventory level and the low adjustment rate.

A significant characteristic of this policy is that it has the best performance in terms of WIP level showing the lowest and the most stable level (Fig. 3c). This performance is expected as the scaling mechanism is based on WIP level adjustment according to demand. Furthermore, this policy has the lowest backlog level (Fig. 3d) indicating a high responsiveness performance. This is due to the high shipment rate (Fig. 4d) of this policy as a result of the high inventory level, as explained earlier.

In general, the WIP-based scalability policy in this sudden change demand pattern confirms the conventional wisdom that WIP is a major factor for manufacturing system stability. This was clear in the positive dynamic behavior of the system's different parameters. In addition, the analysis shows that this policy has the best responsiveness level. This is a very important conclusion for RMS capacity scalability management. However, the previous desired performance was achieved at the expense of the cost of capacity scalability. Therefore, capacity scalability planners must then consider the trade-off between the profit due to stable and responsive systems and the cost associated with higher scalability levels when making their decisions as to the best policy to adopt.

4.1.4 Hybrid scalability policy

In the legends to Figs. 3 and 4, this policy is referred to as number 2. The capacity scalability stock falls at the beginning (since it has a high equilibrium starting point) and then gradually rises with the sudden demand change, showing no overshooting until it reaches the same level of capacity in the WIP based policy, although at 25 weeks later (Fig. 3a). The production rate follows the same behavior (Fig. 4c). The scaling rate rises from a negative value, since the required capacity is less than the actual starting capacity (which has a high value to maintain the simulation equilibrium starting point), and then has a small overshoot to balance for a later drop in the capacity level before reaching equilibrium (Fig. 4a). The high level of capacity is due to the scalability mechanism of this policy that strives to account partially for demand, WIP and inventory levels (based on the selected weights).

The inventory level is subjected to an overshoot at the beginning in response to the sudden fall in the production rate followed by a drop until it gradually rises again to reach the same level of the WIP-based policy (Fig. 3b). Consequently, the inventory adjustment rate has the same behavior but in an opposite direction (Fig. 4b). This small oscillatory behavior in the inventory level and adjustment rate at the early period is not desirable, but it is unavoidable (in this sudden demand change scenario) due to the different delays involved in the system structure.

The WIP level rises above the required level with the same value as that for the chasing demand policy (Fig. 3c). This high WIP level value is due to the compromise between the tendency to continuously increase the WIP resulting from the partial accounting for the inventory level (as in the inventory-based policy) and the desire to keep the WIP at a low level also by partially accounting for WIP (as in the WIP-based policy).

The backlog level is much higher than the required level (as in the case of the chasing demand scenario) indicating a low responsiveness level (Fig. 3d). This is due to the objective of this policy to keep inventory at an acceptable limit, which negatively affects the shipment rate leading to a drop in its value before it rises again (Fig. 4d).

In general, although this policy tires to balance between different performance measures when deciding on the value of the capacity increments, it did not show the best performance among other policies focusing on performance measure in this specific demand scenario. This was shown in the undesirable dynamic behavior of both the capacity and inventory and the high level of WIP and backlog with their associated cost.

4.2 Cyclic demand scenario

The second scenario considered is the cyclic demand. This demand pattern features repeated increase and decrease and provides a good test case for assessing dynamic behavior of the considered capacity scalability policies for RMS. The cyclic demand pattern is shown in Fig. 5 where cycles have a mean of 10 K products/week ± 20%.

Figure 6a–d shows the fluctuation of capacity, inventory, WIP and backlog levels under this cyclic demand change for different policies. Figure 7a–d plots the same dynamic behavior for the scalability, inventory adjustment, production, and shipment rates. The analysis of Figs. 6 and 7 will follow the same scheme as that for the sudden change in demand scenario for each capacity scalability policy:

4.2.1 Chasing demand scalability policy

The capacity stock, following this policy, responds to this demand pattern in the same cyclic manner with a phase lag of 2 weeks due to the capacity scalability delay time (Fig. 6a). The amplitude of the capacity stock cycles is exactly equal to the demand values, indicating a cost effective performance. However, as discussed in the sudden change demand scenario, the production rate does not equal the demand due to the same utilization limitation (Fig. 7c). In addition, the scaling rate in this policy shows the highest amount of amplification (Fig. 7a) due to the desire to exactly chase the demand. This analysis shows that applying demand chasing policy requires a trade-off between profit of supplying the exact capacity needed and amount of effort required for that to happen. Such trade-offs is a typical challenge confronting RMS implementation.

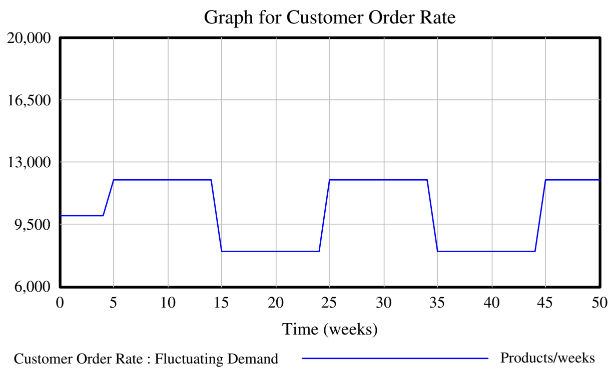

Fig. 5 Cyclic demand pattern

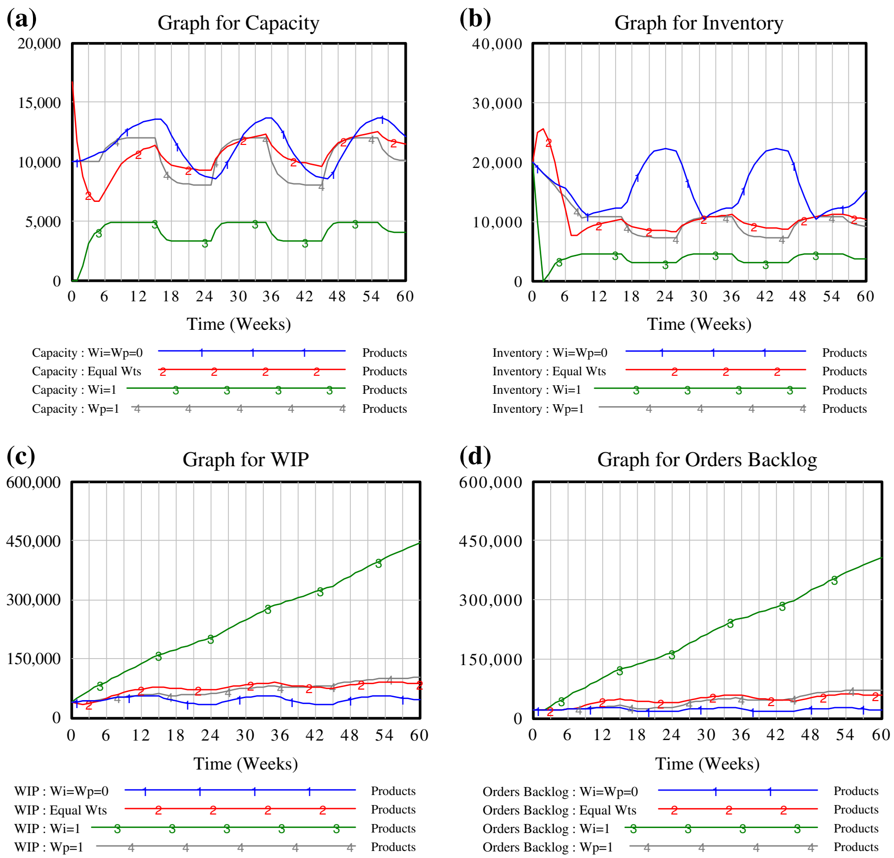

Fig. 6 Dynamic response for different stocks in the developed RMS capacity scalability model in a cyclic demand scenario. (**a**) Capacity Stock, (**b**) Inventory Stock, (**c**) WIP Stock, (**d**) Backlog Stock

The inventory in this policy for this demand scenario follows a similar behavior as the sudden change demand scenario. The inventory stock level drops gradually from its initial value of 20 K products to oscillate around a mean of 9 K products (Fig. 6b). The adjustment of inventory oscillates with the second highest amplification among other policies to fill the gap between desired level and current level of inventory (Fig. 7b).

The shipment rate oscillates around a mean of 9 K products (Fig. 7d), which is less by 10% of the required mean due to the considered Utilization Level of the system. This leads to a continuous gradual increase in the backlog level (Fig. 6d) and thus a responsiveness performance that is below expectation. The WIP level also has a higher value due to the same reason explained in the previous scenario (Fig. 6c).

The general assessment for the performance of this policy under this cyclic demand scenario shows that the only advantage of this policy is the cost savings in the amount of required capacity scalability. It did not show good performance

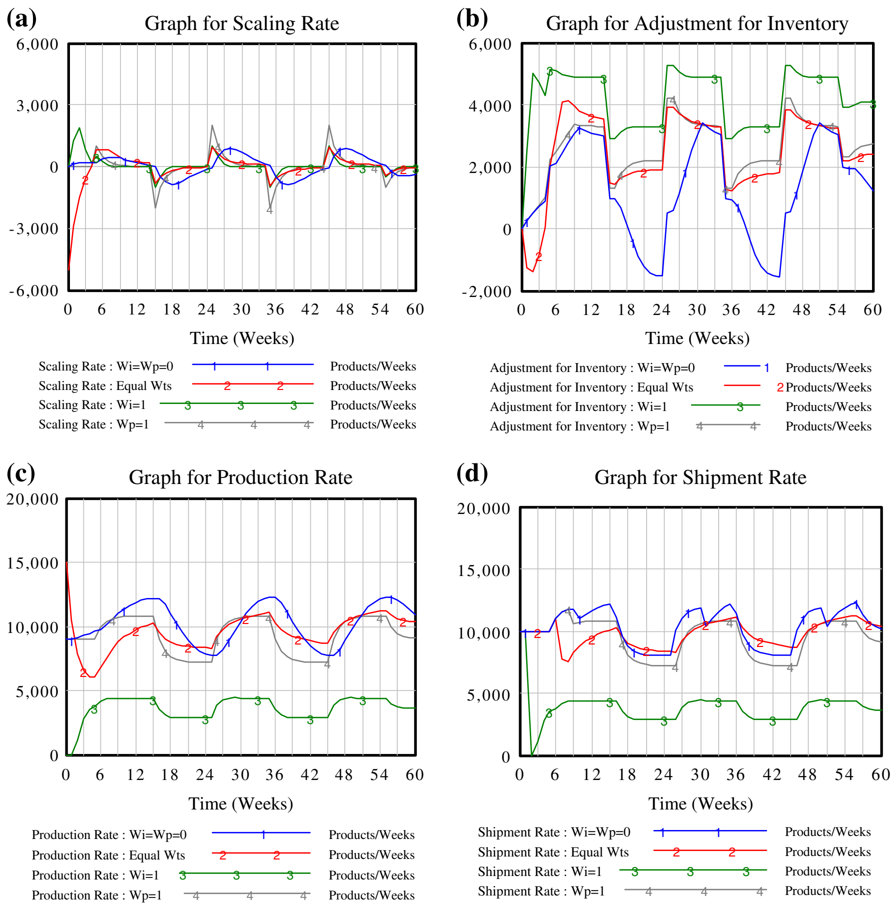

Fig. 7 Dynamic response for different rates in the developed RMS capacity scalability model in a sudden demand scenario. (**a**) Capacity Scaling Rate, (**b**) Inventory Adjustment Rate, (**c**) Production Rate, (**d**) Shipment Rate

considering other investigated measures, such as scalability effort, responsiveness and WIP level.

4.2.2 Inventory based scalability policy

The behavior of the inventory-based policy in this demand scenario is similar to the one with sudden change demand scenario, with the only difference being the cyclic pattern, and thus the same analysis applies.

Although from a dynamic behavior perspective this policy has the lowest cyclic fluctuation (i.e., smallest amplification), the performance measures used for the assessments (including responsiveness, WIP level and cost) show that this policy is the worst among all other considered policies.

4.2.3 WIP-based scalability policy

In contrast to the superior performance of this policy in the sudden change demand scenario, it shows unfavorable dynamic behavior with cyclic demand scenario affecting different performance measures. In terms of capacity level (Fig. 6a) and production rate (Fig. 7c), the highest fluctuation is witnessed in this policy affecting the resulting scalability cost of the RMS system. This fluctuation was reflected in the scalability rate that has the highest value of overall capacity scalability level (Fig. 7a).

The worst performance of this policy is demonstrated with inventory levels as a performance measure by the very high oscillation of both the inventory level (Fig. 6b) and inventory adjustment rate (Fig. 7b). This undesirable dynamic behavior (resulting from the oscillation of the production rate explained earlier) impacts both profit and customer service levels of the system.

However, this policy enjoys the best performance in terms of backlog level and thus responsiveness (Fig. 6d). This can be justified since the shipment rate closely follows the demand cycles (Fig. 7d). In addition, and as expected, the WIP level of this policy is the lowest among other polices and equal to the required level of 40 K products $\pm$ 8 K (Fig. 6c).

The WIP-based capacity scalability policy in the cyclic demand scenario resulted in the lowest performance in terms of the cost and effort required for achieving the required capacity scalability and inventory control while exhibiting the best performance in terms of responsiveness and WIP control.

4.2.4 Hybrid scalability policy

This policy shows the best performance in terms of capacity scalability effort and cost within this demand scenario. Although it is slightly better than the demand chasing policy in terms of capacity level (Fig. 6a) and production rate (Fig. 7c) by having lower amplitudes, it is far better in terms of scaling rate (Fig. 7a).

Concerning inventory and inventory adjustment rate, this policy exhibits a good dynamic performance (excluding the inventory-based policy as it is far beyond the required inventory level) by having the lowest amplitude among other policies (Figs. 6b and 7b).

Although the backlog level is higher than the required, which affects the responsiveness level of the system, it is the second best level in terms of value and steadiness after the one for the WIP-based policy (Fig. 6d). In addition, the shipment rate in this policy has the lowest amplitude among other policies (Fig. 7d). The WIP level also shows a steady and low amplitude behavior ranking the second after the WIP-based policy (Fig. 6c).

In general, the hybrid policy demonstrated a better performance in this demand scenario compared with the sudden change scenario. Although it was the second best in terms of responsiveness and WIP level performance measures, it exhibited a superior dynamic behavior in terms of capacity and inventory levels reducing the cost and effort of capacity scalability of the make-to-order reconfigurable system.

The major reason behind such favorable behavior is the ability of this policy to combine competing measures of responsiveness and cost effectiveness triggered by the considered fluctuating demand.

5 Summary and conclusions

This paper presented simulation results and analyses aiming at helping capacity scalability planners in Reconfigurable Manufacturing Systems to investigate the best scalability policy for various demand scenarios. Modeling was based on a system dynamic approach to better reflect the dynamic nature of both modern market demand patterns as well as the capacity scalability process. The paper contributes to the knowledge of capacity scalability management in make-to-order RMS by considering multiple performance measures that were not considered earlier in this specific field. The measures considered were the scalability effort, inventory level, WIP level and backlog level. These multiple measures were selected due to their direct impact on both cost and responsiveness, which are the main RMS drivers. Table 2 summarizes the recommendations of the different policies under the different demand scenarios using the considered performance measures.

Several dynamic results were demonstrated for the considered performance measures. These results can be classified into either new findings or conventional conclusions that were confirmed through the developed dynamic model for capacity planning. The latter category of results also acts as a validation of the proposed model.

Based on the presented analysis, some new findings are highlighted and can be used by manufacturing systems operation planners to make policy recommendations for capacity scalability in make-to-order RMS as follows:

1. In a sudden change demand pattern that will eventually become steady (if this can be forecasted) and/or in demand patterns where planners reactively respond to sudden changes, the WIP-based capacity scalability policy would be suitable to adopt. It is important to note that this policy scales capacity to maintain a WIP level that is based on both demand and lead-time of the manufacturing system.
2. In a fluctuating demand scenario, adopting a policy that partially accounts for demand, inventory and WIP levels leads to the best results in terms of the

Table 2 Summary of recommended capacity scalability policies'

Measure/Demand	Sudden change	Fluctuating
Capacity level	Demand chase policy	Hybrid policy
WIP level	WIP based policy	WIP based policy
Inventory level	Inventory based policy	Hybrid policy
Backlog level	WIP based policy	Hybrid policy

considered performance measures. This policy is referred to as the hybrid policy. It is important to note that this result applies for the selected values of weights of this policy. Detailed sensitivity analysis would be required to generalize this recommendation.

3. More effort, on both the technical as well as the managerial levels, is required to decrease the delay in achieving the required scalability to enhance the responsiveness of RMS. This was clear in the observed phase lag between demand and the response of the capacity scalability level.
4. Although we are considering RMSs where the basic philosophy is to exactly match the demand, the demand chasing scalability policy was not the best policy to be adopted for capacity scalability (as intuitively expected) when scalability effort, WIP level, inventory level and backlog level are considered as performance measures.

Among the conventional conclusions that were confirmed through the presented results using the dynamic analysis approach were that:

- Through the difference between the production rate and the demand as well as the accumulated backlog in different policies, it was clear that the manufacturing system utilization level affects the responsiveness of the system. Hence, in RMS where there is no need for slack capacity to hedge against uncertainties, manufacturers should aim to maximize the system utilization.
- The inventory-based policy is not recommended for the studied make-to order reconfigurable system under any demand pattern. This aligns with the conventional wisdom that keeping a target inventory level is not a suitable capacity management policy for make-to-order manufacturing systems. However, totally neglecting the inventory level was also shown to be inefficient in turbulent demand and, thus, the partial accounting for inventory, as in the hybrid policy, is a suitable approach for these systems.

An important part of the assessment approach in this paper was to point to the different trade-offs in the capacity scalability planning in RMS. Although these trade-offs are qualitatively or intuitively known, the ability of the proposed model to quantify them serves to provide better insight about the magnitude of the required balance in these trade-off decisions. The two major trade-offs highlighted were:

- The trade-off between the competing objectives of the RMS paradigm, which are cost and responsiveness. It was clear from the dynamic behavior analysis of different policies that keeping the capacity at a level that fully satisfies the demand was achieved at the expense of the effort and costs reflected in the capacity scaling rate of the system (in terms of its magnitude and frequency).
- The trade-off between dynamic stability of the manufacturing system, which affects the performance of the system and the cost. Keeping high levels of capacity and inventory stocks stabilizes the system against demand changes, but again at the expense of cost.

The inclusion of multiple performance measures and different demand patterns in the reported assessments makes them applicable to various operational strategies in make-to-order RMS systems, depending on the adopted market strategy. In addition, although the presented assessment focused on capacity scalability policies in modern RMS, other make-to-order systems that adopt dynamic capacity management policies (such as flexible manufacturing systems) can benefit from the different insights that these assessments provided. Finally, further work is required to generalize this dynamic analysis for capacity scalability in RMS by including systems other than make-to-order with dynamic capacities.

Acknowledgments The contributions from the Canada Research Chairs (CRC) program and the Natural Sciences and Engineering Research Council (NSERC) of Canada, in support of this research, are greatly acknowledged.

References

Anderson E, Morrice D, Lundeen G (2005) The "physics" of capacity and backlog management in service and custom manufacturing supply chains. Syst Dyn Rev 22/3:217–247

Asl R, Ulsoy A (2002) Capacity management via feedback control in reconfigurable manufacturing systems, Proceeding of Japan-USA symposium on flexible manufacturing automation, Hiroshima, Japan

Baines T, Harrison D (1999) An opportunity for system dynamics in manufacturing system modeling. J Prod Plan Control 10/6:542–552

Deif A, ElMaraghy W (2006) A control approach to explore capacity scalability scheduling in reconfigurable manufacturing systems. J Manuf Systems 25/1:12–24

Deif A, ElMaraghy W (2007) Integrating static and dynamic analysis in studying capacity scalability in RMS. Int J Manuf Res 2/4:414–427

Duffie N, Falu I (2002) Control-theoretic analysis of a closed loop PPC system. Ann CIRP 52/1:379–382

ElMaraghy H (2005) Flexible and reconfigurable manufacturing systems paradigms. Int J Flexible Manuf Syst. Special Issue Reconfig Manuf Syst 17/4:261–276

Evans G, Naim M (1994) The dynamics of capacity constrained supply chains. Proceedings of international system dynamics conference, Stirling, pp 28–35

Forrester JW (1961) Industrial dynamics. MIT Press, Cambridge, MA

Fowler A (1999) Feedback and feedforward as systematic framework for operation control. Int J Oper Prod Manage 19/2:182–204

Goncalves P, Hines J, Sterman J (2005) The impact of endogenous demand on push-pull production systems. Syst Dyn Rev 22/3:217–247

Helo P (2000) Dynamic modeling of surge effect and capacity limitation in supply chains. Int J Prod Res 38/17:4521–4533

Holt CC, Modigliani F, Muth JF, Simon HA (1960) Planning production, inventories, and work force. Prentice-Hall, Englewood Cliffs, NJ

John S, Towill DR, Naim M (1994) Dynamic analysis of a WIP compensated support system. Int J Manuf Syst Des 1:283–297

Kim J-H, Duffie N (2004) Backlog control design for a closed loop PPC system. Ann CIRP 54/1:456–459

Kim J-H, Duffie N (2005) Design and analysis for a closed loop capacity control of a multi workstation production system. Ann CIRP 55/1:470–474

Koren Y, Heisel U, Jovane F, Moriwaki T (1999) Reconfig Manuf Syst, Ann CIRP Vol. 48/2/1999

Luss H (1982) Operation research and capacity expansion problems: a survey. Oper Res 3/5:907–947

Manne Alan S (1967) Investments for capacity expansion, size, location, and time-phasing. The MIT Press, Cambridge, MA

Mehrabi M, Ulsoy G, Koren Y (2000) Reconfigurable manufacturing systems: Key to future manufacturing. J Int Manuf 11:403–419

Nyhuis P (1994) Logistic operating curves—A comprehensive method for rating logistic potentials, EURO XIII/OR36, University of Strathclyde, Glasgow

Sethi AK, Sethi PS (1990) Flexibility in manufacturing: a survey. Int J Flexible Manuf Syst 2:289–328

Sethi SP, Thompson GL (2000) Optimal control theory: applications to management science and economics. Kluwer, Boston, MA

Sterman JD (2000) Business dynamic—systems thinking and modeling for a complex world. McGraw-Hill, New York

Towill DR (1993) System dynamics—background, methodology, and applications—part 1: background and methodology. Comput Control Eng J 4/5:201–208

Towill DR, Del Vecchio A (1994) The application of filter theory to the study of supply chain dynamics. Prod Plan Control 5:82–96

Wiendahl HP, Breithaupt J (1999) Modeling and controlling of dynamics of production system. J Prod Plan Control 10/4:389–401

Wiendahl HP, Breithaupt J (2000) Automatic production control applying control theory. Int J Prod Econ 63:33–46

Wiendahl H-P, ElMaraghy HA, Nyhuis P, Zaeh M, Wiendahl H-H, Duffie N, Kolakowski M (2007) Changeable manufacturing: classification, design, operation, Keynote Paper, CIRP Annals, 56/2

Wikner J, Towill DR, Naim M (1991) Smoothing supply chain dynamics. Int J Prod Econ 22:231–248

Dr. Ahmed M. Deif received his Ph.D. and M.A.Sc in Industrial and Manufacturing Systems Engineering from the University of Windsor in (2006) and (2003) respectively. He received his B.Sc. from The American University in Cairo (1999). He currently works as a post doctorate fellow at the intelligent manufacturing systems centre at the University of Windsor. His research interests are in manufacturing capacity management, manufacturing planning and control and dynamic analysis of manufacturing systems. He is a research affiliate of the International Academy for Production Engineering (CIRP).

Hoda A. ElMaraghy, Ph.D., P.Eng. is currently Canada Research Chair in Manufacturing Systems, Professor of Industrial and Manufacturing Systems Engineering and Director of the Intelligent Manufacturing Systems (IMS) Centre at the University of Windsor. She is a Fellow of the Canadian Society of Mechanical Engineers (CSME), Fellow of the Society of Manufacturing Engineers (SME), and Fellow of the International Academy for Production Engineering (CIRP) and Member of the American Society of Mechanical Engineers (ASME). Her teaching and research interests are in manufacturing systems design, modelling, flexibility, reconfiguration, control, quality, inspection, assembly, process planning and robotics. Dr. Hoda ElMaraghy published more than 350 papers in scientific and professional journals and conferences, edited two books and contributed to 13 other books. She is a member of Editorial Boards of four International Scientific and Engineering Journals. She received a mechanical engineering Bachelor degree from Cairo University, Egypt and Master's and Ph.D. degrees from McMaster University, Canada where she became a Professor and founding Director of the Flexible Manufacturing Systems Centre until joining the University of Windsor as the Dean of Engineering in 1994.