**1. Introduction**

**2. Literature review**

**3. Problem description and formulations**

**4. Special Cases**

**5. Experiments**

**6. Proposed Solution**

**5. 1 General Description of the experiments**

5.2 Compare performance in small to medium sized problems

In this experiment a comparative analysis of Guan’s and TBased2 formulations is being made. The objectives of the experiment are:

1. Analyze which parameters affect the solution time for each of the formulations.
2. Compare the two formulations, performance as problem resolutions grows.

5.2.1 Experimental design

A factorial design was chosen to achieve the objectives, factors for this experiment are summarized in <Table:6>, there are three 2 level factors and one 4 level factor. A randomly generated problem for this experiment is generated in the following manner: firstly number of machines and number of jobs are being set by the factors; then for each job a process time,, is drawn from discrete uniform distribution , then the size of the job,, is drawn from discrete uniform distribution , and finally a due date value, is drawn from discrete uniform distribution . The distribution of process time is , so the values will be drawn from or etc., this allows to look at problems with different process times as similar problems but with different time resolution. For each combination of the factors 10 random problems are generated. There are 32different factors (, and 10 problems for each factor so in total 320 problems are generated.

|  |  |
| --- | --- |
| Factor | Set of possible values |
| M – Number of machines | {15,25} |
| – Number of jobs | {15,25} |
| – Max. due date | {15,25} |
| – Max. Process time | {5,10,20,40} |

<Table:6> Factors summary.

Each of the 320 problems will be tested with Guan’s formulation, and with TBased2, so in total 640 responses are received.

5.2.2 Results and Analysis

Three ANOVA tests conducted on three different responses, time received by Guan’s formulation, time received from TBased2 formulation and the difference between them. All models were found significant, with significance level of 0.05. Guan’s most influencing factor is number of jobs, as the number of jobs growth the time it takes to solve the problem also growth with it, this is expected as it is the only factor that affects number of integer variables, further analysis is provided in section 5.2.2.1. For TBased2 all factors has influence on the problem solution time, this is because all the factors has impact on number of integer variables, further analysis is provided in section 5.2.2.2

Summary of solution results shown in <Table:1>, in total TBased2 solved all problems faster than Guan. Despite this to, in almost of the problems, 201, Guan had better solution time. Comparing only optimally solved problems, by both of the models, Guan solved the problems faster, on average Guan solved each problem in 106.2 sec (SD = 261.1), and for TBased2 it took 127.5 sec. (SD = 308) per problem. In both formulations the standard deviations of problems that are optimally solved for both of them, are bigger than the mean, this shows that there is a big variation in solution time.

|  |  |  |
| --- | --- | --- |
|  | **Guan** | **TBased2** |
| **Average solution time** | 339.3 sec. (SD = 624.7) | 225.6 sec. (SD = 471.1) |
| **Average solution time optimal only** | 112.6 sec (SD = 261.4) | 131.9 sec. (SD = 308.9) |
| **Average solution time mutual optimality** | 106.2 sec. (SD = 261.1) | 127.5 sec. (SD = 308.2) |
| **Median of solution time** | 6.6 sec. | 33.3 sec. |
| **Median of solution time for optimally solved problem** | 2.1 sec. | 28.5 sec. |
| **Better solution time** | 201 | 119 |
| **Optimality proved** | 277 | 300 |
| **Exclusive Optimality** | 14 | 37 |
| **Mutual optimality** | 263 | 263 |
| **Better solution value** | 10 | 2 |

<Table:1> Summary of solution time

TBased2 has a little bit more optimally proved problems than Guan, 300 vs. 277, but despite that Guan formulation resulted 10 times in a better solution than TBased2, while TBased2 had only 2 results that are better than Guan. Additionally <TableId:1> shows that for each formulation there is a small set of problems that only it, was able to proof optimality, 14 for Guan, and 37 for TBased2, it indicates that the factors are affect differently on each of the formulation.

As explained in section 5.2.1 above, the data set is also constructed in a way that it is possible to compare what is happening to each formulation when the time resolution changes, i.e. solving the same problem but once when a time unit is hour, and then when the time unit is half an hour, etc.. From <Figure 7>, it is seen that as the resolution is high, for example hours, TBased2 is better in solving the problems, and when moving to a more granular time units, for example quarter hours, TBased2 performs slower than Guan formulation, since Guan’s solution time is not impacted by this factor.

<Figure:7>

5.2.2.1 Analysis of Guan formulation

Most of the problems are being solved in less than 100 seconds, as can be seen in the histogram <Figure:13>.

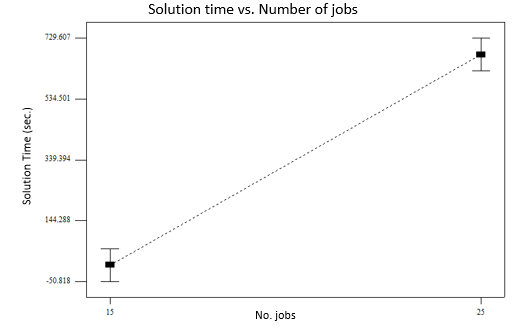
<Figure:13> Histogram of Guan’s solution time.

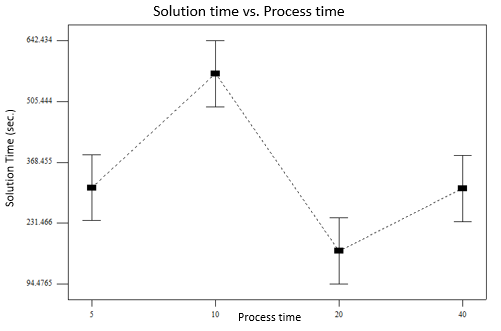
To analyze results, and understand which factors are most influencing solution time of a problem, ANOVA test was used.

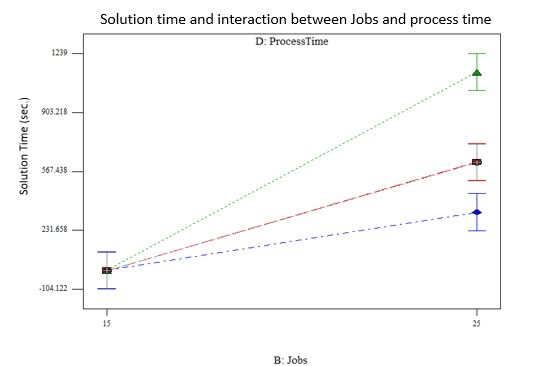
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **Sum of**  **Squares** | **df** | **Mean**  **Square** | **F**  **Value** | **p-value**  **Prob > F** |
| **Model** | 58712306.78 | 28 | 2096868.099 | 9.2112896 | < 0.0001 |
| A-Machines | 374059.3875 | 1 | 374059.3875 | 1.6431979 | 0.2009 |
| **B-Jobs** | **36497227.49** | **1** | **36497227.49** | **160.32793** | **< 0.0001** |
| C-Duedate | 6310.572199 | 1 | 6310.572199 | 0.0277216 | 0.8679 |
| **D-ProcessTime** | **6653822.887** | **3** | **2217940.962** | **9.7431481** | **< 0.0001** |
| AB | 361047.4894 | 1 | 361047.4894 | 1.5860382 | 0.2089 |
| AC | 125440.1285 | 1 | 125440.1285 | 0.5510434 | 0.4585 |
| AD | 1143697.729 | 3 | 381232.5763 | 1.6747089 | 0.1725 |
| BC | 6586.961743 | 1 | 6586.961743 | 0.0289357 | 0.8650 |
| **BD** | **6639642.032** | **3** | **2213214.011** | **9.7223832** | **< 0.0001** |
| CD | 2202615.698 | 3 | 734205.2325 | 3.2252754 | 0.0230 |
| ABC | 126794.6112 | 1 | 126794.6112 | 0.5569935 | 0.4561 |
| ABD | 1144009.051 | 3 | 381336.3502 | 1.6751648 | 0.1724 |
| ACD | 1214321.745 | 3 | 404773.9148 | 1.7781232 | 0.1514 |
| BCD | 2216730.998 | 3 | 738910.3326 | 3.2459443 | 0.0223 |
| Residual | 66243560.4 | 291 | 227641.101 |  |  |
| Lack of Fit | 1191467.451 | 3 | 397155.8171 | 1.7582966 | 0.1553 |
| Pure Error | 65052092.95 | 288 | 225875.3227 |  |  |
| Cor Total | 124955867.2 | 319 |  |  |  |

<Table:2> ANOVA results for Guan formulation

<Table:2> shows received ANOVA results for testing Guan’s solution time dependent variable. There are two significant factor for guan solution time, number of jobs with , indicating that as number of jobs is growing the time it takes to solve the problem is increasing, and a second factor, process time, with . As can be seen from <Figure:12>, process time does not impact the solution time in any specific direction, it is not expected that process time will be significant factor, since it does not have a relation to the number of integer variables in Guan’s formulation. On the other hand number of jobs has a visible impact, <Figure:2>, on solution time, this result is expected since the amount of integer variables in Guan’s formulation is being affected by the number of jobs in the problem, and as the number of jobs is bigger, the number of integer variables is growing and the time for solving a problem is increasing drastically. Additionally <Figure:14> shows the significant interaction between number of jobs and process time factors.

 <Figure:2>. Solution time as a function of number of jobs.

<Figure:12>. Solution time as a function of process time.

<Figure:14> Interaction between Number of jobs and process time.

5.2.2.2 Analysis of TBased2 formulation

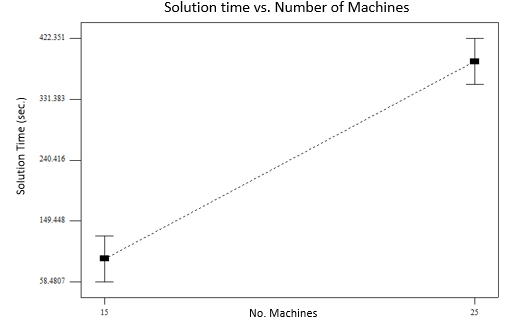
As for Guan, most of the problems were solved in less than 100 seconds, <Figure:15> shows the distribution of results. <Table:3> summarizes the ANOVA results. All factors except due date are affecting the time that will be needed to solve the problem.

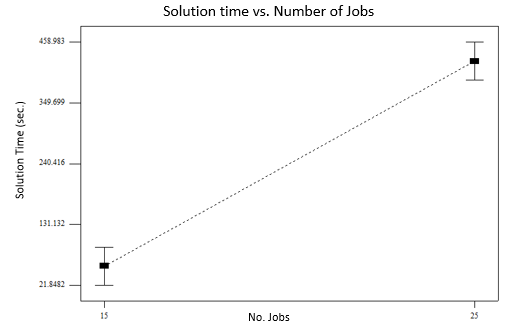
<Figure:15> Histogram TBased2 solution time

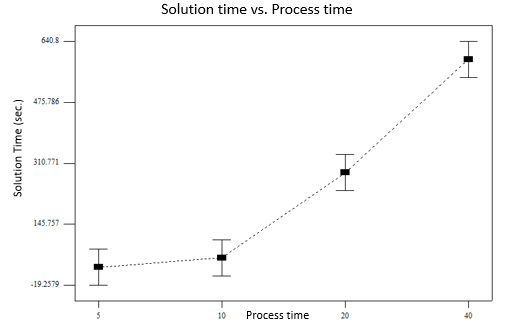
Machines factor is significant with , indicating that increase in the number of machines will cause increase in solution time. Jobs factor is significant with indicating that increase in number of jobs will cause increase in solution time. Process time factor is significant with indicating that increase in average process time for a job will increase problem solution time. <Figure:4>, <Figure:5>,<Figure:6> visualize the change in solution time, as the factor changes. It is trivial to understand why number of jobs and number of machines affect the solution time, as they directly impact the number of integer variables that a formulation will have. Process time is a little bit different, in TBased2 formulation, time horizon is the third factor, besides number of jobs and machines, that impact on the number of integer variables, in conducted tests time horizon was set as , thus increase in average process time would increase time horizon and hence increase the number of integer variables. Additionally any interaction between the three significant factors is also significant, figures 16-18 show this. The general pattern for the interactions is that a higher factor values are a lot more influencing other high factor values. As an example interaction between machines and jobs is studied, in <Figure:16> it is clearly seen, that the influence of jobs with value 25 is higher than for value 15, the gradient is significantly bigger.

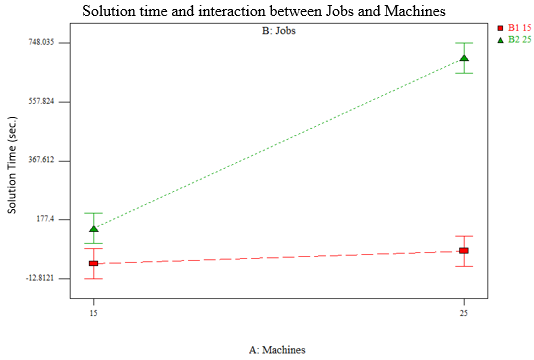
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **Sum of**  **Squares** | **df** | **Mean**  **Square** | **F**  **Value** | **p-value**  **Prob > F** |
| **Model** | **56694865.38** | **28** | **2024816.621** | **20.86244418** | **< 0.0001** |
| **A-Machines** | **6976987.272** | **1** | **6976987.272** | **71.88651359** | **< 0.0001** |
| **B-Jobs** | **10868234.79** | **1** | **10868234.79** | **111.9794945** | **< 0.0001** |
| C-Duedate | 216126.6552 | 1 | 216126.6552 | 2.226833895 | 0.1367 |
| **D-ProcessTime** | **16398807.51** | **3** | **5466269.169** | **56.32101903** | **< 0.0001** |
| **AB** | **5177224.445** | **1** | **5177224.445** | **53.34288295** | **< 0.0001** |
| AC | 161874.3395 | 1 | 161874.3395 | 1.667851961 | 0.1976 |
| **AD** | **4833917.347** | **3** | **1611305.782** | **16.6018871** | **< 0.0001** |
| BC | 4072.922004 | 1 | 4072.922004 | 0.041964841 | 0.8378 |
| **BD** | **7794231.118** | **3** | **2598077.039** | **26.76896103** | **< 0.0001** |
| CD | 543503.4096 | 3 | 181167.8032 | 1.866639746 | 0.1353 |
| ABC | 164739.1568 | 1 | 164739.1568 | 1.697369247 | 0.1937 |
| **ABD** | **3217539.745** | **3** | **1072513.248** | **11.05050578** | **< 0.0001** |
| ACD | 74052.51746 | 3 | 24684.17249 | 0.254330276 | 0.8582 |
| BCD | 263554.1568 | 3 | 87851.3856 | 0.905165737 | 0.4389 |
| Residual | 28243173.78 | 291 | 97055.58002 |  |  |
| Lack of Fit | 82151.27852 | 3 | 27383.75951 | 0.280051008 | 0.8398 |
| Pure Error | 28161022.51 | 288 | 97781.32815 |  |  |
| Cor Total | 84938039.17 | 319 |  |  |  |

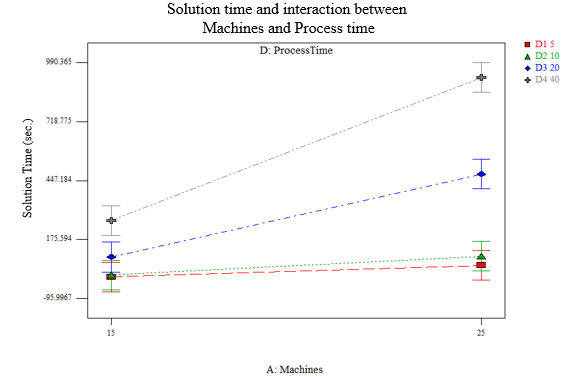
<Table:3> ANOVA results for TBased2 formulation

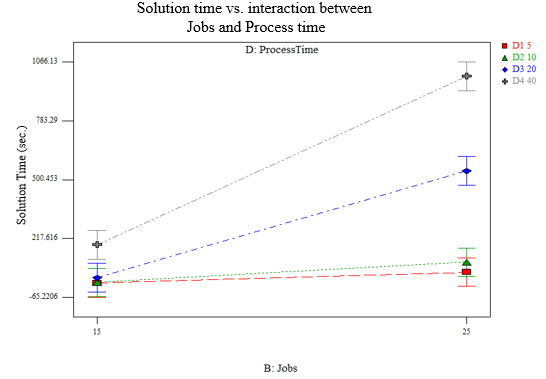
<Figure:4>. Solution time vs number of machines

<Figure:5>. Solution time vs number of jobs

<Figure:6>. Solution time vs max job process time

<Figure:16>. Solution time as function of interaction between jobs and machines

 <Figure:17>. Solution time as function of interaction between machines and process time

 <Figure:18>. Solution time as function of interaction between jobs and process time

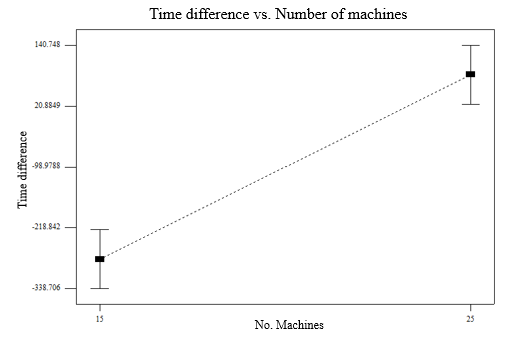
5.2.2.3 Analysis of solution time difference between the two formulations

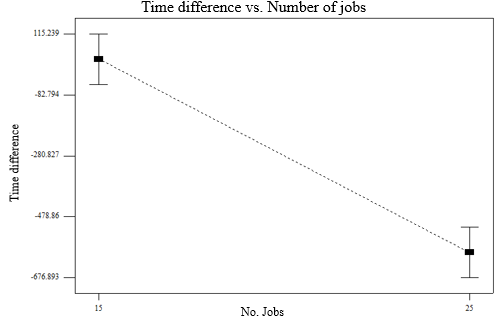
Third dependent variable that is tested is the difference of solution time between TBased2 and Guan. All factors except due date, has a significant impact on the difference between the two formulations. Machines factor is significant

with factor , indicating that as the number of machines is growing the difference changes from negative (meaning TBased2 is faster), to positive (meaning Guan is faster), <Figure:9> visualize this.

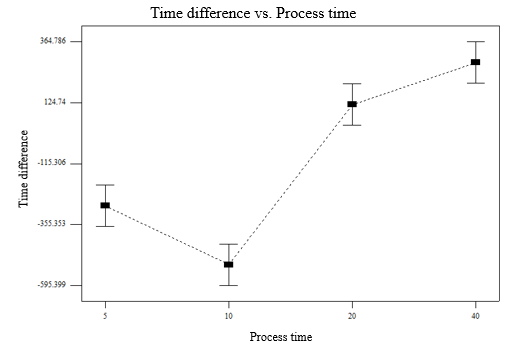
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | Sum of  Squares | df | Mean  Square | F  Value | p-value  Prob > F |
| **Model** | **1.01E+08** | **28** | **3615508** | **13.06458** | **< 0.0001** |
| **A-Machines** | **10582026** | **1** | **10582026** | **38.238** | **< 0.0001** |
| **B-Jobs** | **7532790** | **1** | **7532790** | **27.21963** | **< 0.0001** |
| C-Duedate | 148575.7 | 1 | 148575.7 | 0.536876 | 0.4643 |
| **D-ProcessTime** | **31822591** | **3** | **10607530** | **38.33015** | **< 0.0001** |
| **AB** | **8272663** | **1** | **8272663** | **29.89315** | **< 0.0001** |
| AC | 572309.5 | 1 | 572309.5 | 2.068032 | 0.1515 |
| **AD** | **8419730** | **3** | **2806577** | **10.14152** | **< 0.0001** |
| BC | 300.6982 | 1 | 300.6982 | 0.001087 | 0.9737 |
| **BD** | **21686048** | **3** | **7228683** | **26.12074** | **< 0.0001** |
| CD | 1907233 | 3 | 635744.2 | 2.297252 | 0.0778 |
| ABC | 580587.7 | 1 | 580587.7 | 2.097945 | 0.1486 |
| **ABD** | **6422060** | **3** | **2140687** | **7.735339** | **< 0.0001** |
| ACD | 833771.7 | 3 | 277923.9 | 1.004274 | 0.3912 |
| BCD | 2453538 | 3 | 817846.1 | 2.955275 | 0.0328 |
| Residual | 80531672 | 291 | 276741.1 |  |  |
| Lack of Fit | 818208.6 | 3 | 272736.2 | 0.98538 | 0.4000 |
| Pure Error | 79713464 | 288 | 276782.9 |  |  |
| Cor Total | 181765896.9 | 319 |  |  |  |

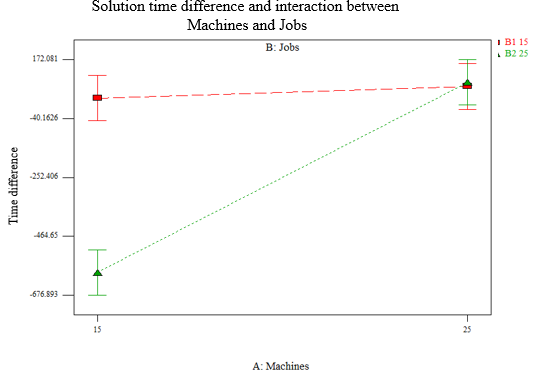
<Table:4>. ANOVA results for time difference test.

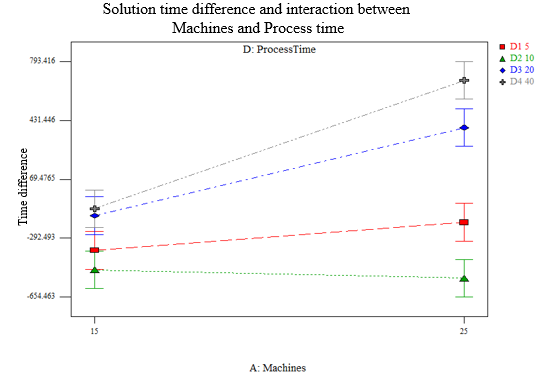
<Figure:9>. Time difference between TBased2 and Guan formulation, as number of machines changes.

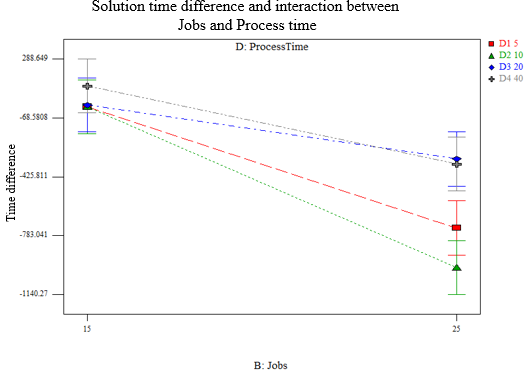
Number of jobs is significant factor with . This factor impacts the opposite way from previous one, as number of jobs is growing difference value is degrading, which means that in cases where we have many jobs, TBased2 is better performing than Guan, <Figure:10> shows this. <Figure:10>. Time difference between TBased2 and Guan formulation, as number of jobs changes.

Last significant factor is the process time, with , indicating that as the process time is getting bigger, Guan will have better solution time, <Figure:11> show the difference change. Summary of other ANOVA analysis is provided in <Table:4>. A correlation between the difference and number of machines or process time is expected, because both of the factors have a significant effect on the solution time, only in TBased2 formulation. Number of jobs has inverse correlation, which means that Guan’s performance degrades more rapidly as the number of jobs is being increased, than TBased2 formulation. So for problems with relative big number of jobs TBased2 has more potential of getting better result.

<Figure:11>. Time difference between TBased2 and Guan formulation, process time changes.

<Figure:19> Solution time difference as a function of interaction between Jobs and Machines.

<Figure:20>. Solution time difference as a function of interaction between Machines and Process time.

<Figure:21>. Solution time difference as a function of interaction between jobs and process time.

5.3 Compare solution quality on large problems

In this experiment a comparative analysis of Guan’s and TBased2 formulations is being made. The objectives of the experiment are:

1. Compare the two formulations, by solution quality.
2. Analyze which parameters affect mostly on solution quality in each formulation

5.3.1 Experimental design

A factorial design was chosen to achieve the objectives, factors for this experiment are summarized in <Table:6>, there is one two level factor, one three level and one four level. A randomly generated problem for this experiment is generated in the following manner: firstly number of machines and number of jobs are being set by the factors; then for each job a process time,, is drawn from discrete uniform distribution , then the size of the job,, is drawn from discrete uniform distribution , and finally a due date value, is drawn from discrete uniform distribution . The distribution of process time is , as described above. For each combination of the factors 10 random problems are generated. There are 24different factors (, and 10 problems for each factor so in total 120 problems are generated.

|  |  |
| --- | --- |
| Factor | Set of possible values |
| M – Number of machines | {15,25} |
| – Number of jobs | {30,40,50} |
| – Max. Process time | {20,40} |

<Table:7> Factors summary.

Each of the 120 problems will be tested with Guan’s formulation, and with TBased2, so in total 240 responses are received.

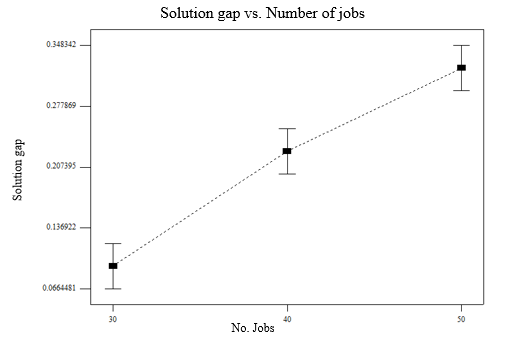
5.3.2 Results and Analysis

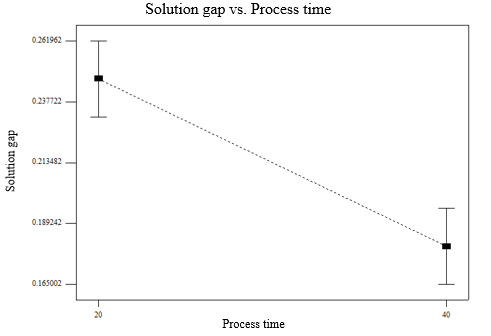
|  |  |  |
| --- | --- | --- |
|  | **Guan** | **TBased2** |
| **Optimally solved** | 17 | 42 |
| **Mutual Optimality** | 11 | 11 |
| **Exclusive Optimality** | 6 | 31 |
| **Better solution value** | 59 | 13 |
| **Better Solution value, for problems solved mutually not optimal.** | 55 | 3 |
| **Avg Gap** | 0.213 (SD = 0.13) | 0.337 (SD = 0.43) |
| **Avg. Gap mutually not optimal** | 0.263 (SD = 0.10) | 0.526 (SD = 0.44) |

<Table:5>. Summary results for experiment 2.

5.3.2.1 Analysis of Guan formulation

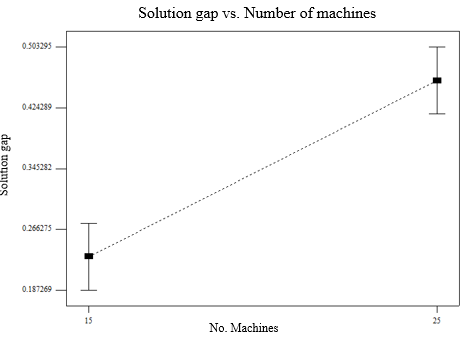
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | Sum of  Squares | df | Mean  Square | F  Value | p-value  Prob > F |
| **Model** | **1.372218192** | **11** | **0.124747108** | **17.88240997** | **< 0.0001** |
| A-Machines | 1.48702E-08 | 1 | 1.48702E-08 | 2.13163E-06 | 0.9988 |
| **B-Jobs** | **1.184154069** | **2** | **0.592077034** | **84.87382517** | **< 0.0001** |
| **C-Process Time** | **0.133601293** | **1** | **0.133601293** | **19.15165103** | **< 0.0001** |
| AB | 0.004367244 | 2 | 0.002183622 | 0.313020666 | 0.7319 |
| AC | 0.005493765 | 1 | 0.005493765 | 0.787527399 | 0.3768 |
| BC | 0.024954208 | 2 | 0.012477104 | 1.788584031 | 0.1721 |
| ABC | 0.019647599 | 2 | 0.009823799 | 1.408234716 | 0.2490 |
| Pure Error | 0.753404475 | 108 | 0.006975967 |  |  |
| Cor Total | 2.125622667 | 119 |  |  |  |

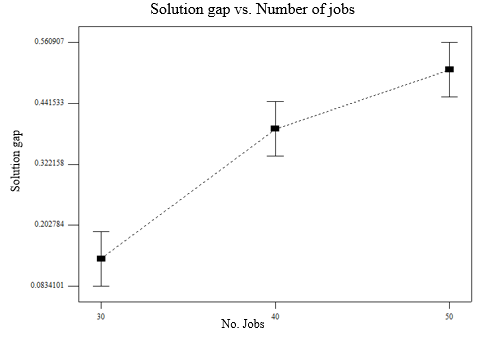


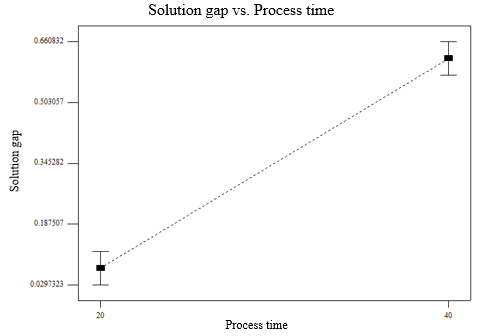


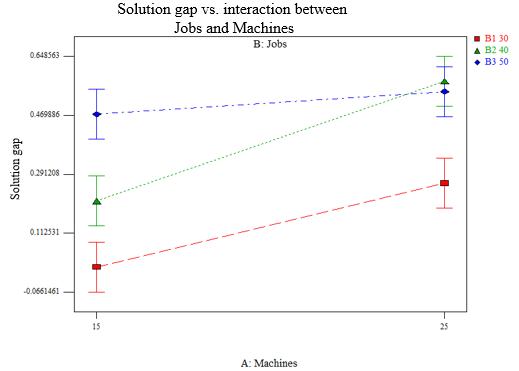
5.3.2.2 Analysis of TBased2 formulation

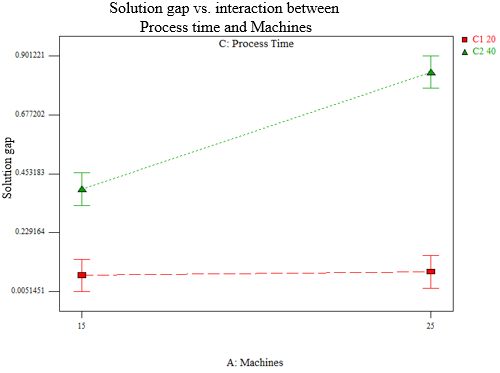
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | Sum of  Squares | df | Mean  Square | F  Value | p-value  Prob > F |
| **Model** | **16.520836** | **11** | **1.5018942** | **25.899319** | **< 0.0001** |
| **A-Machines** | **1.5715659** | **1** | **1.5715659** | **27.10077** | **< 0.0001** |
| **B-Jobs** | **2.8776294** | **2** | **1.4388147** | **24.81155** | **< 0.0001** |
| **C-Process Time** | **8.8765066** | **1** | **8.8765066** | **153.07036** | **< 0.0001** |
| **AB** | **0.4438406** | **2** | **0.2219203** | **3.8268906** | **0.0248** |
| **AC** | **1.3971195** | **1** | **1.3971195** | **24.092539** | **< 0.0001** |
| **BC** | **1.2977202** | **2** | **0.6488601** | **11.189227** | **< 0.0001** |
| ABC | 0.0564534 | 2 | 0.0282267 | 0.4867537 | 0.6160 |
| Pure Error | 6.2628893 | 108 | 0.0579897 |  |  |
| Cor Total | 22.783725 | 119 |  |  |  |

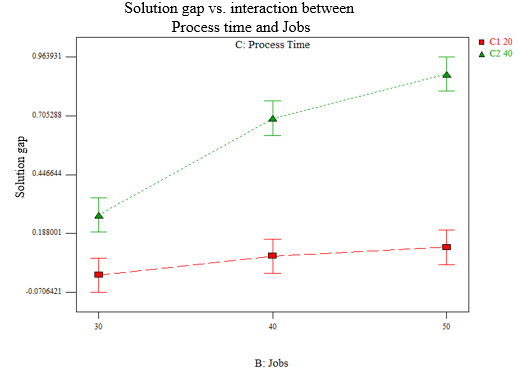












5.3.2.3 Analysis of Gap difference between formulations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | Sum of  Squares | df | Mean  Square | F  Value | p-value  Prob > F |
| **Model** | **16.87780395** | **11** | **1.534345813** | **22.7199555** | **< 0.0001** |
| **A-Machines** | **1.571871689** | **1** | **1.571871689** | **23.27562309** | **< 0.0001** |
| B-Jobs | 0.384793913 | 2 | 0.192396957 | 2.84893422 | 0.0623 |
| **C-Process Time** | **11.18810033** | **1** | **11.18810033** | **165.6687426** | **< 0.0001** |
| **AB** | **0.452331861** | **2** | **0.226165931** | **3.348971162** | **0.0388** |
| **AC** | **1.577832529** | **1** | **1.577832529** | **23.36388872** | **< 0.0001** |
| **BC** | **1.682225498** | **2** | **0.841112749** | **12.45484823** | **< 0.0001** |
| ABC | 0.020648126 | 2 | 0.010324063 | 0.152874439 | 0.8584 |
| Pure Error | 7.293559525 | 108 | 0.067532959 |  |  |
| Cor Total | 24.17136347 | 119 |  |  |  |