

Tilburg University

## Performance metrics in supply chain management

Kleijnen, J.P.C.; Smits, M.T.

*Published in:*  
The Journal of the Operational Research Society

*Publication date:*  
2003

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*  
Kleijnen, J. P. C., & Smits, M. T. (2003). Performance metrics in supply chain management. *The Journal of the Operational Research Society*, 54(5), 507-514.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# Performance metrics in supply chain management

Jack PC Kleijnen\* and Martin T Smits

*Tilburg University (KUB), LE Tilburg, The Netherlands*

This survey paper starts with a critical analysis of various performance metrics for supply chain management (SCM), used by a specific manufacturing company. Then it summarizes how economic theory treats multiple performance metrics. Actually, the paper proposes to deal with multiple metrics in SCM via the balanced scorecard — which measures customers, internal processes, innovations, and finance. To forecast how the values of these metrics will change — once a supply chain is redesigned — simulation may be used. This paper distinguishes four simulation types for SCM: (i) spreadsheet simulation, (ii) system dynamics, (iii) discrete-event simulation, and (iv) business games. These simulation types may explain the bullwhip effect, predict fill rate values, and educate and train users. Validation of simulation models requires sensitivity analysis; a statistical methodology is proposed. The paper concludes with suggestions for a possible research agenda in SCM. A list with 50 references for further study is included.

*Journal of the Operational Research Society* (2003) **00**, 000–000. doi:10.1057/palgrave.jors.2601539

**Keywords:** supply chain; logistics; performance measurement; balanced scorecard; simulation; system dynamics

## Introduction

Performance metrics are treated very differently in business practice and economic theory (see the special *JORS 2002* issue edited by Shutler and Storbeck,<sup>1</sup> especially the contribution by Harrison and New<sup>2</sup>). We shall give an example of this practice, for a specific manufacturing company (for confidentiality reasons, the name of the company is not revealed). Next, we shall summarize how economic theory treats multiple performance measures, including graphical presentations such as Kiviat and spider graphs. Then we account for multiple performance metrics via the BSC, which (by definition) measures customers, internal processes, innovations, and finance. Subsequently, we shall explain how to forecast the consequences of business process redesign (BPR) for the measurements on this balance scorecard (BSC): the forecasting tool is simulation, which may be spreadsheet simulation, system dynamics, discrete-event simulation, or business games. Finally, we shall present our conclusions, suggesting a research agenda for performance metrics in supply chain management (SCM).

## Business examples of logistical performance metrics

One division of a large multinational company evaluates the logistical performance of its SCM systems — and hence its managers — through five key performance metrics. These metrics are measured each month — for each specific

product. The metrics are defined as follows (we paraphrase the original wording).

- (i) *Fill rate*: The percentage of orders delivered ‘on time’; that is, no later than the delivery day requested by the customer.
- (ii) *Confirmed fill rate*: The percentage of orders delivered ‘as negotiated’; that is, delivered no later than the day agreed between the customer and the supplier (the supplier may discover — upon planning a specific order — that the requested day cannot be realized).
- (iii) *Response delay*: The difference (say)  $d$  between the requested delivery day (see (i)) and the negotiated day (see (ii)), expressed in working days. Obviously,  $d$  is a positive integer (however, we may also wish to measure early deliveries — not only late deliveries — in which case  $d$  is negative). We add that we may also measure the frequencies of the various delay values, so that we can estimate the statistical distribution of orders with a particular delay value. In this example, however, the managers are interested only in the probability of exceeding a specific threshold value (say)  $d_0$  (related metrics are discussed by Kleijnen and Gaury<sup>3</sup>).
- (iv) *Stock*: Total work in process (WIP). Obviously, the physical product changes as it moves from workstation to workstation, where value is added. The total WIP can be expressed as a percentage of total sales over the preceding  $m$  months; for example,  $m=3$ . Obviously, the smaller this percentage is, the higher the financial metrics will be — at least, in the short run (in the longer run, a small WIP may lead to low fill rates, so customers will terminate their relationship with this company: loss of goodwill).

\*Correspondence: JPC, Kleijnen, Department of Information Systems and Management, PO Box 90153, 5000 LE Tilburg, The Netherlands  
E-mail: kleijnen@kub.nl

- (v) *Delay*: Actual delivery day minus confirmed delivery day. Note that a fill rate (see (i)) less than 100% implies some delay; metric (v) measures the size of that delay. As with metric (iii), management is interested only in the probability of exceeding a specific threshold value.

Another company — Hewlett-Packard (HP) — emphasises the importance of *shared* performance metrics; that is, metrics shared by all companies in the supply chain (SC). More specifically, in their SCM case study at HP, Callioni and Billington<sup>4</sup> (p 38) mention three metrics:

- (i) *fill rate*: percentage of demand filled from available stock;
- (ii) *sales/inventory ratio*: inventory turnover ratio;
- (iii) *sales*.

From these two case studies we conclude that in practice managers use *multiple* performance measures; a single measure does not suffice.

Note that a large variety of performance measures are also used by Liker and Wu<sup>5</sup> to analyse their survey among automakers. Many performance measures are discussed by Beamon<sup>6</sup> and Gunasekaran *et al.*,<sup>7</sup> based on literature reviews. Kaydos<sup>8</sup> surveys many performance measures, but not in an SC context. So, this section raises the issue: how to deal with multiple measures?

### Economic theory on multiple performance metrics

In economics, *utility* is the final performance measure of a system (economic theory treats physical products and services in the same way). So, a product has multiple characteristics (attributes); for example, a computer has CPU speed, memory size, monitor-display diameter, etc. Assume for simplicity that there is a given number of relevant characteristics (say)  $n$  and that these characteristics can be measured on a cardinal scale (other scales would be an ordinal scale — such as the Likert scale in used by Harrison and New<sup>2</sup> in their survey on performance measurement in SCM — and a nominal scale). This gives the measures (say)  $x_1, \dots, x_n$ . Utility theory then postulates that there is a utility functional  $f(x_1, \dots, x_n)$  that maps these attributes into a single cardinal utility  $u$ .

In practice, this utility function  $f$  is approximated through a *simplified* function. Obviously, the simplest approximation is linear:  $u = a_1x_1 + \dots + a_nx_n$ . A simple alternative is to measure utility and attribute values on a logarithmic scale:  $\ln(u) = a_1 \ln(x_1) + \dots + a_n \ln(x_n)$  or  $u = x_1^{a_1} \times \dots \times x_n^{a_n}$ . Kleijnen<sup>9</sup> discusses these so-called *scoring methods* and their use in computer selection. That monograph also covers Kiviat graphs (see next paragraph), empirical utility measurement (based on Keeney and Raiffa<sup>10</sup>), uncertain attribute values (see the measures — such as fill rate — in the preceding section), mathematical programming (including goal programming), fuzzy set theory, etc. Van Schaik<sup>11</sup> also discusses scoring methods for the evaluation of companies.

The *Kiviat graph* alternates good and bad attributes (for example, fill rate, stock, confirmed fill rate, response delay), so an attractive product results in a star-like graph oriented upwards. An example is Figure 1 where the good attributes are on the axes labelled 1, 3, 5, and 7.

Related to the Kiviat graph is the *spider diagram*. The latter diagram is in fact used by the management of the division of the multinational (not HP) discussed in the preceding section. More specifically, this diagram displays the scores on the five performance metrics, after weighing the two scores for response delay (see (iii)) and stocks (see (iv)); Figure 2 is an example. In this figure we show actual numbers, but hide the identity of the metrics; we do not present the actual weighting of the two metrics because we find it a rather arbitrary weighting scheme. The total logistic performance is expressed as the percentage of the surface of the spider diagram covered by the actual scores. We, however, do not consider this percentage to be the final figure of merit; we prefer the BSC approach — discussed next. (Economic analysis of SCM can also be found in Starbird.)<sup>12</sup>

### Balanced scorecard (BSC)

The BSC was introduced by Kaplan and Norton;<sup>13</sup> also see Smith and Goddard.<sup>14</sup> We, however, base this section on Oasis.<sup>15</sup>

Kaplan and Norton compare the BSC to the dials in an airplane's cockpit or a car's dashboard. The BSC is a tool for implementing a business strategy (as we shall explain in this section and the next section). Traditional business performance metrics are only financial; examples are return on investment (ROI) and price/earnings. The BSC, however, distinguishes the following four different types (dimensions) of performance metrics.

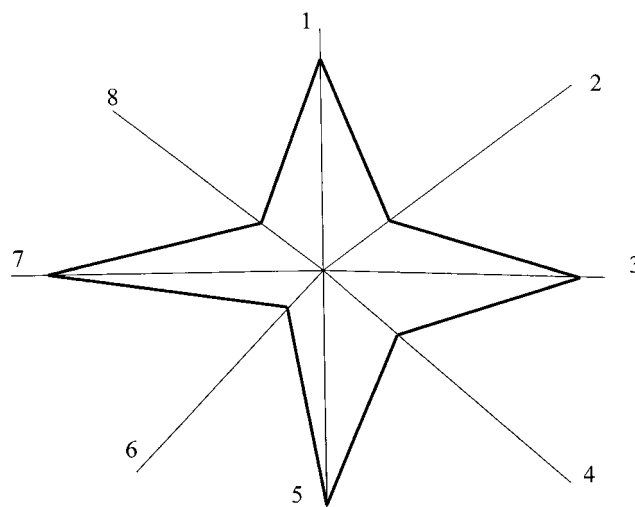


Figure 1 Kiviat graph: an example.

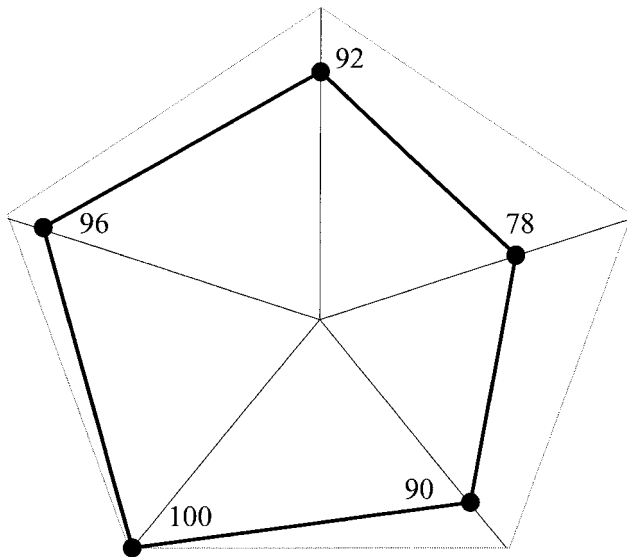


Figure 2 Spider diagram: a case study.

- (i) *Customers*: SCM examples are the fill rate in case of mass products, and conformance to specification in case of ‘built to order’ products.
- (ii) *Internal processes*: Examples are WIP, resource utilization, throughput; also see again Van Schaik.<sup>11</sup>
- (iii) *Innovation*: The division discussed above calculates the ‘best in class’ and the ‘best improver’ among its business units (also see benchmarking discussed by Francis and Holloway<sup>16</sup>); we might add the amount of new IT investments for SCM.
- (iv) *Finance*: WIP may be expressed in value added instead of physical units; SCM should increase profit, market share, and other financial metrics.

For each type of metric, Kaplan and Norton distinguish five to six *key control variables*; these variables we prefer to call (sub)metrics. For example, four of the five metrics used by the multinational’s division relate to customers: fill rate, confirmed fill rate, response delay, delay (stock is not a customer metric). WIP is an internal metric: neither the customers nor the commercial manager care about the company’s WIP; the factory manager, however, does want to minimize WIP.

We also refer to Kleijnen’s<sup>17</sup> case study on a decision support system (DSS) for production planning in a Dutch factory. In that case study, the client gave 37 performance metrics originally. Upon further questioning, however, he reduced these 37 to only two metrics. The latter metrics illustrate the classical friction between the commercial manager and the plant manager: one metric quantified customer lead time, and one quantified the number of productive hours in the plant, which excludes hours needed to switch among product types.

Van Schaik<sup>11</sup> (p 41) points out that the financial metrics — see (iv) above — are more *ex post*, whereas the other

metrics are more *ex ante* — see metric (iii), innovation. In the next section, we shall discuss how to forecast the various metrics.

The BSC was developed outside SCM, so the challenge is to develop a BSC for the companies in a specific SC, given its ‘environment’. For example, does the BSC for a specific company depend on the upstream/midstream/downstream position of that company in the SC? Brewer and Speh<sup>18</sup> also discuss BSCs for SCM.

In a case study, Ashayeri *et al.*<sup>19</sup> consider the use of a BSC in a specific SC within the automotive industry in Europe. In another case study, Olhager *et al.*<sup>20</sup> mention that Ericsson uses the BSC for its SC. And whereas the division discussed above uses a spider diagram, another division of the same company uses a BSC (lately we heard that the former division will switch to a BSC, in the near future).

We emphasise that recently the environment dramatically changed: the global economy is no longer growing. Consequently, markets are now buyers’ markets. We conclude that in the BSC, the customer metric (metric (i)) will become crucial (remember that in a SC, one company’s customer may be another company’s supplier). And survival in the new hostile environment will become a challenge, so improving the internal operations (metric (ii)) and stimulating SC innovations (metric (iii)) — possibly through information technology (IT) — will become necessary. If the SC partners succeed in improving these three metrics, then their financial position (metric (iv)) will improve — at least relative to companies outside this particular SC.

The performance problem becomes simpler when the BSC metrics are *shared* by

- (i) all stakeholders (managers, employees, customers, suppliers, banks, etc);
- (ii) all business units within a company’s division, and all divisions within a company;
- (iii) all companies in the SC.

These different BSC users may place a specific metric (for example, fill rate) in different dimensions (say, customer and internal operations, respectively). Even if these users cannot agree on shared metrics, then each user should at least know which metrics — including definitions — are used by the other stakeholders. Ashayeri *et al.*<sup>19</sup> give an example of shared performance metrics (for example, CO<sub>2</sub> emission) imposed by the European Community (EC).

Obviously, metrics may be *correlated*, either positively or negatively. For example, higher WIP increases costs so it decreases profits, in the short run. Higher WIP, however, definitely increases the fill rate, which increases goodwill so market share may increase — in the long run. Consequently, a manager must account for constraints on various metrics (as in mathematical programming); and different stakeholders may make compromises to create win–win situations. As the HP case study demonstrated, shared metrics facilitate consensus among SC partners.

Moreover, in practice the various SCs become a *value net*; for example, the manufacturing SC and its supporting IT chain are intertwined. We postulate that each company is an independent economic — and legal — entity (agent) so it should have its own BSC; that is, we do not argue for a single BSC for the whole SC. However, communication and coordination within this network require *shared meaning*; that is, the terminology (language) needs standardisation (the necessity of standardisation of non-financial metrics is also discussed by Van Schaik<sup>11</sup> (p 49). An example is the traditional balance and profit and loss account: to determine whether the SC results in a win-win situation, the partners need to agree on the definitions of profits, depreciation (replacement or historical value), etc.

Note that the SC literature seems to disagree with our opinion about the relevance of metrics for the SC as a whole. For example, Hausman<sup>21</sup> describes a number of metrics expressly designed to monitor performance across the *whole* SC. As an example, he mentions that Compaq measures both its own inventory and the downstream inventories at its distributors. We, however, assert that downstream inventories are important input information (to be shared among the links of the SC) for predicting Compaq's BSC, but these inventories are not relevant output performance measures on Compaq's BSC! (Gunasekaran *et al*<sup>7</sup> and Lai *et al*<sup>22</sup> seem to agree with Hausman.) However, if Compaq has contracts with distributors stipulating target values for downstream inventories, then vendor managed inventories (VMI) become a relevant performance metric on Compaq's BSC.

### BSC and simulation

In the preceding section, we saw how the consequences of a given strategy may be evaluated through a BSC — rather than focusing on financial metrics only. However, the manager's dashboard (the BSC) displays only the current status: 'driving via the rear mirror'. It is necessary to *forecast* how the BSC will look — once strategic choices have been made and implemented (to continue the dashboard metaphor, we mention that a global positioning system or GPS adds forecasting). This forecasting requires insight into *causes and effects* of the SC performance: which inputs (or factors) significantly affect which outputs (metrics)? Fortunately, simulation can help to understand causality: simulation is a methodology that does not treat a system (for example, a SC) as a black box. Actually, simulation can help to design alternative strategies, and to quantify their costs and benefits. (Smith and Goddard,<sup>14</sup> however, completely ignore simulation in their discussion of 'analytical techniques' for performance measurement.)

### Simulation types

We distinguish the following four types of simulation.

(i) *Spreadsheet simulation*: Corporate modelling has become popular with the introduction of spreadsheet software (see Plane<sup>23</sup> and Powell<sup>24</sup>). This type of simulation has made simulation acceptable to managers. A simple example of a spreadsheet relation is

$$\text{newinventory} = \text{old inventory} + \text{production} - \text{sales} \quad (1)$$

This equation illustrates positive and negative correlations between the performance metrics WIP and production, and WIP and sales, respectively. Spreadsheets have been used to implement manufacturing resource planning (MRP). However, to evaluate the resulting MRP proposals, spreadsheet simulation is too simple and unrealistic; see (iii) below.

(ii) *System dynamics (SD)*: SD was developed by Forrester,<sup>25</sup> who viewed companies as systems with six types of flows: materials, goods, personnel, money, orders, and information. Examples of these input flows are production and sales; example output flows are fill rate and average WIP (which are also performance metrics in SCM). Besides flows, he distinguishes stocks, such as WIP at a given point in time. He assumes that managerial control is realized through the changing of rate variables, such as production and sales rates, which change flows — and hence stocks. A crucial role in SD is played by the *feedback* principle: a managerial target value for a specific performance metric is compared with its realization, and in case of undesirable deviation, the manager takes corrective action. The SD model is first formulated as an *influence diagram*, which is a picture with these stocks and flows (each flow type may have its own colour in this diagram) and feedback loops. A change of strategy may be modelled as a change in this picture: change some feedback loops, etc. For example, management may decide to change performance metrics, namely to switch from an internal metric (say, WIP) to a customer performance metric (say, fill rate). An alternative is to change only the target value of the current performance metric, so the old feedback loop remains in place: quantitative instead of qualitative policy change. In general, an influence diagram provides a mental map and facilitates *shared meaning* (remember the importance of shared performance metrics, mentioned above). SD is detailed in a recent textbook with 982 pages, namely by Sterman.<sup>26</sup>

In fact, Forrester had already employed a model for the following SC — without using the term 'SC'. There are four links in the SC, namely retailer, wholesaler, distributor, and factory. Forrester studied how the SC links react to deviations between actual and target inventories. He found that 'common sense' strategies may amplify fluctuations in the demand by final customers — up the SC. Later, Lee *et al.*<sup>27</sup> identified this amplification as one of the *bullwhip* effects; also see Beamon<sup>6</sup>.

We emphasise that this bullwhip effect implies that companies should monitor the *final* customers in the SC, not only their immediate customer (the next link in the chain). We may illustrate this phenomenon through the

metaphor of ‘driving a truck in a column’: traditionally, each driver sees only the truck in front, whereas in the new design each driver communicates with the driver at the head of the column (the retailer in the SC). In a SC, however, physical distances are so big that electronic data interchange (EDI) may be used for business-to-business communications. In turn, EDI may be replaced by the Internet and extranets. These IT nets may allow access to the (decentralised) databases of the individual participants in the SC. (Below, we shall return to these IT issues.)

In another SC case, study Ashayeri *et al.*<sup>28</sup> use SD for the distribution chain of Edisco — the European distribution arm of the US company Abbott Laboratories. For a review of research on SD models for SCM, we refer to Angerhofer and Angelidis,<sup>29</sup> Beamon,<sup>6</sup> and Otto and Kotzab.<sup>30</sup>

(iii) *Discrete-event dynamic system (DEDS) simulation*: A DEDS simulation is more detailed than the preceding two simulation types: it represents *individual events*, such as the arrival of an individual customer order and the departure of a production lot. Moreover, such a simulation incorporates *uncertainties*; for example, customer orders are placed at random points in time and are of random size; machines may break down and require random repair times. The most popular (83 000 copies sold) textbook on this type of simulation is by Law and Kelton.<sup>31</sup>

DEDS simulation is already part of the MRP/ERP toolbox for quantifying the costs and benefits of strategic and operational policies (ERP: enterprise resource planning); see Vollmann *et al.*<sup>32</sup>

Viswanadham and Srinivasa Raghavan<sup>33</sup> state that there has been a ‘tremendous amount of research’ in DEDS simulation of SCM. We now summarize two recent examples of SCM simulations.

Rao *et al.*<sup>34</sup> develop a rapid-response SC for Caterpillar, using simulation — and other management science/operational research (MS/OR) techniques — to construct and analyse different SC configurations. They use the following performance metrics: expedited deliveries, partial backlogging of orders, lead times, inventories, realized sales, lost sales, profits, costs, etc.

A second example is Persson and Olhager,<sup>35</sup> who simulate three alternative designs for a SC in the mobile communications industry in Sweden centred on Ericsson. These authors use five performance metrics: costs, inventory, quality, lead-time (including a so-called scrap-inflated lead-time), and lead-time variability.

(iv) *Business games*: It is relatively easy to simulate technological and economic processes (see Equation (1) above), but it is difficult to model human behaviour. Therefore, it may be more realistic to let managers themselves operate within the simulated ‘world’, which consists of the SC and its environment. Such an interactive simulation is called a business or management game (also see Ten Wolde<sup>36</sup>).

We distinguish two subtypes, namely strategic and operational games.

(a) *Strategic games*: There are several teams of players representing companies that compete with each other in the simulated world. These players interact with the simulation model during (say) five to ten rounds. These games are very popular. The simulation model may be a SD model; a famous example is the beer game, which illustrates the bullwhip effect (see Simchi-Levi *et al.*,<sup>37</sup> Sodhi,<sup>38</sup> and Sterman<sup>26</sup>). The game may also be a corporate, economic, business model that illustrates the effects of prices, sales promotion, and research development decisions on profits (see again Kleijnen,<sup>9</sup> pp 157–186).

(b) *Operational games*: A single team (which may consist of a single player) interacts with the simulation model during several rounds or in real time; that is, it is a game against nature. Examples are games for training in production scheduling. We could not find any publications on these games, but we know that such games do exist; also see the web (<http://www.synthesis.org/>).

#### *More on BSC and simulation*

Which of these four simulation types is needed in SCM depends on the question to be answered. For example, SD does not aim at exact forecasts, but at qualitative insight; SD can demonstrate the bullwhip effect, as Forrester<sup>25</sup> has already shown. DEDS simulation can quantify fill rates, which are random variables. Games can educate and train users, since the players are active participants in the simulated world. Moreover, games can be used in research to study the effects of qualitative factors (such as type of DSS) on profits, etc (see Van Schaik<sup>39</sup>).

All simulation types require *sensitivity analysis* or *what-if analysis*: which factors are critical? In principle, the various SC performance metrics may be affected by many factors, because of the large scale of the SC. In practice, however, we have to come up with a short list of really important factors. Therefore, we developed a statistical methodology for sensitivity analysis of large-scale simulation models. This methodology applies the theory on design of experiments (DOE). Traditionally, simulation analysts change one factor at a time in their sensitivity analysis. DOE, however, uses (fractional) factorial designs (such as  $2^{k-p}$  designs), which give more accurate estimates of first-order factor effects — and enable the estimation of interactions among factors; see Kleijnen.<sup>40</sup> Moreover, this methodology can create a set of scenarios, in a systematic, scientific way; see Van Groenendaal.<sup>41</sup> Finally, in simulation — as opposed to real-world experiments — hundreds of factors may be changed in a controlled way; see Campolongo *et al.*<sup>42</sup>

This sensitivity analysis reveals which information is crucial and should be shared, and which decisions should be coordinated in the SC. This analysis also enables us to test assumptions about the SC and its environment; this helps us to validate the mental model of the world in which SCM operates. Kleijnen<sup>43</sup> discusses how to check whether simulations are realistic models of reality.

BPR is enabled by simulation, because simulation quantifies costs and benefits so optimisation becomes practical. For example, Rosenwein<sup>44</sup> discusses BPR and MRP at Philips and Lucent. He mentions (p 27) that through ‘client/server architecture ... managers shared data and plans’. And continues (p 29): ‘However, the optimisation engine, designed to support manufacturing planning was shelved. ... lessons are: 1. Keep initial models simple. 2. Ensure that data are readily available to support any models that are developed’. BPR and simulation is also discussed by Oasis.<sup>15</sup>

So, optimisation is a major methodological and practical challenge. Actually, a simulation program is a non-linear — possibly random (stochastic) — model with multiple outputs (performance metrics). Angün *et al*<sup>45</sup> minimise a random goal function (for example, expected total costs) with several random constraints (for example, the fill rate should be at least 95%); those authors use Interior Point methods developed in Mathematical Programming. Some commercial simulation software packages (such as Arena) combine tabu search and scatter search (marketed under the name OptQuest) to optimise simulation models.

Moreover, once the optimum solution for the SC is found, the environment will change! Therefore, the ‘optimal’ solution should be *robust*. This robustness issue is related to Taguchi’s approach (which he originally developed for designing cars at Toyota). Robustness has already been investigated for individual companies within the SC by Kleijnen and Gaury.<sup>3</sup> Research on the robustness of the Ericsson SC mentioned above will be reported by Bettonvil, *et al*<sup>46</sup> (for a preprint see Working Paper #10 on <http://center.kub.nl/staff/kleijnen/papers.html>).

Related to this robustness is *flexibility*. Hamblin<sup>47</sup> defines flexibility as ‘the ability to respond effectively to changing circumstances’; he distinguishes two types of flexibility: (i) the capability of reconfiguring the business to meet changing needs, and (ii) what we call ‘robustness’. Also see Olhager *et al*<sup>20</sup> and Gunasekaran *et al*.<sup>7</sup>

Whereas analytical models — such as Starbird<sup>12</sup>’s model and the Petri model in Viswanadham and Srinivasa Raghavan<sup>33</sup> (p 1163) — focus on steady-state performance, simulation may also estimate *transient* performance. Performance metrics do show transient behaviour in dynamic, changing environments! See again Law and Kelton.<sup>31</sup>

We conclude that simulation is a technique that can indeed predict and support the design of SCs in practice, including BPR.

Note that performance evaluation may use — besides simulation — analytical models, case studies, and real-life experiments; see the review of empirical techniques for evaluating IT systems by Adelman<sup>48</sup> and the taxonomic reviews of SCM research by Beamon,<sup>6</sup> Ganeshan *et al*,<sup>49</sup> and Otto and Kotzab.<sup>30</sup>

## Conclusions: suggestions for a research agenda

We reviewed the state of the art in performance evaluation of SCs. Our survey suggests the following major challenges for researchers in SCM.

- (i) Select a specific SC. This SC should consist of so many links that we can distinguish upstream, midstream, and downstream companies. This SC may be a benchmark for similar SCs (again see Francis and Holloway,<sup>16</sup> also see: <http://www.benchmarkingreports.com/>). Later on, we may study other types of SCs; examples are: assembly-to-order *versus* make-to-stock; buyer’s market for the SC’s final product *versus* seller’s market; single SC *versus* value net (including an intertwined IT SC).
- (ii) Determine a list of recommended performance metrics, submetrics, and sub-submetrics. Next, apply the BSC approach to determine the SC’s main metrics. Which BSCs are used by different stakeholders, and how do they relate? How often should these metrics be monitored: per month, per quarter, or in real time (at the strategic level, less frequent measuring of aggregated performance suffices; see Kleijnen<sup>9</sup>)? Which metrics suffice (other metrics may be strongly correlated with the ‘critical’ metrics)? When do the BSC metrics enter a ‘danger’ zone, requiring managerial action (a dashboard’s display may have a ‘red’ zone, similar to the control limits in quality control)? Remember that Hausman<sup>21</sup> and others recommend metrics for the SC as a whole, whereas we assert that such information may be important input for predicting the BSC (output) metrics of a specific company.
- (iii) Design a simulation model that explains how the SC’s performance metrics react to environmental and managerial control factors. The type of simulation (spreadsheet, SD, DEDES, game) depends on the type of questions to be answered by the model. For example, SD — possibly run as a game — suffices for demonstrating the bullwhip effect to SC stakeholders. DEDES is needed to estimate the probability of realizing the pre-specified fill rate — especially in a turbulent environment.
- (iv) Perform sensitivity analysis, optimisation, and robustness analysis of the SC simulation model. Sensitivity analysis helps validate the simulation model, provides insight into the behaviour of the SC, and gives the critical control factors. Optimising these factors in the simulation model may support BPR. In practice, however, it is more important to find robust solutions than the optimal solution. The relation between robustness and flexibility need more research.

Such a research agenda may result in both an integrated methodology for performance evaluation (cost/benefit analysis) of SCs, and general results on the main drivers of these costs and benefits. (The need for research on the role of IT in

SCs is also emphasised by the research agenda in Kauffman and Walden.<sup>50</sup> Another research agenda is proposed by Beamon<sup>6)</sup>

*Acknowledgements*—We thank three anonymous referees and the Joint Editor of *JORS*, and also Jakko Kleijnen (Philips, Sunnyvale), Ryan R Peterson (Instituto de Empresa, Madrid), Robert-Jan Streng (Atos Origin, Utrecht), and Frans van Schaik (Deloitte & Touche, The Hague) for their comments on earlier versions.

## References

- 1 Shutler M and Storbeck J (2002). Part special issue editorial: performance management. *J Opl Res Soc* **53**: 245–246.
- 2 Harrison A and New C (2002). The role of coherent supply chain strategy and performance measurement in achieving competitive advantage: an international survey. *J Opl Res Soc* **53**: 263–271.
- 3 Kleijnen JPC and Gaury E (2001). Robustness in simulation: a practical methodology, with a production-management example. *Eur J Opl Res*, accepted for publication.
- 4 Callioni G and Billington C (2001). Effective collaboration. *OR/MS Today* **28**: 34–39
- 5 Liker JK and Wu Y (2000). Japanese automakers. US suppliers and supply-chain superiority. *Sloan Mngt Rev* **42**: 81–93.
- 6 Beamon BM (1998). Supply chain design and analysis: models and methods. *Int J Prod Econ* **55**: 281–294.
- 7 Gunasekaran A, Patel C and Tittiroglu E (2001). Performance measures and metrics in a supply chain environment. *Int J Opns Prod Mngt* **21**: 71–87.
- 8 Kaydos W (1999). *Operational Performance Measurement; Increasing Total Productivity*. CRC Press LLC: Boca Raton, FL.
- 9 Kleijnen, JPC (1980). *Computers and Profits: Quantifying Financial Benefits of Information*. Addison-Wesley: Reading, MA.
- 10 Keeney RL and Raiffa H (1976). *Decisions with Multiple Objectives, Preferences and Value Tradeoffs*. John Wiley & Sons, Inc.: New York.
- 11 Van Schaik FDJ (2000). *Operationele kengetallen hebben de toekomst bij het beoordelen van ondernemingen*. Oratie: Tilburg University, pp 46–47.
- 12 Starbird SA (2001). Penalties, rewards, and inspection: provisions for quality in supply chain contracts. *J Opl Res Soc* **52**: 109–115.
- 13 Kaplan RS and Norton DP (1992). The balanced scorecard: measures that drive performance. *Harvard Bus. Rev* **70**: 71–79.
- 14 Smith PC and Goddard M (2002). Performance management and operational research: a marriage made in heaven? *J Opl Res Soc* **53**: 247–255.
- 15 Oasis (1999). *Management Ondersteunende Systemen Anno 1999*. Oasis BV: Nieuwegein.
- 16 Francis G and Holloway J (2002). Beyond comparisons — the role for the operational researcher in benchmarking. *J Opl Res Soc* **53**: 283–291.
- 17 Kleijnen JPC (1993). Simulation and optimization in production planning: a case study. *Dec Support Sys* **9**: 269–280.
- 18 Brewer PC and Speh TW (2000). Using the balanced scorecard to measure supply chain performance. *J Bus Logist* **21**: 75–93.
- 19 Ashayeri J, Van Eekelen B and Vosseveld R (2001). Lean & clean; maak het Europese distributie-netwerk e-efficient. *STATOR* **2**: 15–21.
- 20 Olhager J, Person F, Parborg B and Rosen S (2002). Supply chain impacts at Ericsson from production units to demand-driven supply units. *Int J Technol Mngt* **23**: 40–59.
- 21 Hausman WH (2002). Supply chain performance metrics. In: Billington C, Harrison T, Lee H and Neale J (eds). *The Practice of Supply Chain Management*, Kluwer: Boston.
- 22 Lai K, Ngai EWT and TCE Cheng (2002). Measures for evaluating supply chain performance in transport logistics. *Transp Res Part E* **38**: 439–456.
- 23 Plane DR (1997). How to build spreadsheet models for production and operations management. *OR/MS Today* **24**: 50–54.
- 24 Powell SG (1997). Leading the spreadsheet revolution. *OR/MS Today* **24**: 8–10
- 25 Forrester JW. (1961). *Industrial Dynamics*. MIT Press: Cambridge, MA.
- 26 Sterman JD (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill: Boston.
- 27 Lee HL, Padmanabhan V and Whang S (1997). The bullwhip effect in supply chains. *Sloan Mngt Rev* **38**: 93–102.
- 28 Ashayeri J, Keij ER and Broker A (1998). Global business process re-engineering: a system dynamics based approach *Int J Opns Prod Mngt*
- 29 Angerhofer BJ and Angelidis MC. (2000). System dynamics modelling in supply chain management: research view. In: Joines JA et al. (eds). *Proceedings of the 2000 Winter Simulation Conference*, pp 342–351.
- 30 Otto A and Kotzab H (2003). Does supply chain management really pay? *Six perspectives to measure the performance of managing a supply chain*. *Eur J Opl Res* **144**: 306–320.
- 31 Law AM and Kelton WD (2000). *Simulation Modeling and Analysis*, 3rd ed, McGraw-Hill: Boston.
- 32 Vollmann TE, Berry WL and Whybark DC (1997). *Manufacturing Planning and Control Systems*, 4th ed. Irwin/McGraw-Hill: New York.
- 33 Viswanadham N and Srinivasa Raghavan MR (2000). Performance analysis and design of supply chains: a Petri net approach. *J Opl Res Soc* **51**: 1158–1169.
- 34 Rao U, Scheller-Wolf A and Tayur S (2000). Development of a rapid-response supply chain at Caterpillar. *Opns Res* **48**: 189–204.
- 35 Persson F and Olhager J (2002). Performance simulation of supply chain designs. *Int J Prod Econ* **77**: 231–245.
- 36 Ten Wolde H (2000). Building blocks of education. *OR/MS Today* **27**: 12.
- 37 Simchi-Levi D, Kaminsky P and Simchi-Levi E (2000). *Designing and Managing the Supply Chain: Concepts, Strategies, and Case studies*. Irwin/McGraw-Hill: Boston.
- 38 Sodhi MS (2001). Applications and opportunities for operations research in internet-enabled supply chains and electronic marketplaces. *Interfaces* **31**: 56–69
- 39 Van Schaik FDJ (1986). *Een laboratorium van informatiesystemen. Handboek informatieverzorging*. Samson Uitgeverij: Alphen aan den Rijn
- 40 Kleijnen JPC (1998). Experimental design for sensitivity analysis, optimization, and validation of simulation models. In: Banks J (ed). *Handbook of Simulation*. Wiley: New York, pp 173–223
- 41 Van Groenendaal WJH (1998). *The Economic Appraisal of Natural Gas Projects*. Oxford University Press: Oxford.
- 42 Campolongo F, Kleijnen JPC and Andres T (2000). Screening methods. In: Saltelli A, Chan K and Scott EM (eds). *Sensitivity Analysis*. Wiley: Chichester, England, pp. 65–89
- 43 Kleijnen JPC (2000). Strategic directions in verification, validation, and accreditation research: a personal view. In: Joines JA et al (eds). *Proceedings of the 2000 Winter Simulation Conference*.

Q1

Q2



- 44 Rosenwein M (1997). The optimization engine that couldn't. *OR/MS Today* 24: August 26–29.
- 45 Angün E, Gürkan G, Den Hertog D and Kleijnen JPC (2002). Response surface methodology revisited. *Proceedings of the 2002 Winter Simulation Conference*, in press.
- 46 Bettonvil B, Kleijnen JPC and Persson F (2003). *Robust solutions for supply chain management: simulation and risk analysis of the Ericsson case study*. Working Paper: Tilburg University.
- 47 Hamblin DJ (2002). Rethinking the management of flexibility — a study in the aerospace defence industry. *J Opl Res Soc* 53: 272–282.
- 48 Adelman L (1991). Experiments, quasi-experiments, and case studies: a review of empirical methods for evaluating decision support systems. *IEEE Trans Syst Man Cybernet* 21: 293–301.
- 49 Ganeshan R, Jack E, Magazine MJ and Stephens P (1999). A taxonomic review of supply chain management research. In: Tayur S, Ganeshan R and Magazine M (eds). *Quantitative Models for Supply Chain Management*. Kluwer Boston, pp 839–879.
- 50 Kauffman RJ and Walden EA (2001). Economics and electronic commerce: survey and research directions. *Int J Electron Commerce* 5: 25–26.

*Received June 2002;  
accepted December 2002  
after one revision*