A Lower Bound Analysis for the Flowshop Scheduling Problem with Makespan Minimization

Bruno de Sousa Alves
Electrical Engineering Department
Polytechnique de Montréal
Montréal, Canada
bruno.alves@polymtl.ca

Carlos Ernani Fries
Department of Production and Systems Engineering
Federal University of Santa Catarina
Florianópolis, SC, Brazil
carlos.fries@ufsc.br

Abstract

Flowshop Scheduling Problem (FSP) is a common but not trivial problem in manufacturing scheduling. Lower bound (*LB*) measures can provide a reasonable estimation considering makespan minimization. This paper deals with an analysis of a *LB* measure comparing with the CDS heuristic and the optimal solution of FSP. Performed simulations varying the number of jobs (*N*) and machines (*M*) with processing times following uniform and exponential distributions show that the discrepancies between the solutions tend to increase until $N<M$ and decrease for $N>M$, with largest discrepancy observed for $N=M$. The divergences tend to be larger when greater variability on processing times is considered.

Keywords
Lower bound analysis, Flowshop scheduling; CDS heuristic.

1. Introduction

In Flowshop Scheduling Problems (FSP), a set of N jobs has to be processed in a predefined same sequence of M machines. The job processing time in each machine is known, and in most cases, the main problem consists to find a schedule that minimizes the makespan, i.e., the time difference between the end of the last job on the last machine, and the start of the first job on the first machine (Taillard, 1990). According to Gupta and Stafford (2006), the makespan minimization is the most studied criteria since the publication of Johnson's seminal paper in 1954. In FSP the sequence of machines of all jobs is known and each operation starts only after the end of the last one. Generally, finding the optimal solution of this problem is often difficult due its complexity regarding computation time. The complexity to find the best processing sequence of *N* jobs over the set of *N!* plans is considered NP-hard (Askin and Standridge, 1993). Scheduling is one of the most important decisions in production control systems. Hence, this paper aims to develop a makespan Lower Bound (*LB*) analysis comparing the LB measure defined by Askin and Standridge (1993) and the performance of a heuristic proposed by Campbell, Dudek and Smith (1970) (CDS heuristic). Further, optimal solutions for FSP problems with reduced number on jobs were evaluated and compared with those obtained with the heuristic and the *LB* measure. The discrepancies between these three makespan measures (*LB*, heuristic and optimal solution) for different number of jobs, number of machines and probability distribution of processing times may be the starting point for the gross estimation of production capacity necessary to process a set of jobs.

2. Mathematical Models and Heuristic Methods

The FSP with makespan minimization has been a recurring subject of studies since Johnson (1954) proposed an exact

procedure for the particular case of the two-machine FSP. According to Gupta and Stafford (2006), since Johnson's seminal work, over 1,200 articles related to the FSP were published until 2006 and today, beyond 2,500 citations could be observed as bibliometric analysis performed by the authors.
The criterion (or measure of performance) set for the scheduling may vary according to industry needs and/or research intent. Minimizing makespan is the most studied criterion (Gupta; Stafford, 2006), however, minimizing deviations from the deadline (Vallada et al., 2008) has also significant attention. Several mathematical programming models have been formulated in attempt to determine exact solutions for this problem. Morton and Pentico (1993) for example, have proposed a mixed linear programming model using the makespan minimization as performance measure. Another mixed integer programming model, suggested by Zhu and Heady (2000) aims to minimize punishments for due date advances and delays. A detailed presentation of several mathematical models facing the FSP is presented by Souza (2009). Without exceptions, these models have in common the fragility of the inherent FSP complexity, which is known as NP-hard. For this reason, solving the FSP with a large number of jobs requires the application of heuristic procedures whose main purpose is the evaluation of good scheduling plans in acceptable search time.
Campbell et al. (1970) suggested a heuristic that uses principles of Johnson's exact algorithm for a flowshop with more than two machines. According to Johnson (1954) if a job has shorter processing time in the first machine, then it has to be inserted in the first position of the sequence of jobs to be processed in the shop; in case a job has the longest processing time in the second machine, then it has to be included as late as possible in the sequence of jobs to be processed in both machines.
The heuristic proposed by Campbell, Dudek and Smith, (or simply CDS heuristic), reduces the original problem with *M* machines in (*M-1*) artificial FSP problems with two pseudo machines each. Thus, the artificial FSP can be solved by Johnson's algorithm. The pseudo machines processing times would result from the additive aggregation of the original machines processing times. The authors suggest solving (*M-1*) problems with two pseudo machines each that are made up of groups of machines as follows:

Problem 1 = {Machine 1}, {Machine M}
Problem 2 = {Machines 1 and 2}, {Machines M-1 and M}
... ...
Problem M-1 = {Machines 1, 2, ... and M-1}, {Machines 2, 3, … and M}

Lower bound measures for combinatorial problems of high complexity have also been object of research for other purposes. Several authors use measures of *LB* to evaluate the performance of their new algorithms comparing results obtained with those minimum values that can be effectively determined, i.e., the LB. In general, the evaluation of a *LB* measure is quite simple. The *LB* ensures that there will not be a solution under the given performance, thus providing a reference for comparing heuristics, and cut additional limits on the branch-and-bound search in mixed linear programming models (integer or binary).
Mastrolilli and Sveson (2009) defined a trivial *LB* measure: If *D* represents the size, (in terms of processing time), of the most time consuming job (dilation), and *C* denotes the time required to process all jobs in the most loaded machine (congestion), the *LB* for the system can be calculated as: $LB=max[D,C]$. On the other hand, Askin and Standridge (1993) proposed an alternative *LB* measure: for each machine *j* a *LB* for the makespan can be given by the sum of: (1) total time of jobs on machine *j*; (2) minimum of the sum of the processing times of each job on the machines upstream of machine *j*; (3) minimum of the sum of the processing times of each job on the downstream machines of the machine *j*. Mathematically, this measure *LB* for each machine *j* is expressed as:

$$L_j = \min_i \left\{ \sum_{r=1}^{j-1} p_{ir} \right\} + \sum_{i=1}^{N} p_{ij} + \min_i \left\{ \sum_{r=j+1}^{M} p_{ir} \right\} \tag{1}$$

where: p_{ij} is the sum of processing and setup times of the job *i* in the machine *j* and L_j the estimation of makespan lower bound considering machine *j*. The lower bound of a shop of *M* machines is then defined by

$$LB = \max_{j=1,M} [L_j] \tag{2}$$

Gharbi and Mahjoubi (2013) proposed variants for the makespan LB measure of FSP with makespan minimization. Salmasi et al. (2011) proposed another *LB* measure for this problem by promoting a grouping of the jobs into small groups which are treated as independent jobs. The authors of these recent proposals for makespan *LB* measures plead for some advantages in reducing time when solving the mixed linear programming models, in the case such measures would be included in the formulation.

By the authors early performed simulations it was shown that the makespan *LB* measure suggested by Askin and Standridge provides, in approximately 93% of the simulated cases, stricter results, i.e., lower bounds with higher values when compared with those obtained with the *LB* measure proposed by Mastrolilli and Sveson (2009). For this reason, the *LB* measure proposed by Askin and Standridge will be considered in this study.

3. Methodology

The methodology is based on the simulation of randomly generated FSPs. According Banks (1998), simulation is a technique of solution given by the analysis of a model that describes the behavior of a system using computer equipment. Simulation allows understanding the dynamics of a system and thus analyzes and predicts the effect of changes taking place on the simulated model. In this study, instances may vary by (a) number of jobs (*N*) being processed, (b) number of machines (*M*) available in the shop and, (c) job processing times.

Figure 1 shows a flowchart of the methodological outline that was formulated to analyze the effects for one random generated instance. For each instance with *N* jobs and *M* machines, its makespan *LB* as well as the makespan obtained with the application of the CDS heuristic are compared. Further, if an optimal solution for the mixed integer programming model of the instance could be found then its makespan is included in the analysis. The mixed integer programming model of the random instance was formulated after Gueret et al. (2000) using *GAMS* code. Simulations for the implementation of the CDS heuristic and *LB* evaluation were developed using *Matlab*.

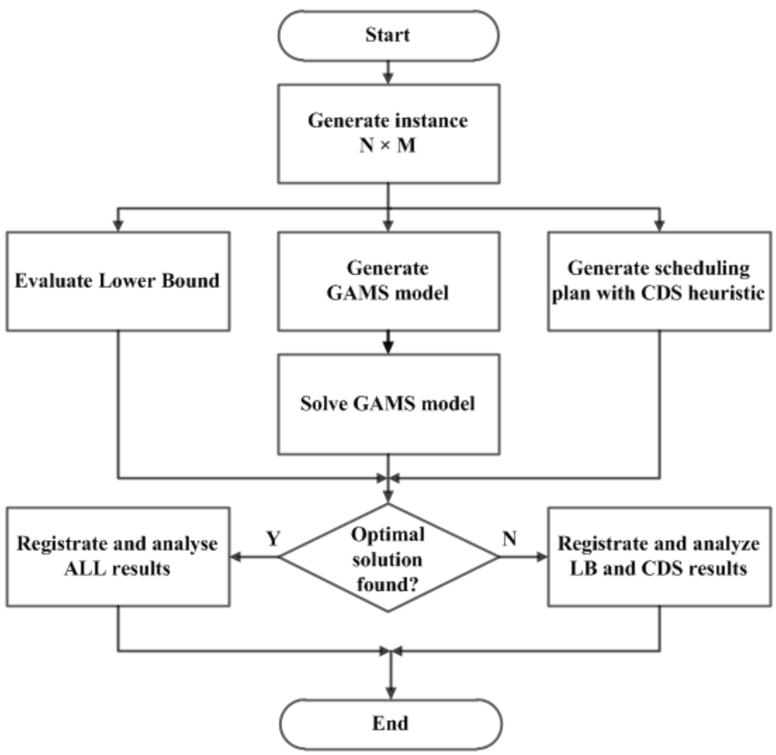

Figure 1. Flowchart of methodological outline

4. Results and Discussions

The performance analysis was evaluated by varying the number of machines (*M*) of the shop, the number of jobs (*N*) and the probability distribution (uniform and exponential) of their processing times in the machines. For each configuration, 1,000 random instances were generated. The makespan *LB* as well as the makespan obtained by the scheduling plan generated by the heuristic CDS were evaluated for each instance. For the same FSP, mixed integer linear programming codes (in *GAMS*) were generated automatically in order to determine the optimal solution for each instance.

4.1. Lower Bound and CDS Heuristic Comparison

Table 1 summarizes the average results obtained for the relationship between the makespan of scheduling plans evaluated with the heuristic CDS and the makespan *LB* for processing times following a uniform distribution. The values listed in Table 1 are plotted in Figure 2 to allow visualization of changes according to the number of jobs on a logarithmic scale.

Table 1. Average ratio between CDS and LB for a selection of number of jobs (N) and machines (M) considering uniformly and exponentially distributed processing times

# jobs (N)	Uniformly distributed processing times				Exponentially distributed processing times			
	M=10	M=50	M=100	M=300	M=10	M=50	M=100	M=300
3	19%	14%	11%	7%	25%	21%	17%	11%
10	21%	27%	23%	15%	30%	45%	39%	26%
50	12%	36%	39%	33%	18%	59%	68%	59%
100	9%	32%	40%	40%	13%	53%	69%	72%
300	5%	22%	32%	44%	7%	36%	56%	80%
500	4%	17%	27%	43%	5%	29%	46%	77%
1000	2%	12%	20%	36%	4%	21%	34%	64%

Results show that keeping the same number of machines and increasing the number of jobs in the shop a higher relative discrepancy of the makespan value obtained with CDS heuristic and the corresponding makespan lower bound can be observed. The maximum discrepancy is reached when the number of jobs is identical to the number of machines in the shop. Considering uniform distributed processing times and 10 machines, for example, the maximum discrepancy of 21% occurs for 10 jobs. Similarly, for 50 machines this point occurs for 50 jobs with maximum discrepancy of 36%; for 100 machines it happens at 100 jobs with maximum discrepancy of 40% while for 300 machines the maximum at 300 jobs is observed with a maximum discrepancy of 44%. As shown in the Figure 2, increasing the number of jobs over the number of machines tends to decrease considerably the relative discrepancy.

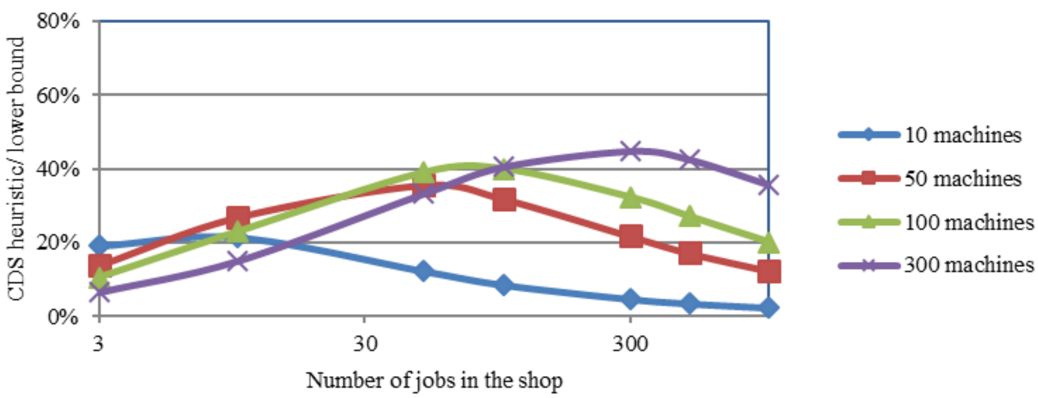

Figure 2. Average ratios between CDS and LB with uniformly distributed processing times

A high idle manufacturing system can be observed in instances with number of jobs less than the number of machines. Inserting more jobs in a relatively idle system tends to increase the use of the machines and therefore a higher relative discrepancy of the CDS makespan and its corresponding *LB* can be observed. In instances with number of jobs higher than the number of machines, the number of possible scheduling solutions is higher. In this case, it should be possible to find a diminishing ratio to the increase in the number of jobs in the system, given that it will be possible to identify

a job that has less processing time on the first machine and put it before other jobs to be processed or otherwise, if the processing time is shorter in the last machine.
Similar generated instances, but exponentially distributed processing times produced results shown in Figure 3. Analogously with the previous situation, the behavior of the ratios is similar but shows higher relative discrepancies than the former. The maximum point is also observed when *N=M*, but at a higher level than that obtained for the analysis, and lower dispersion obtained with uniformly distributed processing times.

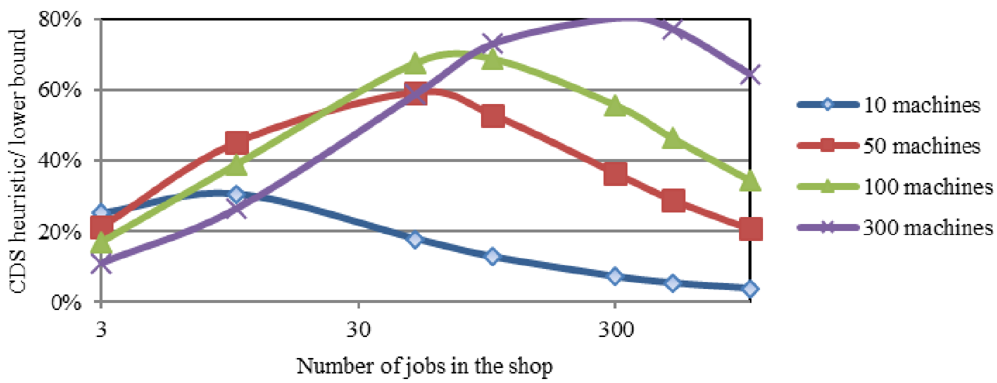

Figure 3. Average ratios between CDS and *LB* for exponentially distributed processing times

Whilst for uniform distributed processing times 10 jobs in 10 machines provided average heuristic solution 21% higher than its *LB*, exponentially distributed processing times lead this gap to 30%. By comparing the graphics of Figures 2 and 3 it can be seen that the relative discrepancy is almost doubled for instances with 300 jobs and 300 machines (=80%), when comparing results obtained with exponentially distributed processing times and those obtained (=44%) for jobs with uniformly distributed processing times. Figure 4 shows the evolution of the average absolute makespan values obtained with the CDS heuristic and its corresponding makespan *LB* for a manufacturing system with 100 machines and progressive increase in the number of jobs. It is obvious that increasing the number of jobs in system, the maximum makespan tends to increase. This increment is most strongly observed for a number of jobs less than the number of machines, as mentioned before. With the progressive increase in the number of jobs from 100 jobs (*N=M=100*), there is a relatively steady makespan growth evaluated with the CDS heuristic. The value of the *LB* also increases, but at a lower rate when compared to the heuristic solutions. However, with more jobs in the system and consequent increase in makespan, the ratio to makespan *LB* tends to decrease.
Problems with many jobs and many machines need more computation time to be solved. The total computational time for the execution of 1,000 instances was 9,181 seconds, or 2 hours and 33 minutes on a regular personal computer. Its indicates that each instance needed about 9.2 seconds for the determination of a scheduling plan by using the CDS heuristic, the evaluation of the *LB*, and the compilation of results.

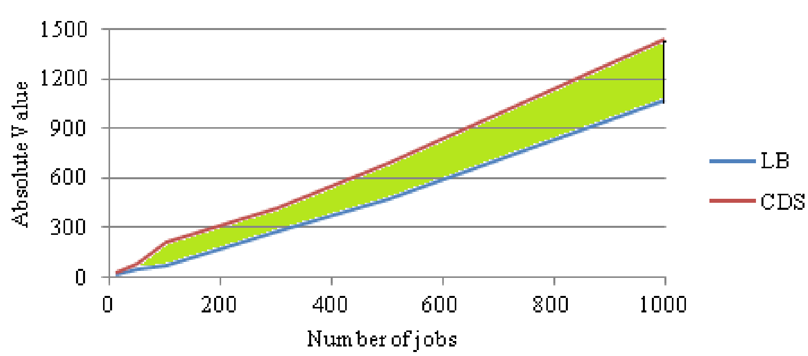

Figure 4. Mean absolute values for the makespan LB and CDS heuristic obtained makespan considering exponentially distributed processing times (*M=100*).

4.2. Lower bound, CDS heuristic and optimal solution

To verify the feasibility of the MIP solver available, a program was developed using Matlab to generate instances that can be solved using a MIP model with GAMS. An automated batch system was also designed to process more than one MIP model in sequence without the need of user interaction. The results were compiled in a table and exported to Excel spreadsheet for further analysis. The limit for the simulation GAMS system time was set at 1,000 seconds. If the system did not find the optimal solution for an instance in this time interval, it abandons the current model and goes to search the optimal solution of the next generated instance.
Initially, 500 instances with 10 jobs and 10 machines were generated and processed in the time limit for the GAMS model. Instances with 10, 20, 40 and 50 jobs, 10 machines and random exponentially distributed processing times were generated and submitted to solve. Results are summarized in Table 2.

Table 2. Comparison between optimal solution (MIP), LB, and CDS heuristic solution for different configurations

Size of Problems	# instances generated	# of solved instances	Relative performance	
			CDS/LB	MIP/LB
10 jobs × 10 machines	500	500	30%	21%
20 jobs × 10 machines	200	92	24%	9%
40 jobs × 10 machines	200	75	14%	1%
50 jobs × 10 machines	120	44	14%	1%

For each configuration, the number of instances and number of optimal solved instances diverge with increasing number of jobs. All instances with 10 jobs and 10 machines were solved in the time limit of 1,000 seconds while for a FSP with 50 jobs and 10 machines, only 44 of 120 generated instances were solved. The relative performances of the heuristic CDS and MIP model with respect to its *LB* are presented in the same frame.
Obviously the optimal solution found for each case must lie between LB and CDS heuristic makespan solution. For problems with 10 jobs and 10 machines, the optimal solution was 21% higher of the *LB*, on average, while the heuristic solution remained above 30% of this same parameter. For problems involving a larger number of jobs, the optimal solution tends to the value of makespan *LB*. On average, MIP provided solutions that were only 0.89% above *LB* for problems with 50 jobs and 10 machines. In this case, the heuristic solution was, on average, 14% above *LB*. This indicates that, although the processing times follow an exponential distribution, the optimal solution is much “closer” to the LB than the CDS heuristic solution.
For FSP with fewer jobs, the heuristic solution is “closer” to the optimal solution. It must be observed that this does not reflect the absolute makespan value of these instances but its relative discrepancies. For shops with many jobs or many machines, the discrepancy in absolute values may be high, although the relationship MIP/LB is probably very low.
As already mentioned, when the number of jobs tends to infinity, the relative discrepancy obtained by the CDS heuristic and its LB tends to decrease. Simulations show that for 1,000 jobs and 10 machines, the heuristic solution is only an average of 3% above the LB, but in this case, the optimal solution could not be evaluated due its complexity taking too much computational time. Since the difference between optimal makespan solution and its LB decreases by increasing the number of jobs - with its solution being always better than the CDS heuristic solution – it can be said that the optimal makespan solution will be proportionally very close to LB for problems with a large number of jobs.

5. Concluding Remarks

FSP with makespan minimization can hardly be optimal solved when the number of jobs involved is considerably high. To undertake the remarkable variety and complexity of FSPs with large *N* and *M* machines (*M*>3), several heuristics and optimization models have been developed throughout the past few decades. Lower bound measures have also been included in complex search algorithms to derive interesting solutions in acceptable time. Most of these procedures consider makespan minimization as natural criterion to generate scheduling plans. In this paper, a study on the relationship between the makespan obtained by applying the heuristic CDS and its corresponding *LB* for different manufacturing instances is presented. Additionally relations between these values and the makespan obtained from exact scheduling solutions were also included in the analysis.

Simulation results show that with the progressive increase of the number of jobs in the shop, the relative discrepancy of the makespan value obtained with the CDS heuristic tends to increase to the value of the respective makespan *LB* until the number of jobs is equal to the number of machines in manufacturing ($N=M$). Additional increments of jobs from $N=M$ lead to diminishing the relative discrepancy. On the other side, with more jobs in the shop higher makespan values are expected even if lower ratios between *LB* and CDS makespan are observed.

Additionally, further simulations analyses with lower number of jobs and machines were performed to the aim to identify how close the CDS heuristic and the *LB* would be situated from optimal makespan solution. Instances with 10, 20, 40 and 50 jobs and 10 machines were simulated and solved with a GAMS model. In terms of ratios, the values obtained for the relative makespan discrepancies were similar to those observed for the relationship between the values obtained with the CDS heuristic and its *LB*. Because the makespan value of optimal solution is always included between the makespan found by the heuristic and its *LB*, the value of the optimal solution tends to be closer the former with increasing number of jobs in the shop. Thus, it can be said that, in a shop with many machines and few jobs, the heuristic shows average results near optimal solution while for shops with many jobs and few machines, the lower bound seems to be a good estimate for the optimal makespan.

References

Askin, R. G., and Standridge, C. R., *Modeling and Analysis of Manufacturing Systems*, John Wiley & Sons, 1993.

Banks, J., *Handbook of simulation: principles, methodology, advances, applications, and practice*, New York: John Wiley & Sons, 1998.

Campbell, H. G. Dudek, R. A., and Smith, M. L., A heuristic algorithm for the *n* job, *m* machine sequencing problem, *Management Science*, vol. 16, no. 10, pp. B630–B637, 1970.

Gharbi, A., and Mahjoubi, A., New lower bounds for flowshop scheduling, *International Journal of Humanities and Management Sciences*, vol 1, no. 1, pp. 28-32, 2013.

Gueret, C., Prins, C., and Sevaux, M., *Applications of Optimization with Xpress-MP*, Translated and revised by Susanne Heipcke, Dash Optimization, 2000.

Gupta, J. N. D., and Stafford, E. F., Flowshop scheduling research after five decades, *European Journal of Operational Research*, vol. 169, no. 3, pp. 699-711, 2006.

Johnson, S. M., Optimal two- and three-stage production schedules with setup times included, *Naval Research Logistics Quarterly*, vol. 1, no. 1, pp. 61–68, 1954.

Mastrolilli M., Svensson O., Improved Bounds for Flow Shop Scheduling. In: Albers S., Marchetti-Spaccamela A., Matias Y., Nikoletseas S., Thomas W. (eds) Automata, Languages and Programming. ICALP 2009, *Lecture Notes in Computer Science*, vol. 5555, pp. pp 677-688, Springer, Berlin, Heidelberg, 2009.

Morton, T. E., and Pentico, D. W., *Heuristic scheduling systems with applications to production systems and project management*, John Wiley & Sons Inc., New York, 1993.

Salmasi, N., Logendran, R., and Skandari, M. R., *Makespan minimization of a flowshop sequence-dependent group scheduling problem*, The International Journal of Advanced Manufacturing Technology, vol. 56, no. 5-8, pp. 699–710, 2011.

Souza, E. C., Programação de tarefas em um flow-shop, *Tese de doutorado*, Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia Naval e Oceânica, São Paulo, 2009.

Taillard, E., Some efficient heuristic methods for the flowshop sequencing problem, *European Journal of Operational Research*, vol. 47, no. 1, pp. 65-74, 1990.

Vallada, E., Ruiz, R., and Minella, G., Minimising total tardiness in the m-machine flowshop problem: A review and evaluation of heuristics and metaheuristics, *Computers & Operations Research*, vol. 35, no. 4, pp. 1350-1373, 2008.

Zhu, Z., and Heady, R. B., Minimizing the sum of earliness and tardiness in multi-machine scheduling: a mixed integer programming approach, *Computer & Industrial Engineering*, vol. 38, no. 2, pp. 297-305, 2000.

Biography

Bruno de Sousa Alves received the B.S. degree in Electrical and Industrial Engineering, and the M.Sc. degree in Electrical Engineering, both from the Federal University of Santa Catarina (UFSC), Brazil, in 2014 and 2016, respectively. He earned the best student award in 2014 from UFSC and concluded his Master's degree with the maximum GPA (4/4). He is currently a Ph.D. Candidate in Electrical Engineering at Polytechnique de Montreal, Canada, and his research mainly deals with the numerical modeling of extremely nonlinear electromagnetic thin layers, including both numerical techniques and physical models. Particular applications include electromagnetic modeling, design, and optimization of high-performance electrical devices.

Carlos Ernani Fries is currently Associate Professor in the Department of Production and Systems Engineering at UFSC. Mr. Fries holds a Master and PhD in Production Engineering from UFSC. He has taught courses of Operations Research applied to Manufacturing, Decision Theory, and Statistics among others. His research interests include Simulation, Optimization, Management Games, and Data Analysis. He is member of INFORMS.