Graduate Program
Department of Applied Informatics
University of Macedonia

# Investigation of Business Process Optimization Using Design of Experiments

M.Sc. Thesis of Dimitrios Paganias
Supervisor: Kostas Vergidis

Thessaloniki,
29 October 2019

# Research aim and objectives

- Research aim:

  The aim of this thesis is the systematic investigation of the business process optimization problem parameters as they were introduced at the business process optimization framework ($bpo^F$) (Vergidis,2008).

- Research objectives:

  1. Studying and understanding Business Process Optimization
  2. Reviewing the results of BPOF reported by Vergidis (2008)
  3. Studying and understanding Design of Experiments (DoE)
  4. Determination of the problem parameter limits that generate reliable results
  5. Characterization of the problem parameters significance & their influence on the results

# Main Terms

- <u>Business Process</u>:

  *Business process* is a collective set of tasks that when properly connected and sequenced perform a business operation.

  The aim of a business process is to perform a business operation ( i.e. any service-related operation that produces value to the organization )

- <u>Business Process Modeling</u>:

  Business process modeling is the activity of representing processes of an enterprise, so that the current process may be analyzed, improved and automated.

# Main Terms

Techniques of
business process modeling

Diagrammatic models

(diagrams, defined notation e.g. shapes ,lines, arrows)

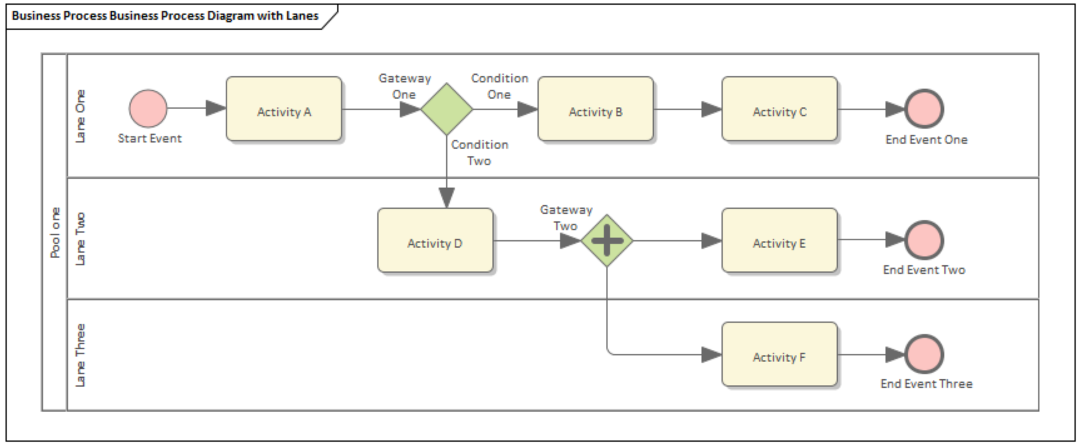
Flowchart, IDEF, UML, BPMN

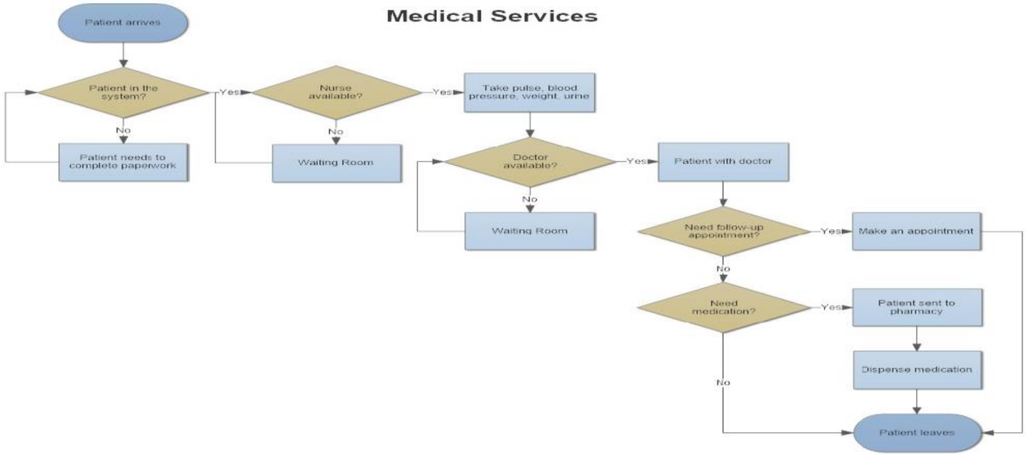
Mathematical models

(mathematical parameters, objective functions)

Approaches by: Hofacker and Vetschera, Vergidis

# Main Terms




- <u>Business process optimization</u>:

  Business process optimization is the problem of constructing feasible business process designs with optimum attributes

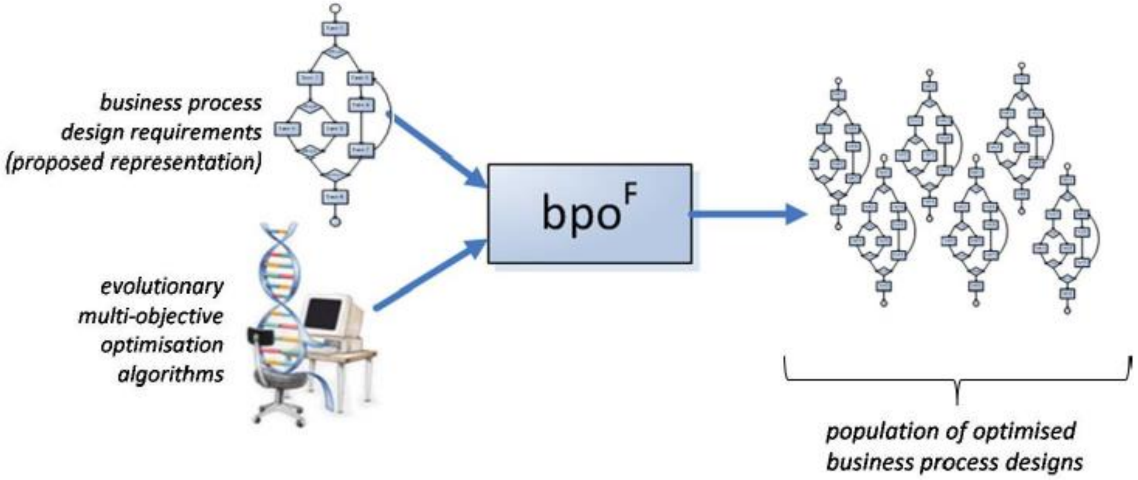
  Hofacker and Vetschera approach

  Vergidis approach

# Business Process Optimization Framework ($bpo^F$)

- The business process optimization framework ($bpo^F$) is an approach for the Evolutionary Multi-objective business process optimization

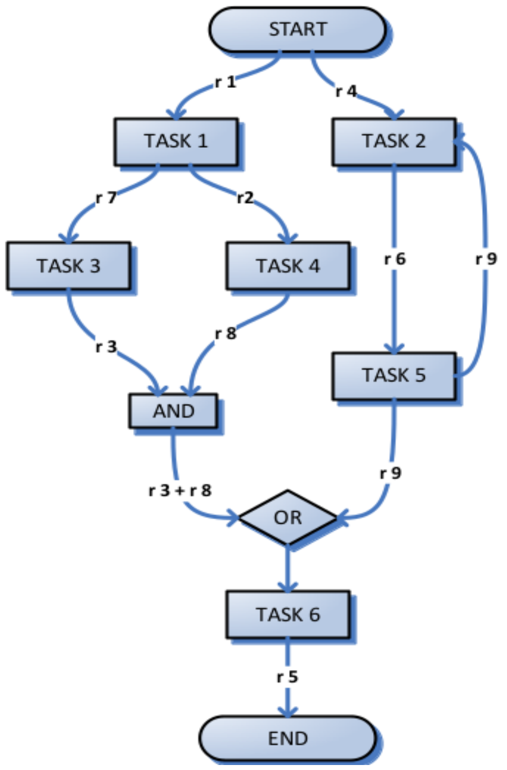
- The ($bpo^F$) utilizes the proposed business process representation and EMOAs in order to generate alternative optimized business process designs



# Business Process Optimization Framework ($bpo^F$)

## Business Process Representation

*Visual representation*




*Mathematical parameters*

| Parameter | Description | Parameter | Description |
|---|---|---|---|
| $n$ | Number of tasks in the library | $N$ | Set of the n tasks |
| $n_d$ | No. of tasks in the design | $N_d$ | Set of the $n_d$ tasks (subset of N) |
| $n_{min}$ | Minimum number of tasks in the design | $N_{in}$ | Set of library tasks to be included in the process design (subset of N) |
| $r$ | No. of available resources | $N_{ex}$ | Set of library tasks to be excluded for the process design (subset of N) |
| $t_{in}$ | No. of task input resources | $S_d$ | Set of the different process sizes |
| $t_{out}$ | No. of task output resources | DoI | Degree of Infeasibility (as calculated by the PCA algorithm) |
| $r_{in}$ | No. of process input resources | TAM | Matrix that stores the task attribute values for each of the $n_d$ tasks in the process design |
| $r_{out}$ | No. of process output resources | PA | Set of the p process attribute values |
| $p$ | No. of task/process attributes | | |

# Business Process Optimization Framework (bpo$^F$)

## Business Process Representation

### Task Attributes Matrix (TAM)

| Tasks \ Attributes | $A_1$ | $A_2$ |
|---|---|---|
| Task 1 | 100 | 300 |
| Task 2 | 120 | 302 |
| Task 3 | 117 | 324 |
| Task 4 | 178 | 308 |
| Task 5 | 145 | 356 |
| Task 6 | 157 | 389 |

| | | |
|---|---|---|
| PROCESS | 817 | 1979 |

### Task Resources Matrix (TRM)

| Tasks \ Resources | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_5$ | $r_6$ | $r_7$ | $r_8$ | $r_9$ |
|---|---|---|---|---|---|---|---|---|---|
| Task 1 | 1 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| Task 2 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 1 |
| Task 3 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| Task 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| Task 5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 |
| Task 6 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 1 | 1 |

# Business Process Optimization Framework ($bpo^F$)

Business Process Representation
& Design Composition

- Process Composition Algorithm (PCA):
  The PCA constructs the business process diagram and checks whether the result corresponds to a feasible business process or not.

  A design is considered infeasible when:
  1. One or more process input resources cannot be utilized from the tasks in the TRM
  2. One or more process output resources cannot be produced from the tasks in the TRM
  3. There is no task in TRM than can be attached to the process diagram based on its input and output resources

# Business Process Optimization Framework (bpo$^F$)

## Process Composition Algorithm (PCA)

- PCA Inputs

*1. Process input and output requirements ($R_{in}$ and $R_{out}$)*

*2. Participating tasks in the design (TRM)*

*3. Task library (N)*

- PCA Outputs

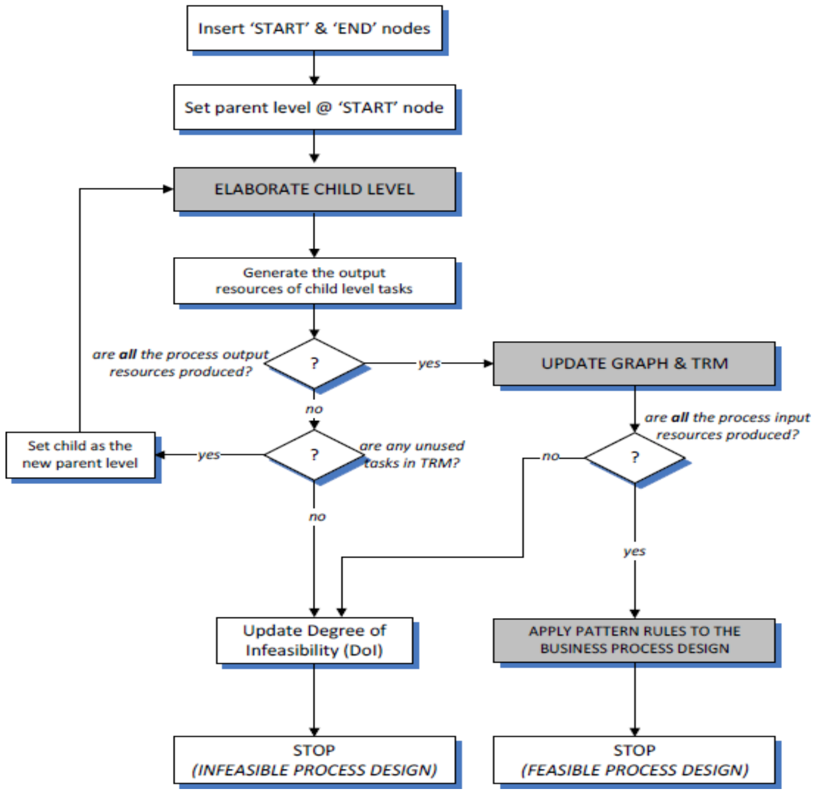
*1. Business process design (process graph)*

*2. Updated set of tasks in the design ($N_d$)*

*3. Degree of Infeasibility (DoI)*

$$DoI = 1 \cdot n_{in} + 5 \cdot (\sim r_{out}) + 3 \cdot (\sim r_{in})$$

- Main steps of the PCA



# Business Process Optimization Framework ($bpo^F$)

## $bpo^F$ inputs

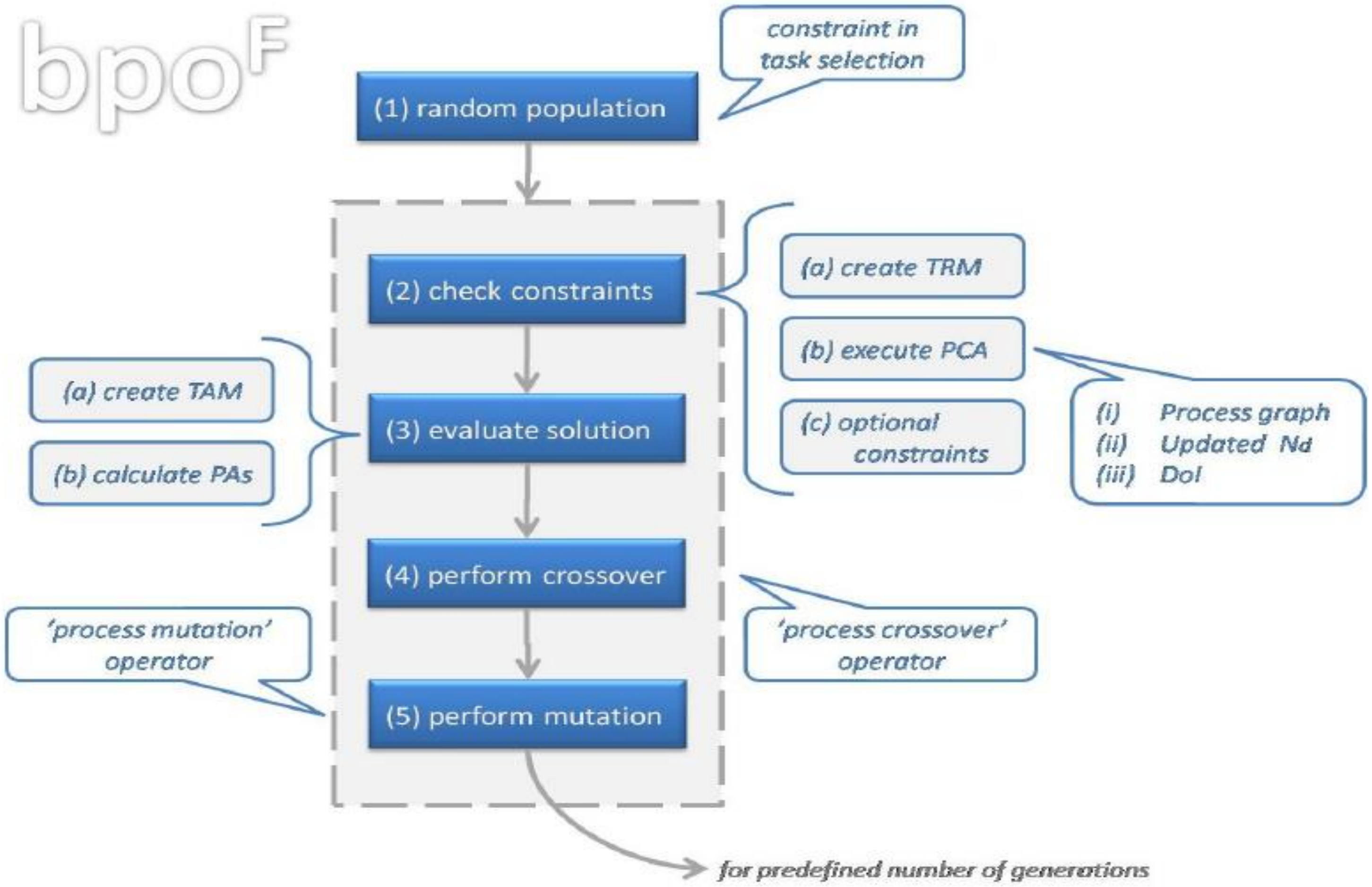
1. The process requirements for the design {process inputs ($R_{in}$) / process outputs ($R_{out}$)}.
2. The process size ($n_d$).
3. The library of tasks (N).
4. The process attribute functions

## $bpo^F$ outputs

1. The tasks in the design, stored in the $N_d$ set.
2. The process graph
3. The Degree of Infeasibility (DoI)
4. The process attribute values

# Business Process Optimization Framework (bpo$^F$)

Main steps of bpo$^F$



Investigation of business process optimization

Investigation of business process optimization problem parameters

Scalable business process tests

Design of Experiments (DoE)

# Scalable business process test problems

## Problem parameters

| Parameter | Description |
|---|---|
| $n$ | Number of tasks in the library |
| $r$ | No. of available resources |
| $n_d$ | No. of tasks in the design |
| $n_{min}$ | Minimum number of tasks in the design |
| $t_{in}$ | No. of task input resources |
| $t_{out}$ | No. of task output resources |
| $r_{in}$ | No. of process input resources |
| $r_{out}$ | No. of process output resources |
| $p$ | No. of task/process attributes |

## Scenario A

| | |
|---|---|
| $r_{in}$ | 1 |
| $r_{out}$ | 1 |
| $t_{in}$ | 1 |
| $t_{out}$ | 1 |

## Scenario B

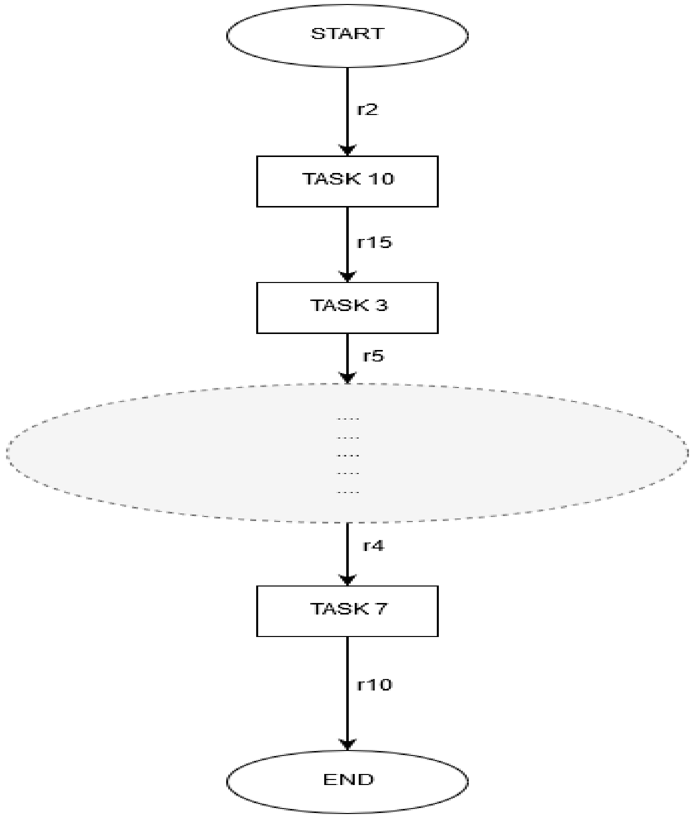
| | |
|---|---|
| $r_{in}$ | 5 |
| $r_{out}$ | 5 |
| $t_{in}$ | 3 |
| $t_{out}$ | 3 |

$n_{min} = n_d$ , $p = 2$ : A=[100,115] B=[200,230]

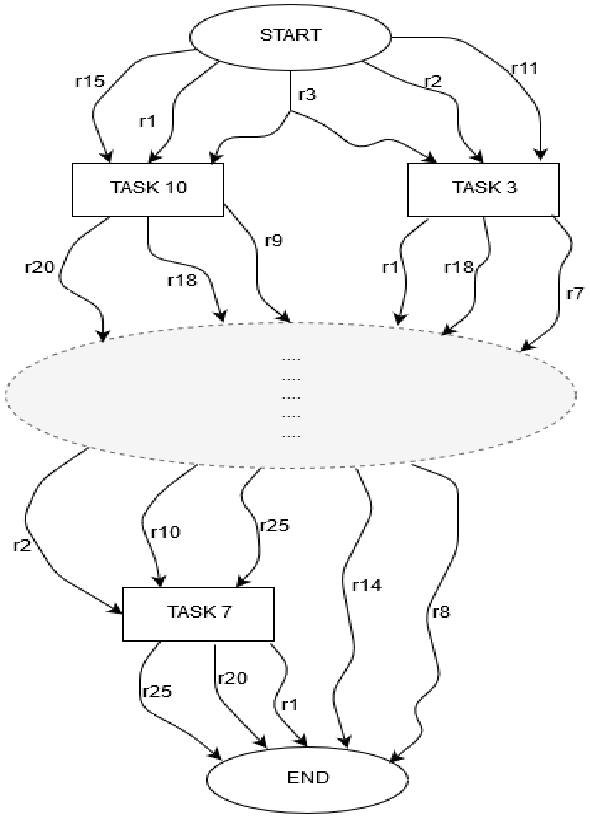
| Parameter | NSGA2 |
|---|---|
| Population | 500 |
| Generations | 25,000 |
| Crossover prob. | 0.8 |
| Mutation prob. | 0.2 |

# Scalable business process test problems

## Process design of Scenario A




## Process design of Scenario B



# First part of Scenario A

| | $n_d$=10 | | r | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 10 | 20 | 30 | 40 | 50 | 60 |
| Test1 | n | 50 | 0 | 1 | 1 | - | - | - |
| Test2 | | 75 | 0 | 1 | 1 | - | - | - |
| Test3 | | 100 | 0 | 2 | 1 | 0 | - | - |
| Test4 | | 150 | 0 | 1 | 2 | 0 | - | - |
| Test5 | | 200 | - | 4 | 5 | 2 | 1 | 0 |
| Test6 | | 300 | - | 4 | 3 | 2 | 2 | 0 |

# Second Part of Scenario A

| | n | r | $n_d$ | |
|---|---|---|---|---|
| | | | 15 | 20 |
| Test7 | 200 | 20 | 0 | 0 |
| Test8 | 200 | 30 | 0 | 0 |
| Test9 | 300 | 20 | 0 | 0 |
| Test10 | 300 | 30 | 0 | 0 |

# First part of Scenario B

| | $n_d$=10 | | r | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 50 | 60 |
| Test1 | n | 30 | 0 | 10 | - | 3 | 0 | - | - | - | - |
| Test2 | | 100 | 0 | 9 | - | 10 | 8 | 3 | 3 | 0 | - |
| Test3 | | 125 | - | 0 | 8 | - | 5 | 3 | 3 | 0 | - |
| Test4 | | 150 | - | 0 | 0 | 3 | 4 | 7 | 2 | 1 | 0 |
| Test5 | | 175 | - | - | 6 | - | 5 | 4 | 3 | 0 | - |
| Test6 | | 200 | - | - | 0 | 0 | 7 | 0 | - | - | - |

# Second part of Scenario B

| | n | r | $n_d$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 9 | 8 | 7 | 6 | 5 | 4 | 11 | 12 | 13 |
| Test7 | 30 | 15 | 12 | 7 | 8 | 10 | 6 | 2 | | | |
| Test8 | 30 | 25 | 1 | 2 | 2 | 0 | - | - | | | |
| Test9 | 30 | 15 | | | | | | | 6 | 5 | 0 |
| Test10 | 30 | 25 | | | | | | | 0 | - | - |

| | n | r | $n_d$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 9 | 8 | 7 | 6 | 5 | 4 | 11 | 12 | 15 | 20 | 25 | 30 | 35 | 40 |
| Test11 | 200 | 30 | 5 | 5 | 4 | 4 | 5 | 2 | | | | | | | | |
| Test12 | 200 | 30 | | | | | | | 4 | 6 | 6 | 3 | 5 | 2 | 1 | 1 |

# Reliable results of the tests

<table>
<tr><th rowspan="2"></th><th colspan="3">Parameter</th></tr>
<tr><th>$n$</th><th>$r$</th><th>$n_d$</th></tr>
<tr><td rowspan="2">Scenario A<br>$\begin{Bmatrix} r_{in} = r_{out} = 1 \\ t_{in} = t_{out} = 1 \end{Bmatrix}$</td><td>200</td><td>20-30</td><td rowspan="2">10</td></tr>
<tr><td>300</td><td>20-30</td></tr>
<tr><td rowspan="6">Scenario B<br>$\begin{Bmatrix} r_{in} = r_{out} = 5 \\ t_{in} = t_{out} = 3 \end{Bmatrix}$</td><td>30</td><td>15-25</td><td rowspan="6">10</td></tr>
<tr><td>100</td><td>15-40</td></tr>
<tr><td>125</td><td>20-40</td></tr>
<tr><td>150</td><td>25-35</td></tr>
<tr><td>175</td><td>20-40</td></tr>
<tr><td>200</td><td>30</td></tr>
</table>

<table>
<tr><th rowspan="2"></th><th colspan="3">Parameter</th></tr>
<tr><th>$n$</th><th>$r$</th><th>$n_d$</th></tr>
<tr><td rowspan="4">Scenario A<br>$\begin{Bmatrix} r_{in} = r_{out} = 1 \\ t_{in} = t_{out} = 1 \end{Bmatrix}$</td><td rowspan="2">200</td><td>20</td><td>10</td></tr>
<tr><td>30</td><td>10</td></tr>
<tr><td rowspan="2">300</td><td>20</td><td>10</td></tr>
<tr><td>30</td><td>10</td></tr>
<tr><td rowspan="3">Scenario B<br>$\begin{Bmatrix} r_{in} = r_{out} = 5 \\ t_{in} = t_{out} = 3 \end{Bmatrix}$</td><td>30</td><td>15</td><td>5-12</td></tr>
<tr><td>30</td><td>25</td><td>10</td></tr>
<tr><td>200</td><td>30</td><td>5-25</td></tr>
</table>

# Design of Experiments (DoE)

- DoE is a statistical method which can be used for the investigation and the determination of the relationship between the factors affecting a process and the output of that process.

- DoE provides:
  - ❑Strategies to design an experiment and collect data (single factor design, factorial design)
  - ❑Tools to analyze and interpret the results (Hypothesis testing, P-Value, Analysis of Variance)

# Application of DoE in BPO problem

The application of DoE in business process optimization was based on the following guideline:

1. Recognition of the problem for which DoE is used
2. Selection of the response variable
3. Choice of factors and levels
4. Choice of experimental design
5. Performing the experiment
6. Statistical analysis of the data
7. Conclusions and recommendations

# Application of DoE in BPO problem

1. Recognition of the problem for which DoE is used

Parameter Characterization

2. Selection of the response variable

No of solutions that are generated by $bpo^F$

# Application of DoE in BPO problem

3. Choice of factors and levels

| Factor | Description |
|---|---|
| n | Number of tasks in the library |
| r | No. of available resources |
| $n_d$ | No. of tasks in the design |
| $n_{min}$ | Minimum number of tasks in the design |
| $t_{in}$ | No. of task input resources |
| $t_{out}$ | No. of task output resources |
| $r_{in}$ | No. of process input resources |
| $r_{out}$ | No. of process output resources |
| p | No. of task/process attributes |

| Category | Factor |
|---|---|
| Design factors | $n$ , $r$ , $n_d$ |
| Held-constant factors | $r_{in}$, $r_{out}$, $t_{in}$, $t_{out}$ , $n_{min}$, $p$ |

| Factor | Level | |
|---|---|---|
| | (-) | (+) |
| n | 30 | 100 |
| r | 15 | 25 |
| $n_d$ | 8 | 10 |

# Application of DoE in BPO problem

4. Choice of experimental design

$2^3$ Factorial Design

| | Factor | | |
|---|---|---|---|
| **Run** | *A* | *B* | *C* |
| 1 | - | - | - |
| 2 | + | - | - |
| 3 | - | + | - |
| 4 | + | + | - |
| 5 | - | - | + |
| 6 | + | - | + |
| 7 | - | + | + |
| 8 | + | + | + |

# Application of DoE in BPO problem

5. Performing the experiment

| | Factor | | | Solutions | | | |
|---|---|---|---|---|---|---|---|
| **Run** | *n* | *r* | $n_d$ | *Replicate1* | *Replicate2* | *Replicate3* | *Replicate4* |
| 1 | 30 | 15 | 8 | 8 | 6 | 8 | 7 |
| 2 | 100 | 15 | 8 | 8 | 11 | 7 | 11 |
| 3 | 30 | 25 | 8 | 2 | 2 | 1 | 1 |
| 4 | 100 | 25 | 8 | 9 | 7 | 10 | 7 |
| 5 | 30 | 15 | 10 | 8 | 9 | 11 | 10 |
| 6 | 100 | 15 | 10 | 7 | 10 | 9 | 8 |
| 7 | 30 | 25 | 10 | 2 | 2 | 1 | 1 |
| 8 | 100 | 25 | 10 | 10 | 10 | 5 | 7 |

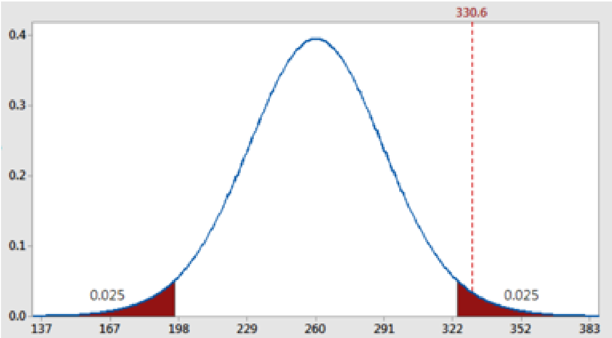
# Application of DoE in BPO problem

6. Statistical analysis of the data

Ho: $\beta_1 = \beta_2 = \beta_3 = \beta_{12} = \beta_{13} = \beta_{23} = \beta_{123} = 0$

H1: at least one $\beta \neq 0$

P- Value



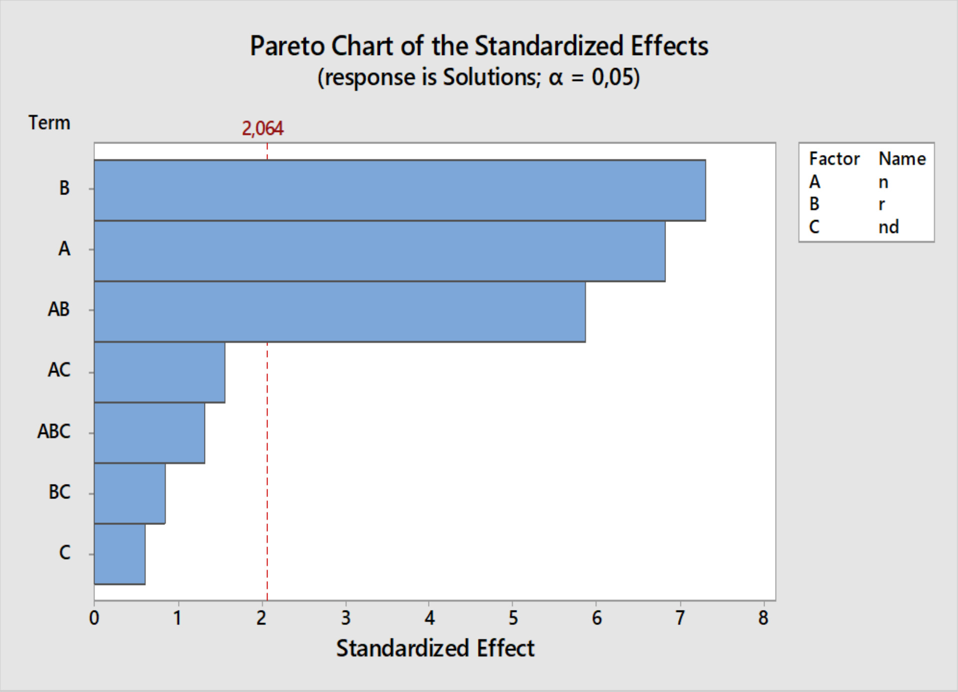

| Source | DF | SS | MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Model | 7 | 304,219 | 43,460 | 19,96 | 0,000 |
| Linear | 3 | 218,594 | 72,865 | 33,47 | 0,000 |
| n | 1 | 101,531 | 101,531 | 46,64 | 0,000 |
| r | 1 | 116,281 | 116,281 | 53,41 | 0,000 |
| $n_d$ | 1 | 0,781 | 0,781 | 0,36 | 0,555 |
| 2-Way Interactions | 3 | 81,844 | 27,281 | 12,53 | 0,000 |
| n*r | 1 | 75,031 | 75,031 | 34,46 | 0,000 |
| n*$n_d$ | 1 | 5,281 | 5,281 | 2,43 | 0,132 |
| r*$n_d$ | 1 | 1,531 | 1,531 | 0,70 | 0,410 |
| 3-Way Interactions | 1 | 3,781 | 3,781 | 1,74 | 0,200 |
| n*r*$n_d$ | 1 | 3,781 | 3,781 | 1,74 | 0,200 |
| Error | 24 | 52,250 | 2,177 | | |
| Total | 31 | 356,469 | | | |

# Application of DoE in BPO problem

## Factor Effects

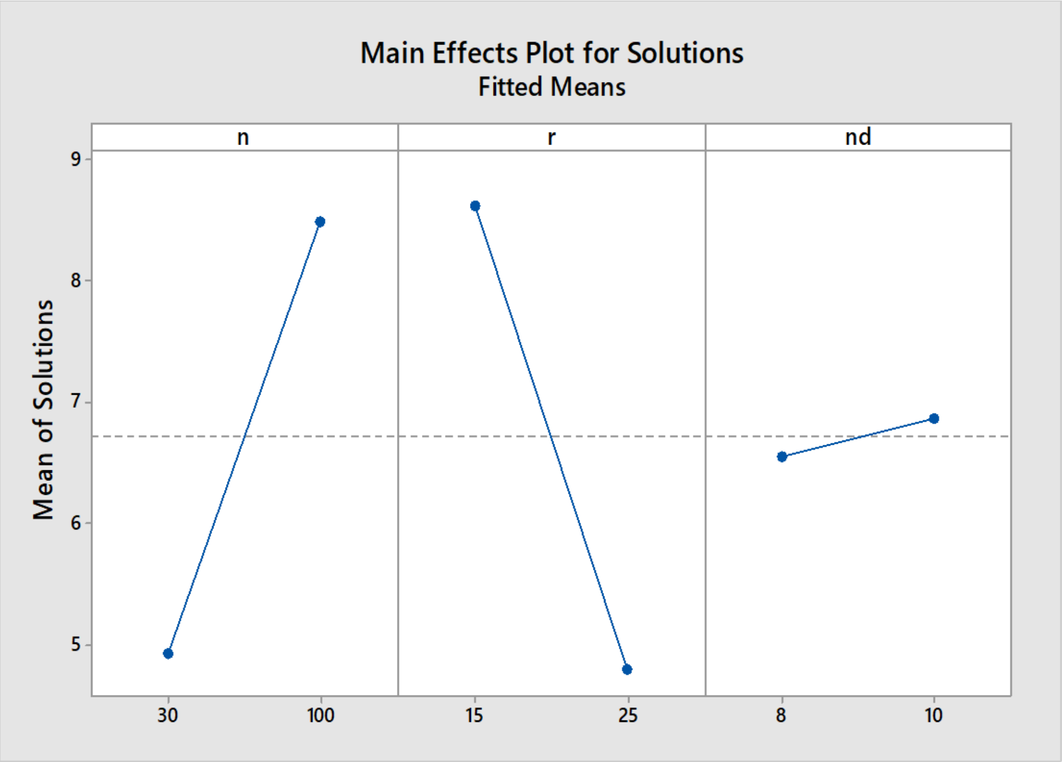
| Term | Effect | P-Value |
|---|---|---|
| $n$ | 3,562 | 0,000 |
| $r$ | -3,813 | 0,000 |
| $n_d$ | 0,312 | 0,555 |
| $n*r$ | 3,062 | 0,000 |
| $n*n_d$ | -0,813 | 0,132 |
| $r*n_d$ | -0,438 | 0,410 |
| $n*r*n_d$ | 0,688 | 0,200 |

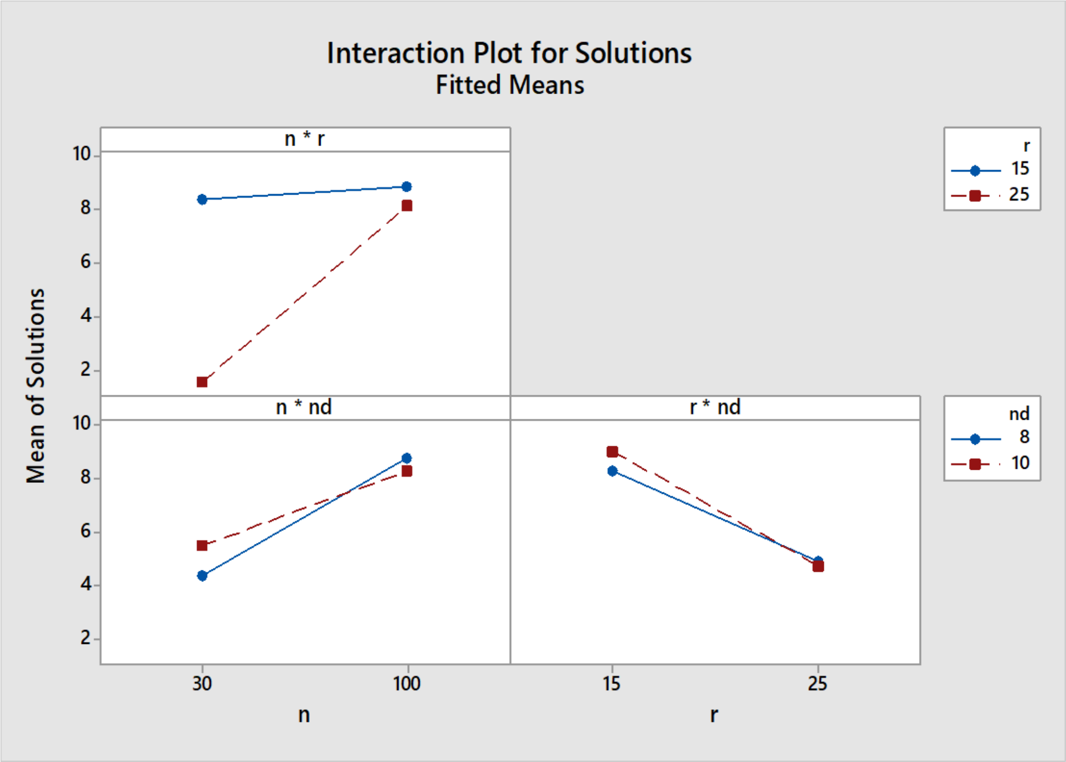
## Pareto chart of the effects



# Application of DoE in BPO problem

*Mean No of solutions and effects*



# Application of DoE in BPO problem

The R- squared quantity:

$$R^2 = \frac{SS_{Model}}{SS_T} = 0{,}8534$$

The regression model:

$$\text{Solutions} = -16{,}6 + 0{,}334\, n + 0{,}593\, r + 4{,}34\, n_d - 0{,}0089\, n{*}r - 0{,}0509\, n{*}n_d - 0{,}171\, r{*}n_d + \\ + 0{,}00196\, n{*}r{*}n_d$$

# Research Contribution

To sum up, through this research the following are determined:

- The parameter values for which the $bpo^F$ generates reliable results and as a result the parameters limits.
- The parameters that have a significant influence on the results that the $bpo^F$ generates.
- The parameter that has the biggest influence.
- The magnitudes of these influences.
- The proportion of variability on the results that the $bpo^F$ generates, which is described by the $n$ , $r$ , $n_d$ parameter group.

# Research limitations & Future work

- Research limitations:
  - Parameters group $r_{in}/r_{out}$ & $t_{in}/t_{out}$
  - Parameter spaces during DoE procedure

- Future work:
  - Investigation of all the possible combinations of the n, r, $n_d$ parameters in the spaces of the n, r parameter examined in this research.
  - Application of the *Response Surface Design* in the BPO problem.

# END OF PRESENTATION

Thank you!