PROPOSAL OF A FRAMEWORK FOR THE EVALUATION AND COMPARISON OF PRODUCTION SCHEDULES

R. Bandinelli[1], S. Cavalieri[2], S. Terzi[2]

(1) Università di Firenze, Dipartimento di Energetica "S. Stecco", Italy
Via Lombroso, 6/17. I-50134 Firenze - Ph: +39 055 4796710
E-mail: r.bandinelli@ing.unifi.it

(2) University of Bergamo, Department of Industrial Engineering, Italy
viale Marconi, 5. I-24044 Dalmine BG – Ph.: +39 035 2052384
E-mail: sergio.cavalieri@unibg.it, sergio.terzi@unibg.it

Abstract: Production Planning and Control (PP&C) has been increasingly becoming a critical activity, since competition in the markets is leveraging on a multitude of factors ranging from product quality, to delivery times and pre-sales and after-sales services. Among PP&C activities, scheduling decisions are the final temporal decision-making phase where plant and supply chain managers have to act for fixing any short noticed variations and maintaining satisfying overall production system performances, "assigning scarce resources to competing activities over a given time horizon to obtain the best possible system performance". In particular, lot of work has been done in the past (and is currently on-going) on Performance Measurement for manufacturing systems at a strategic level. However, at a more operative scheduling level, a comprehensive approach seems to be still missing.
In order to provide an answer to this main issue, the paper illustrates the main distinctive features of the PMS-ESS, a performance measurement system for the evaluation of production scheduling systems, and its application in a test case.

Keywords: Performance Measurement, Scheduling systems, Benchmarking.

1. INTRODUCTION

Production Planning and Control (PP&C) has been increasingly becoming a critical activity, since competition in the markets is leveraging on a multitude of factors ranging from product quality, to delivery times and pre-sales and after-sales services. Among all PP&C activities, scheduling decisions are the final temporal decision-making phase where plant and supply chain managers have to act for fixing any short noticed variations and maintaining satisfying overall production system performances, "assigning scarce resources to competing activities over a given time horizon to obtain the best possible system performance" (Kempf et. al., 2000).
In Europe, a Special Interest Group (SIG 4) on Benchmarking and Performance Measures - coordinated by KULeuven and University of Bergamo - has been formed within the European IMS Network of Excellence (IMS-NoE, 2001), in order to provide a solution to improve the PP&C activities performance. The SIG 4 community is provided with the Benchmarking Service (BS), a web-based environment where different types of scheduling solutions could be tested and executed, using a distributed simulation arena, over multiple scenarios and industrial test cases, in order to verify and

validate the performances of the scheduling algorithm. Into the Benchmarking Service environment one (or more) production scheduling systems could be tested and evaluated on an "emulated" production system in order to identify the best scheduling solution for the due test case, but also to evaluate how (and if) a scheduling approach could be applied to different production systems. In the BS idea, an industrial actor provides a description of the production system. The provided test case is then emulated into an automatically generated computational model, which will be connected to one scheduling system, for example provided by academic researchers or by scheduling software vendors. The execution of the emulated plant and control logic is then analyzed in terms of performance measures.

The Benchmarking Service needs the design and development of an adequate Performance Measurement System (PMS), for enabling the benchmarking action itself: different scheduling policies might be compared into the BS over one (or more) industrial test cases, in order to find the best solution. This comparison could be enabled only setting up a PMS reference model specifically addressed to the scheduling problem, but, at the same time, with the large perspective needed by industrial actors for taking decisions.

The Benchmarking Service is structured into three interconnected elements, each related to a specific project objective and integrated on the same web-enabled virtual environment: (i) Test-Bench Assistant (TBA), which is a visual interactive environment for assisting the designer of a test bench case in inputting all the main data of the industrial case, (ii) Test-Bench Emulator and Evaluator (TBE&E), which is a web-based remote emulation/simulation service for the experimentation, testing and performance analysis of submitted scheduling proposals, and (iii) Test-Bench Virtual Library (TBVL), where data are collected. TBA and TBVL are already available on line (www.ims-noe.org/benchmark), while TBE&E prototypes are currently under development (Terzi et *al.*, 2004).

In the future, in the TBE&E module the execution of the plant emulation will be elaborated in a distributed manner: the plant emulation code will reside on the server of the BS, while the execution of the on-line scheduling and control logic could be resident on a client computer, physically distributed on the web.

Within this context, the paper aims to illustrate the implementation of a PMS for production scheduling systems, named PMS-ESS (Performance Measurement System for the Evaluation of Scheduling Solution) - which has been elaborated by the authors taking into account suggestions and ideas discussed during the SIG 4 meetings - and to report its implementation in a test case which is currently carried out in an automotive company.

2. LITERATURE REVIEW

Performance measurement is a large research topic: performance measures are used to evaluate, control and improve production processes, but are also used to compare the performance of diverse organizations, plants, departments, teams and individuals.

In the area of scheduling evaluation, basic and simple performance scheduling metrics are widely accepted and used (e.g. makespan, tardiness, lateness, flow time, setup time, working time…). Usually, even if most of them are defined for a single job (job is the main reference unit of the scheduling phase), they are used in an aggregate way, in order to calculate mean and total value among all processed jobs.

However, though the understanding of what constitutes a "good" production schedule is central to the development of scheduling systems, few works in literature have given contributions on this aspect in a comprehensive way. In particular, only the work of Kempf et a. (2000), provides an exhaustive and theoretical approach to scheduling evaluation. Some interesting contributions to the definition of a PMS–ESS can be found in the research community currently interested in the adoption of advanced scheduling approaches, such as multi-agent systems or soft-computing techniques.

Shortly, two main areas of contribution could be identified:

A) Stochastic measurement and disruption facing– The production floor is not a static environment, but a large variety of dynamic events occurs (e.g. machine breakdowns, deliveries delayed, absenteeism), affecting the feasibility of proposed schedules. The inability to accurately respect proposed schedules is referred to as scheduling nervousness or disruption. Facing with disruptions, a technique can be defined as good if it is capable to guarantee the maintenance of certain desired system characteristics despite fluctuations in the behaviour of its component parts or its environment. In literature, diverse measures have been proposed, like predictability in Bongaerts et al. (1999), or schedule robustness under uncertainty in Mignon et al., or the relative concepts of utility, stability and robustness in Ouelhadi et al (2003). Other contributions on robustness, flexibility and stability assessment of a scheduling solution cam be found in Jensen, Gören and Daniels and Kouvelis (1995).

B) Rescheduling effort - Schedules generated in practice cannot be used for a long time period because of unexpected disruptions and random events. Thus, it is necessary to revise the existing schedule at some points in time. In literature there are several alternative ways to decide on timing of schedule decisions: (i) the periodic scheduling approach, (ii) the continuous scheduling, (iii) the adaptive scheduling or controlled response, or some (iv) hybrid approaches. In literature, some efforts have been spent in order to define measures capable to evaluate the changes caused by rescheduling, like the Hamming Distance or the Schedule overlap in Jensen (2001). In order to investigate diverse dimensions of rescheduling, Cavalieri et al. (2000) have proposed the analysis of scheduling techniques under different conditions: in a stable manufacturing system and in a system with exogenous and/or endogenous disruptive events.

3. PMS-ESS DEFINITION AND SPECIFICATION

The proposal for evaluation of scheduling solutions presented in this paper is a three-layered framework (named PMS ESS).

The three layers are so defined:

- *Effectiveness Domain* - This part of the framework regards measures and indicators for the evaluation of the effectiveness of the production plant according to the control of a scheduling solution. Effectiveness does not stand merely for efficiency of sub-systems (as single machines and workforce). It comprehends the overall set of measures capable to describe how the manufacturing system works in an aggregate way.
- *Robustness Domain* - This domain is composed by a set of indicators capable to provide a systematic measure of the robustness level of scheduling solutions. Robustness stands for the ability of scheduling systems to perform graceful degradation of their performance in face of external or internal disruptions.
- *Flexibility Domain* - This domain allows users of the Benchmarking Service to get relevant results of a series of experimental tests conducted with the same scheduling system in different production environments. The set of measures pointed out in this domain tries to provide an answer to 'how the scheduling solution acts in a larger or different type of manufacturing system or according to a different production plan?'

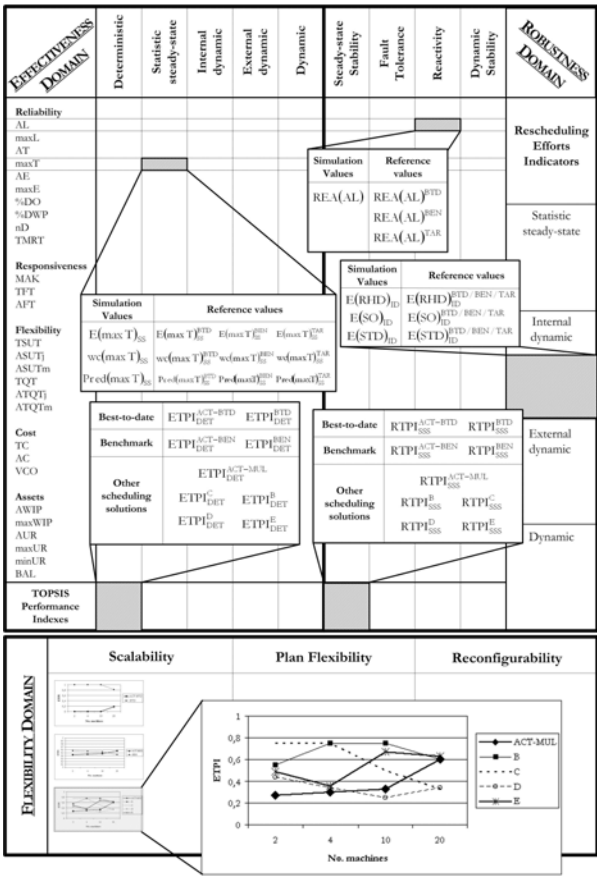

Figure 1 – Overview of the PMS ESS

The architecture of the framework is a general structure designed to be adopted for diverse typologies of manufacturing systems, while the contents are "lay-out dependent". In fact according to the peculiarity of a production lay-out (e.g. manufacturing cells, transfer lines, job-shop, docks,….), it is necessary to define a specific array of measures. During the first development phase of the PMS ESS, a set of measures for a job-shop environment has been designed and implemented in the Test-Bench Assistant.

3.1 Effectiveness Domain

Effectiveness Domain shows the effectiveness of the scheduling solutions, i.e. how the specific manufacturing system performs following the plan proposed by the scheduling system in a steady-state situation. The categories chosen to organize the series of measures are drawn from the SCOR (Supply Chain Council, 2003) framework in terms of *reliability*, *responsiveness*, *flexibility*, *cost*, *assets*.

- *Reliability*, quoting the definition of SCOR, can be defined as the ability to deliver 'the correct product, to the correct place, at the correct time, in the correct condition and packaging, in the correct quantity, with the correct documentation, to the correct customer'.
- *Responsiveness* is the speed at which a manufacturing system provides products to the customer.
- *Flexibility* is the agility of a manufacturing system in responding to market changes in order to gain or maintain competitive advantage.
- *Cost* category, according with the SCOR definition, lists those costs associated with operating the manufacturing system.

Our work focuses the attention and the applicability to a job-shop environment, and the following tables show some of the defined indicators with a short description of them.

<u>Table 1 Reliability Subdomain</u>

Measure	Description
AL (Average Lateness)	The mean value of Lateness calculated over the jobs
maxL (maximum Lateness)	The maximum value of Lateness among all jobs
AT (Average Tardiness)	The mean value of Tardiness calculated over all jobs
maxT (maximum Tardiness)	The maximum value of Tardiness among all jobs
AE (Average Earliness)	The mean value of Earliness calculated over the jobs
maxE (maximum Earliness)	The maximum value of Earliness among all jobs
%DO (percentage of Delayed Orders)	The percentage of orders/jobs that are completed after the due date
%DWP (percentage of Defective Work-Pieces)	The percentage of work-pieces that must be discarded

nD (number of machine disruptions)	The number of disruption due to machine breakdowns
TMRT (Total Machine Repairing Time)	The sum of repairing times for all machines, where they are not available for production

Given these definitions, we can explain the types of measures calculated through simulation campaigns. This way each of the presented measures will be calculated and presented in five different types, due to five different types of simulation campaigns. These campaigns are named *Deterministic value*, *Statistic steady-state*, *Internal Dynamic*, *External Dynamic*, and *Dynamic*.

In the *Deterministic campaigns*, the value of all the measures is simply the one obtained after a simulation replica in a deterministic environment. Given the deterministic values, a simulation campaign in this case is made up of only one replica.

The *statistic steady-state* value of a measure is calculated after more simulation replicas of one experimental campaign are conducted. This value is statistic, since stochastic variations are allowed and used to emulate a real plant. Nevertheless, the value is also steady-state, since probability distributions describe the length of operations on machine or setup times, but no disruptions can occur. Machines breakdowns and arrivals of new jobs are not considered.

As for the statistic steady-state value, the *internal dynamic* value of a measure is calculated after some simulation replicas. The term 'internal dynamic' means that, during a simulation replica, besides statistic variations in operational variables like length of operations on machine or setup times, also unexpected 'internal' disruptions in a production system can occur, i.e. machine breakdowns. In other terms, with this kind of measures we intend to analyse the behaviour of the scheduling solution whenever internal perturbing events occur.

The *external dynamic value* of a measure is calculated at the end of the external dynamic simulation, in which stochastic variations manufacturing operations and disruptions due to demand nervousness (new orders, cancelled orders, urgent jobs) are allowed. In other terms, with this kind of measures we intend to analyse the behaviour of the scheduling solution whenever perturbing events external to the manufacturing system occur.

The *dynamic value* of a measure is the value that is obtained at the end of the dynamic simulation, where all perturbing events (both external and internal) can occur.

3.2 Robustness Domain

Robustness stands for the ability of scheduling systems to have graceful degradation in face of disruptions. If in the *Effectiveness Domain* a user of the *Benchmarking Service* can verify and quantify the performance of scheduling techniques and methodologies, in the *Robustness Domain* he can analyze how the scheduling system is able to maintain the obtained performance when in the manufacturing system perturbing events or disruptions can occur.

Robustness Indicators have been mainly drawn from an in-depth analysis of the literature. The proposed indicators, however, have an original formulation, and particularly an effort to merge experience from scheduling measurement (e.g. the concept of stability by Onelhadi et al., 2003) and scheduling testing (e.g. the different experimental campaigns proposed in Cavalieri et al., 2000) has been done.

Robustness indicators are four, *Steady-State Stability*, *Fault Tolerance*, *Reactivity* and *Dynamic Stability*, and for each measure defined in the *Effectiveness Domain* of the framework, the four indicators are calculated.

Steady-State Stability is an indicator capable to describe the ability of the scheduler not to degrade its performance.

Fault Tolerance is an indicator that can highlight the behavior of a scheduling system in face of internal disruptions as machine breakdowns.

Reactivity aims to evaluate how much the scheduler is capable to react to demand perturbing events (i.e. the ability not to degrade its performance changing from deterministic to external dynamic environment).

Dynamic Stability is an indicator that provides a user of the Benchmarking Service with a tool for the evaluation of the degree of the scheduler's ability facing with stochastic variations and generic disruptions.

3.3 Flexibility Domain

The *Flexibility Domain* tries to answer to a question: 'how the system acts in a greater or different manufacturing system or according to a different production plan?' One of the quoted definitions of flexibility is the ability to respond effectively to changing circumstances. Changing circumstances can be a change in size of the manufacturing system, in the "type" of manufacturing system or a change in production plan. The user of the *Benchmarking Service*, through this area, will be able to understand how performances of the studied scheduling solution can be affected if such a change occurs. An increase in the number of machines or a change in the production mix, in fact, can not be excluded during the manufacturing environment lifecycle.

In this domain a high degree of customization is taken into account. *Flexibility Domain* is composed like a graphical area, where three main dimensions of flexibility are compared using graphs:

Scalability means the ability of a scheduling technique not to degrade its performance if the size of manufacturing system increases or decreases. A series of simulation/emulation is executed and results are compared with results obtained with other scheduling techniques or considering the best-to-date results.

Plan Flexibility is intended to be the ability of a scheduling technique not to degrade its performance following different production plans. The graphs show the trend of indicators *Plan Flexibility* for the technique under study, comparing it to best-to-date values, benchmark values and result values from other scheduling techniques.

Reconfigurability is the ability of the scheduling technique to act effectively in different manufacturing environments. A user can test the excellence of his/her scheduling methodology in different types of production systems.

4. APPLICATION OF PMS ESS

In order to illustrate how PMS ESS works, an application example is hereafter reported. In particular, the example, applied to a simple test case, compares two different scheduling techniques: *Market-like multi-agent architecture and Multi-agent architecture with supervisor*.

The analysis has been conducted in various simulation campaigns focusing on four scheduling measures of the effectiveness domain: *Average Lateness* (AL), *Average Tardiness* (AT), *Average Flow-Time* (AFT) and *percentage of Delayed Orders* (%DO). The plant emulated is a flow-shop, with four couples of manufacturing machines: 2 lathes, 2 fraises, 2 drills and 2 grinders. Each machine has an infinite capacity buffer. The data considered are the output values of simulation replicas. For each scheduling system and for each simulation campaign, the values are obtained from seven replicas. In the deterministic case, only one replica has been run.

The Market-like multi-agent architecture is the technique to test, whereas the results from the other technique are supposed to be the reference data to perform the comparison with. The following tables show the data calculated for some measure of the reliability axes in the effectiveness domain. Due to the short space available, only the first replica of the *Market like architecture* and of the *MAS with supervisor architecture* is shown.

Table 2 Results for MAS - market like architecture

MAS - Market like	repl	Average Late-ness	Average Tardi-ness	% Delayed Orders	Average Flow-Time
Deterministic	1	-34,50	0,75	5,00	37,40
Steady-state	1	-31,24	1,34	12,01	38,80
	mean	-30,29	1,36	11,26	38,94
	std dev	1,27	0,24	1,70	0,88
	worst-case	-28,85	1,75	13,36	40,43
Internal dynamic	1	-28,24	2,24	16,06	40,54
	mean	-28,50	1,75	13,95	39,40
	std dev	1,21	0,49	2,55	1,58
	worst-case	-27,35	2,39	17,22	40,90
External dynamic	1	-12,35	4,05	29,81	40,04
	2	-12,25	4,51	31,20	40,28
	mean	-12,74	4,52	28,70	39,54
	std dev	1,95	0,98	3,36	1,54
	worst-case	-9,27	6,39	33,66	42,18
Dynamic	1	4,27	9,56	54,52	40,80
	2	5,23	11,25	53,67	41,42
	mean	4,43	10,22	51,67	40,86
	std dev	1,62	1,30	2,09	1,32
	worst-case	7,23	12,43	54,52	43,10

Table 3 Results for MAS-with supervisor architecture

MAS - with supervisor	rep	Average Late-ness	Average Tardi-ness	% Delayed Orders	Average Flow-Time
Deterministic	1	-20,40	2,88	15,44	37,01
Steady-state	1	-15,23	4,19	22,13	38,76
	mean	-16,07	4,82	19,25	38,53
	std dev	1,59	0,90	1,90	0,91
	worst-case	-13,73	6,38	22,13	39,96
Internal dynamic	1	-14,27	4,19	25,60	41,11
	mean	-14,75	4,82	23,67	41,82
	std dev	1,53	0,90	1,87	1,77
	worst-case	-12,26	6,38	25,60	44,91
External dynamic	1	-4,44	6,45	34,82	47,95
	mean	-4,46	6,87	33,33	47,82
	std dev	1,83	1,37	4,15	1,39
	worst-case	-1,42	9,38	39,38	50,03
Dynamic	1	4,59	9,36	45,06	54,45
	mean	5,39	10,30	44,58	53,16
	std dev	2,00	1,85	3,98	2,01
	worst-case	9,03	13,48	49,43	55,98

Firstly, mean, predictability and worst-case values for each measure, for each campaign are calculated. Secondly, Robustness Indicators are carried out. Then reference values (best-to-date and benchmark) for effectiveness values and Robustness indicators are calculated considering the data of the *Multi-agent architecture with supervisor* scheduling techniques. These values are compared with the result of the *Market-like multi-agent architecture* scheduling techniques. At last, the *Flexibility Domain* indicators are calculated. The elaboration has been carried out using a spreadsheet application.

5. CONCLUSIONS

The content of the paper is inherent the area of performance measurement for the establishment of a Benchmarking Service. A Performance Measurement System for the Evaluation of Scheduling Solutions (PMS-ESS) has been proposed. This framework can be used in general for the evaluation of scheduling techniques in simulation environments. In this work, however, only a simple validation of the PMS has been carried out. Further developments deal with a detailed industrial application. For the PMS-ESS, the next research step could be its software implementation in BS and the evaluation by Benchmarking Service's stakeholders.

For the *Benchmarking Service*, the next research step is to provide a collection of industrial test cases, in order to study their Performance Measurement criteria and to individuate 'best practices'.

Currently a real application of the PMS-ESS, is under development in a automotive company. The firm's plant consists of a job shop environment where a family of automotive brakes is produced. The production plan of the company presents a high variability due to the changes in the demand, and because of the high price of the product, stocks have to be avoided or reduced. A subsystem of the plant has been modeled with the TBA, a real production plan has been inserted and the model has been simulated with the TBE&E. In order to validate the above framework, a simulation model in ARENA@ 7.0 has been realized. Once the TBA – TBE&E framework will be tested, the PM will be calculated with several production plan scenarios, in order to evaluate the behavior of the plant and the indicators of the *Effectiveness Domain*.

DEBITS

The content of the paper derives from the efforts spent by authors within the IMS-NoE community, Special Interest Group 4. In particular, authors aims to thank Paul Valckenaers, coordinator of SIG 4, and prof. Marco Taisch, coordinator of IMS-NoE for their support.

REFERENCES

Bititci, U.S. (1995), 'Modelling of performance measurement systems in manufacturing enterprises', International Journal of Production Economics, Vol. 42, pp. 137-147.

Bititci, U.S., Carrie, A.S. and McDevitt, L.G. (1997), 'Integrated performance measurement systems: a development guide', International Journal of Operations and Production Management, Vol. 17, No. 6, pp. 522-535.

Bititci, U.S., Suwignjo, P. and Carrie, A.S. (2000), 'Strategy Management through Quantitative Modelling of Performance Measurement Systems', International Journal of Production Economics, Vol. 69, pp. 15-22.

Bititci, U.S., Carrie, A.S. and Turner, T. (2002), 'Integrated performance measurement systems: Structure and dynamics', in Neely, A.D. (ed), Business Performance Measurement: Theory and Practice, Cambridge University Press, Cambridge, pp. 174-197.

Bongaerts, L., Y. Indrayadi, H. Van Brussel, and P. Valckenaers, "Predictability of Hierarchical, Heterarchical, and Holonic Control". Proceedings of the 2nd International Workshop on Intelligent Manufacturing Systems, Leuven, Belgium, pp. 167-176, 22-24 September 1999.

Cavalieri, S., Taisch, M., Garetti, M. and Macchi, M. (2000), 'An experimental benchmarking of two multi-agent systems for production scheduling and control', Computers in Industry, Vol. 43, pp. 139-152.

Daniels, R., and p. Kouvelis, "Robust Scheduling to Hedge Against Processing Time Uncertainty in Single Stage Production". In: Management Science, Vol. 41, No. 2, pp. 363-376, 1995

Gören, S., "Robustness and Stability for Scheduling Policies in a Single Machine Environment". M.Sc. thesis, Bilkent University, Ankara, Turkey, 2002.

IMS-NoE project, (2003), IMS NoE: Network of Excellence on Intelligent Manufacturing Systems, IMS-2001-00002, IST-2001-65001, www.ims-noe.org

Jensen, M.T. "Robust and Flexible Scheduling with Evolutionary Computation". Ph.D. thesis, Department of Computer Science, University of Aarhus, Denmark, 2001.

Kaplan, R.S. and D.P. Norton "The balanced scorecard - measures that drive performance". In: Harvard Business Review, January-February, pp. 71-79, 1992.

Kempf, K., R. Uzsoy, S. Smith and K. Gary (2000). Evaluation and comparison of production schedules. In: Computers in Industry, Vol. 42, pp. 203-220.

Kim, G., Park, C.S. and Yoon, K.P. "Identifying investment opportunities for advanced manufacturing system with comparative-integrated performance measurement", International Journal of Production Economics, Vol. 50, pp. 23–33, 1997.

Mignon, D.J., S. J. Honkomp, and G. V. Reklaitis. "Framework for Investigating Schedule Robustness Under Uncertainty". In: Computers in Chemical Engineering, Vol. 19, Supplement, pp. S615-S620, 1995.

Ouelhadj, D., Cowling, P. and Petrovic, S. (2003), 'Utility and Stability Measures for Agent-Based Dynamic Scheduling of Steel Continuous Casting', Proceedings of the IEEE International Conference on Robotics and Automation, Taipei, Taiwan, 14-19 September.

Supply-Chain Council, "Supply Chain Operations Reference Model - SCOR Version 5.0", Supply Chain Council, Pittsburgh, Pennsylvania, 2003.

Terzi S., Cavalieri S., Bandinelli R., Tucci M., (2004) Experiences in distributed simulation for benchmarking production scheduling systems, Proocedings of international IMS Forum 2004, May 17th -19th 2004, Villa Erba - Cernobbio - Lake Como (Italy)

Terzi, S., P. Brambilla, and S. Cavalieri, (2003), Evaluation and validation of scheduling systems: issues and trends, In: Proceedings of the 3rd International Workshop on Performance Measurement, Dalmine, Italy, 19-20 June, pp. 155-168.