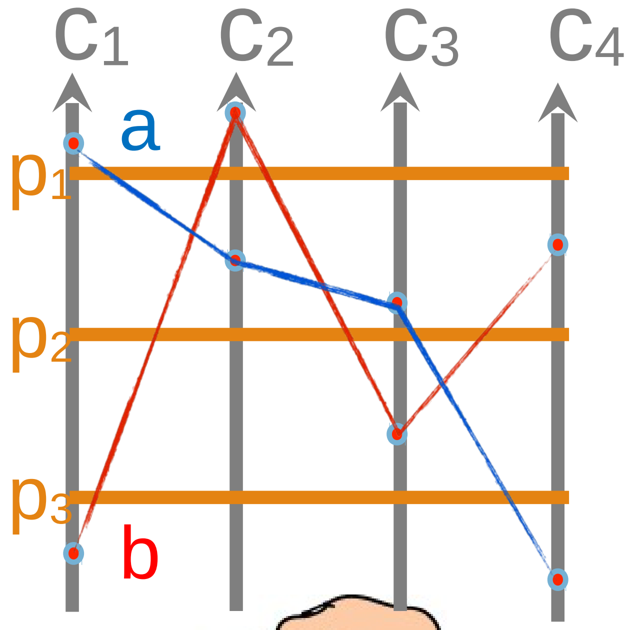
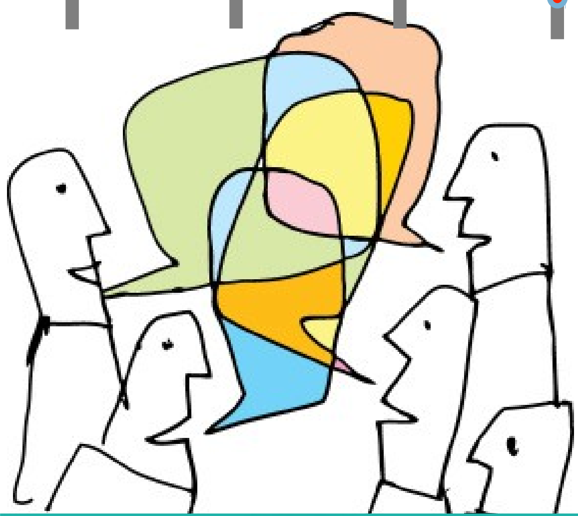




81st meeting of the European working group on
Multicriteria Decision Aiding

Annecy, March 26-28, 2015

Learning the Parameters of a Ranking Model Using Multiple Reference Points: A Case Study Dealing with Large Datasets

Ferretti Valentina, Liu Jinyan, Mousseau Vincent

Outline

- Objective of the research
- Ranking Model Using Multiple Reference Points
  Methodological background
- Case study
  Siting an urban waste landfill in the Province of Turin (Italy)
- Model development
  Presentation of the alternatives
  Elicitation of preference information
  Application of the metaheuristic
- Discussion of the results
- Conclusions

Objective of the research

✓ **Large data-sets**

Nowadays, decision making problems in many contexts involve large datasets, which require adapting and improving the algorithms and techniques to construct acceptable recommendations

✓ **Preference learning**

Preference elicitation thus becomes a challenging task. A very promising line of research in this context consists in inferring indirectly values and preferences from pair-wise comparisons or decision examples supplied by the DM

Which method can be used?

The objectives of the research are twofold:

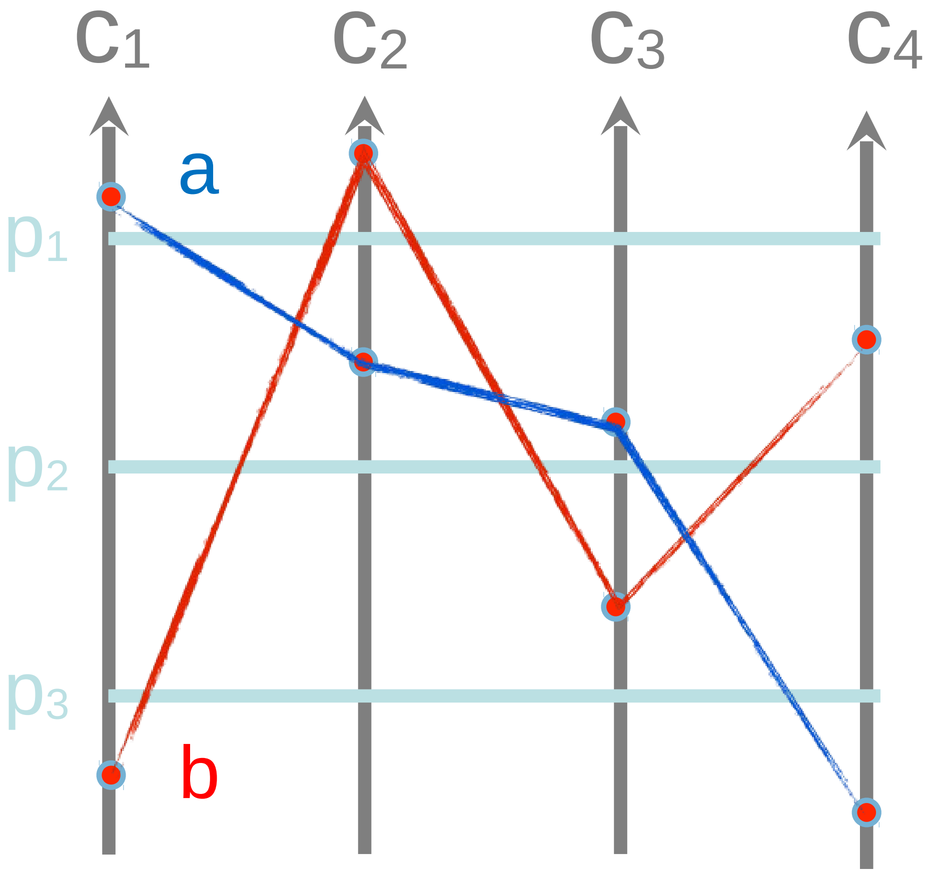
(i) **exploring** the contribution of **Ranking Models Using Multiple Reference Points** for dealing with large data-sets; and
(ii) **validating the innovative algorithmic features** of the tool on an empirical case study

Ranking Model Using Multiple Reference Points

1st step: Compare alternatives to reference point on each criterion

2nd step: Aggregate the comparisons and deduce the relative preference

3rd step: Aggregate the relative preference into the global preference



Case study

This case study concerns the choice of the **most suitable location** for a **Municipal Solid Waste landfill**, which has to be constructed in the Province of Torino (Italy).

The analysis is based on a scientific study that was developed by the Provincial Administration in 2007, where **39 potentially suitable sites** have been identified.

Case study

Landfill location procedure

PHASE 0

Regional planning (competence of the Region)

Definition of the criteria for the definition of unsuitable areas for the location of waste treatment and disposal plants

PHASE 1

Macro level localization (competence of the Province)

- Definition of excluding factors
- Criteria to be minimized
- Criteria to be maximized

Unsuitable areas | Potentially suitable areas

Participation

PHASE 2

Micro level localization (competence of local authorities in charge for waste management)

Screening of the potentially suitable areas (using criteria to be minimized and criteria to be maximized)

Identification and ranking of the suitable sites

PHASE 3

Project (competence of the the specific bodies in charge for the construction of the plants)

Environmental Impact Assessment

PHASE 4

Authorization (within the jurisdiction of the Province)

Permission to build and operate

Case study

criteria definition

Permanent population

The number of people living within a range of 1500 m from each site

Transitory population

The number of people that use the schools, the hospitals and the companies located within a range of 1500 m from each site

Farms

The number of organic farms in surrounding area

Operating costs

Costs for the management and the operation of the plant in each of the considered sites

Ground water vulnerability

The criterion assesses the vulnerability of the groundwater aquifer, considering also the depth of the water table that lies under each site

Land Use capacity

The criterion indicates the potential productive capacity of the soil.

Interference with traffic

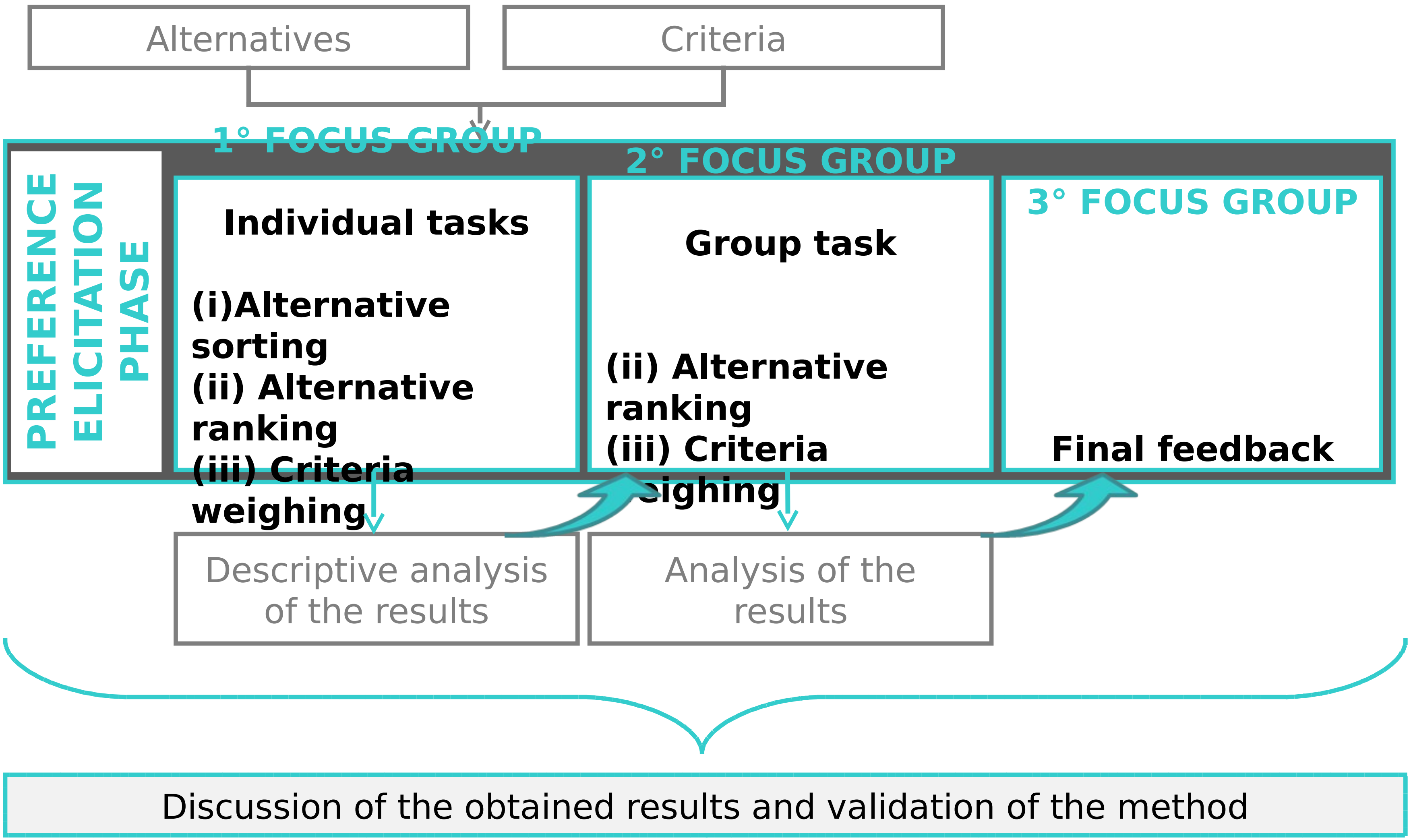
This criterion measures the level of use of road infrastructures in the area surrounding the plant

Case study – Performance matrix

39 POTENTIALLY SUITABLE SITES

SITE	Permanent population [number]	Transitory population [number]	GOD vulnerability index [class]	Land use capacity [class]	Farms [number]	Interference with traffic [m]	Operating costs [€]
Air_A	1461	1484	3	2	0	11050	3.768.283
Air_B	3170	1757	3	2	1	10450	3.561.756
Bri_A	1356	974	4	2	2	6750	2.186.531
Bur_A	867	341	3	1	2	8000	2.864.316
Bur_B	623	225	3	1	1	6500	2.050.974
Caf_A	1356	693	3	3	4	15150	5.179.629
Cav_A	384	69	4	3	0	16650	5.695.972
Crc_A	345	15	3	1	0	9200	3.131.470
Cum_A	1859	684	2	4	2	7850	2.782.982
Cum_B	313	148	3	2	0	11450	3.905.984
Frs_A	140	507	3	4	0	8400	2.856.088
Frs_B	192	563	3	3	0	8000	2.810.093
Mac_A	1062	438	4	3	2	8200	2.918.539
Non_A	337	182	3	2	0	20550	7.038.463
Osa_A	981	569	4	2	5	7450	2.566.091
Pin_A	643	90	4	2	1	4150	480.037
Pin_B	1472	777	4	2	2	6600	2.105.197
Pis_A	1398	1242	3	2	2	8750	2.976.568
Ssp_A	3969	1397	4	2	2	5694	1.613.938
Vig_A	248	20	4	2	1	15000	5.127.995
Vil_A	433	25	4	2	0	19200	6.573.754
Vol_A	1139	445	3	2	2	18650	6.384.429
Air_2	2759	2072	3	2	2	10450	3.681.756
Air_3	1974	1561	3	2	0	10950	3.389.642
Air_4	1699	1527	3	2	0	10950	3.389.642
Non_1	242	369	3	3	0	21570	7.389.576
Fros_1	792	1128	3	2	1	5250	1.373.188
Fros_2	918	1530	3	2	0	5250	1.373.188
Pin_1	494	279	3	1	2	4700	1.074.963
Pin_2	525	125	3	1	2	4350	885.183
Pin_3	485	119	3	1	2	5050	1.264.742
Pin_4	1043	455	2	2	3	4950	1.454.522
Pin_5	445	96	2	2	3	4950	1.454.522
Rol_1	1021	1486	3	2	0	5400	1.454.522
Sca_1	491	53	3	2	3	9850	3.355.219
Sca_2	454	42	3	2	3	9850	3.355.219
Sca_3	535	89	3	2	3	9850	3.355.219
Sca_4	310	15	3	1	0	9200	3.131.470
Vol_2	550	464	3	2	0	17350	5.936.932

The process



Alternatives
Criteria
PREFERENCE ELICITATION PHASE
1° FOCUS GROUP
Individual tasks
(i)Alternative sorting
(ii) Alternative ranking
(iii) Criteria weighing
2° FOCUS GROUP
Group task
(ii) Alternative ranking
(iii) Criteria
eighing
3° FOCUS GROUP
Final feedback
Descriptive analysis of the results
Analysis of the results
Discussion of the obtained results and validation of the method

The focus groups - SORTING

Alternative card

Expert:
Please annotate here the time at which you started analyzing this alternative:

SITE AIR_A

Permanent population: **1461 inhabitants**

Transitory population: **1484 people**

Groundwater vulnerability: **class 3**

Land use capacity: **class 2**

Farms: **0**

Interference with traffic: **11050 meters**

Operating costs: **3.768.283 Euros**

	Class 1	Class 2	Class 3
To which category will you assign this site?			

	Very unconfident	Somehow confident	Absolutely confident
How confident are you with this assessment?			

If you assigned this site to class 1, would you also consider class 2 as a possible category for this site?	If you assigned this site to class 2, would you also consider class 1 or 3 as a possible category for this site? If yes, which one?	If you assigned this site to class 3, would you also consider class 2 as a possible category for this site?

Check procedure

?

Class 1
Highly suitable sites

Class 2
Suitable sites

Class 3
Unsuitable sites

The focus groups - WEIGHING

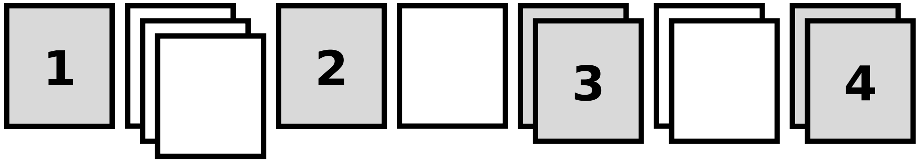
We used the SFR method (Figueira et Roy, 2002) for eliciting the weights of the criteria

1.Give a set of cards to the user: the (i) criteria cards
(ii) blank cards

2.Ask the user to rank the criteria cards from the least important to the most important;

3.Ask the user to think about the importance of two successive criteria and **to introduce blank cards** between them

4.Ask the user to tell **how many times the most important** criterion is more important than the least important one

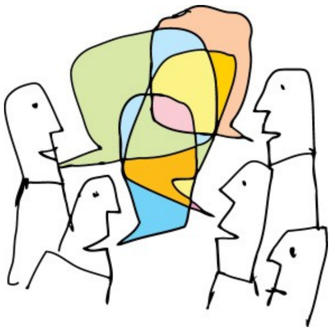


The focus groups' results

Individual preferences (site ranking)

Common preferences

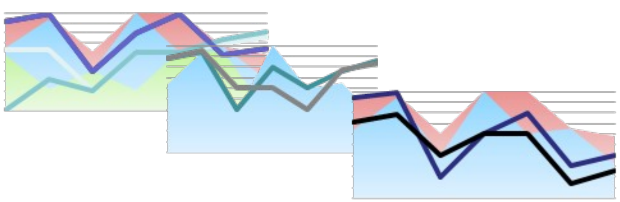
Pin_2
Sca_4
Crc_A
Fros_2
Bur_B
Pin_3
Pin_1
Bur_A
Cum_B
Rol_1
Non_A
...



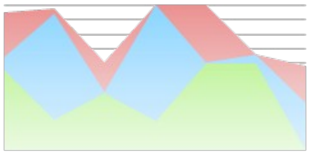
COMPUTING PHASE

Exact algorithm | **Metaheuristic algorithm**

Inferred SRMP models (individual)



Inferred SRMP models (common)



Quick view of the algorithms

MIP algorithm

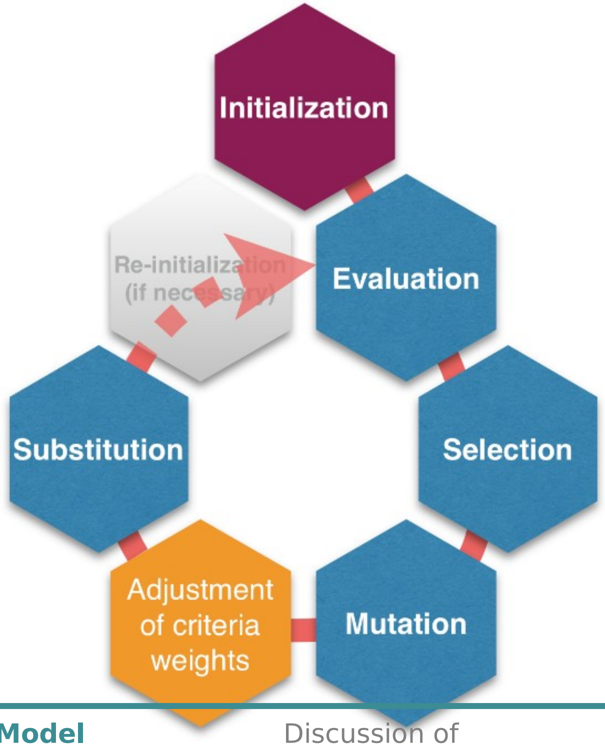
- Based on linear optimization method
- Mixed Integer Programming

  - ✓ Variables
  - ✓ Linear constraints
  - ✓ Objective function

  - Maximize the number of restored pairwise comparisons

Metaheuristic algorithm

- Based on Evolutionary Algorithm



Question No.1

- Q: How does the inferred S-RMP model look like?
  - Reference points
    - The number of reference points?
    - The values of reference points?
    - Their lexicographic order?
  - Criteria weights

e.g. for common ranking

The Inferred S-RMP Model

- Reference points

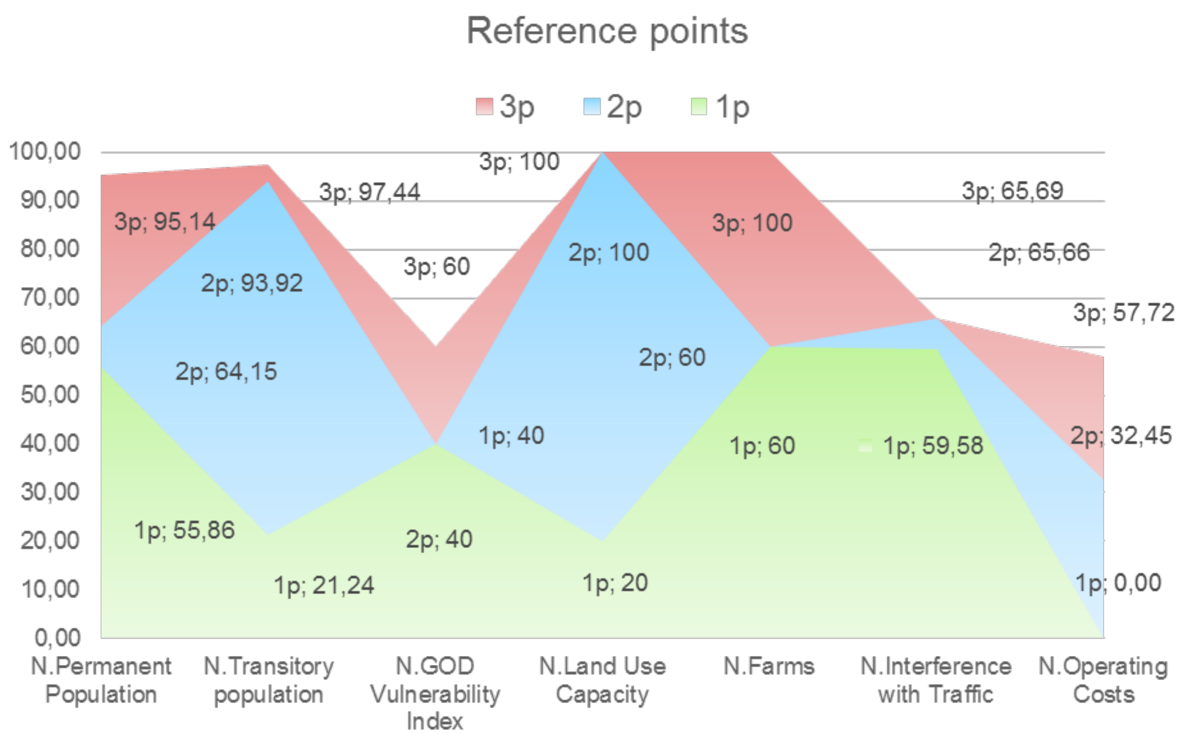
	C1	c2	c3	c4	c5	c6	c7
1p	1752	1632	3	4	2	8719	8561059
2p	1423	126	3	0	2	7407	5782690
3p	193	53	2	0	0	7400	3619347

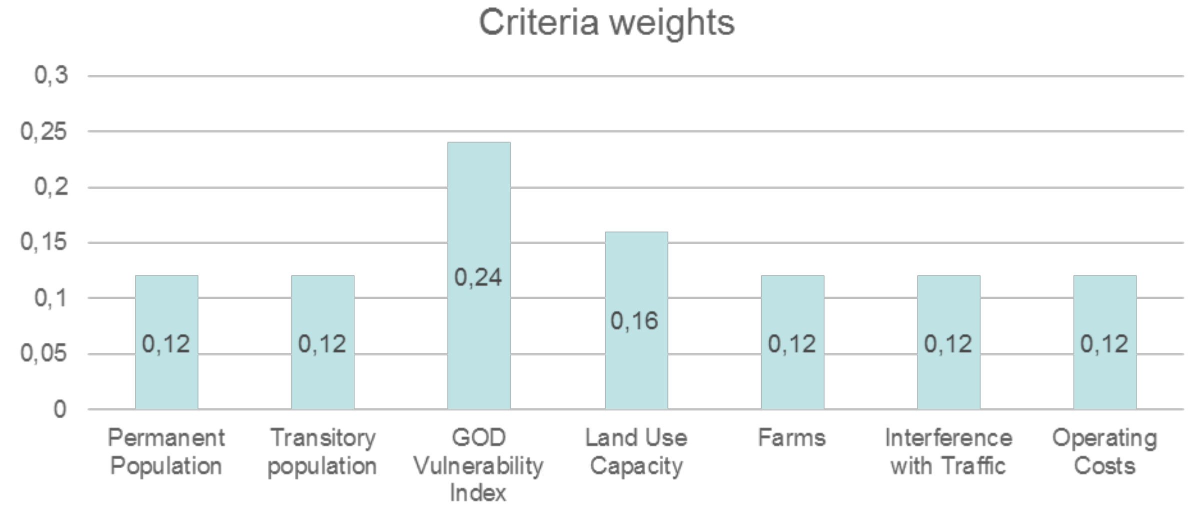
- The inferred lexicographic order is p2 > p3 > p1

	C1	c2	c3	c4	c5	c6	c7
w	0,12	0,12	0,24	0,16	0,12	0,12	0,12

The Inferred S-RMP Model

- The inferred lexicographic order is **p2 -> p3 -> p1**
- The values are linearly transformed to 100-point scale on each criteria.
- The preference direction after transformation is positive on each criteria.
- The weights are normalized to 1.





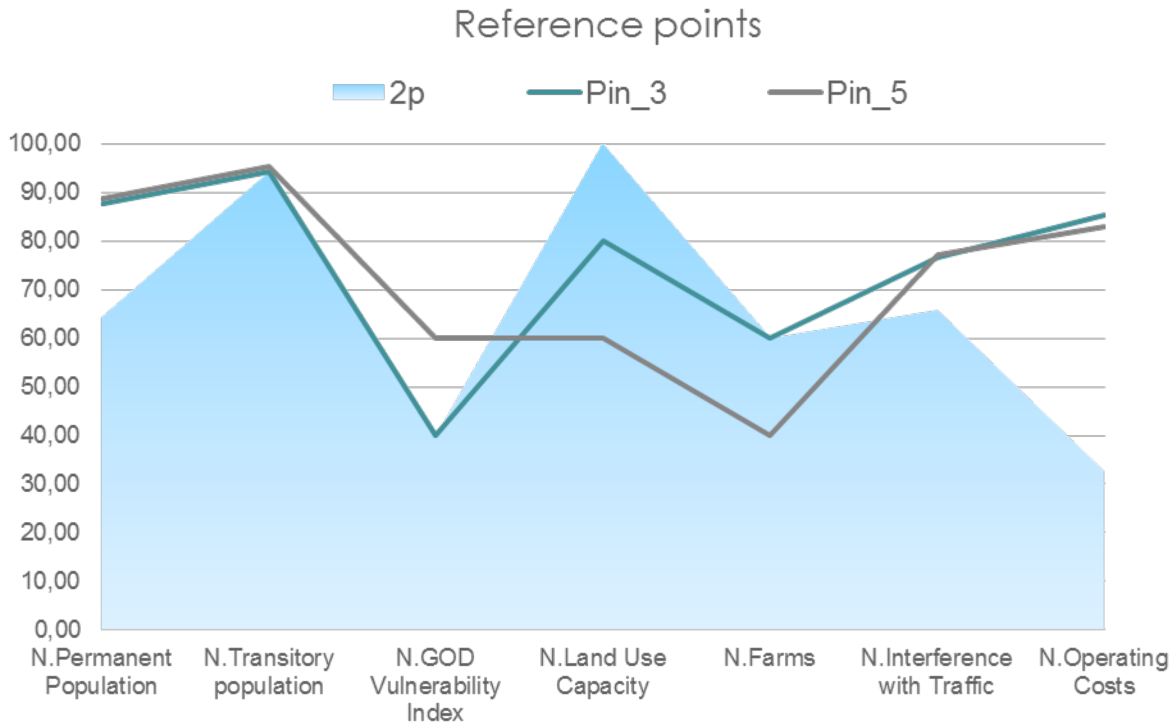
Question No.2

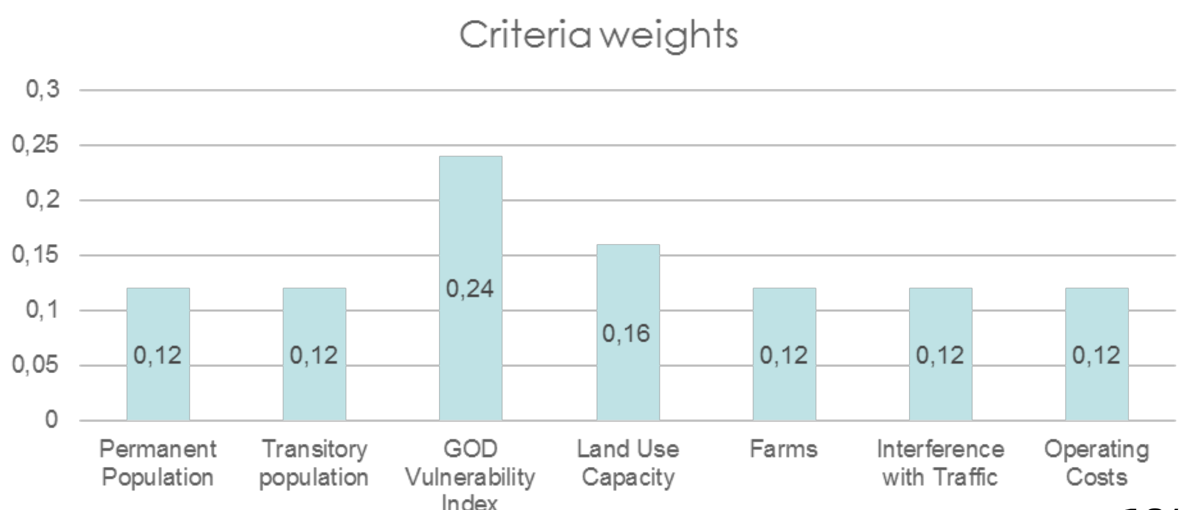
- Q: How does S-RMP method work with the reference points in our case?

By the 1st Reference Point

We can differentiate the pairwise comparisons below:

Pin_3 > Pin_5
Pin_4 > Cum_A
Cum_B > Air_A
Ssp_A > Osa_A
Pis_A > Air_4
Fros_1 > Frs_B
Sca_3 > Vol_2
Bur_B > Frs_A
Vol_A > Vil_A
Pin_2 > Fros_2
Bri_A > Pin_B
Cav_A > Mac_A

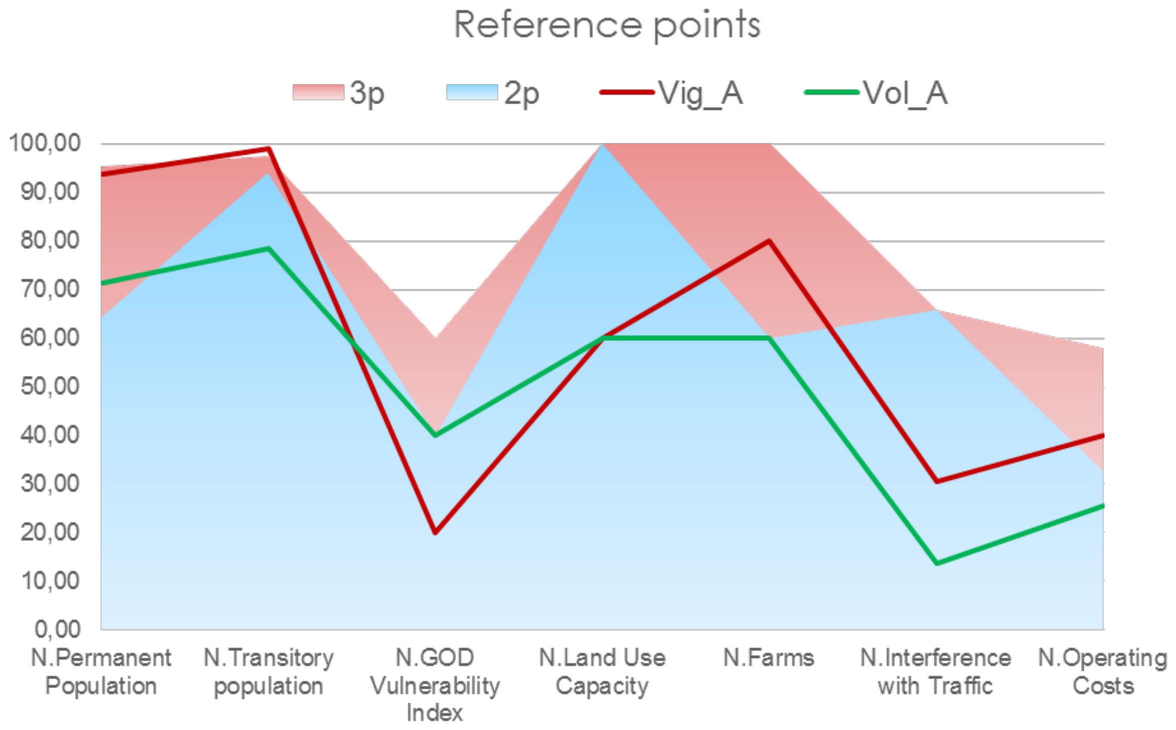


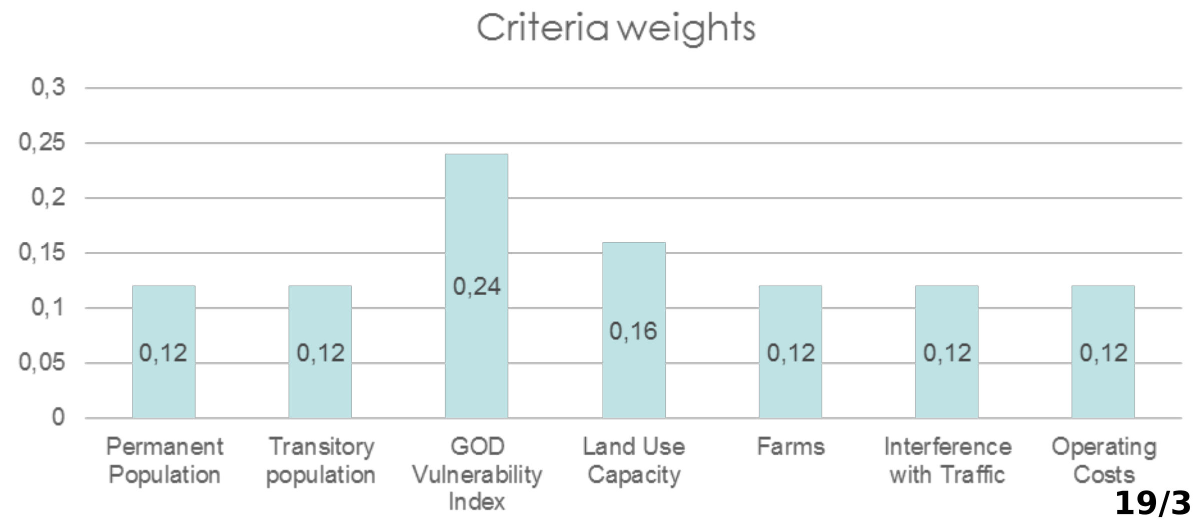


By the 2nd Reference Point

We can differentiate the pairwise comparisons below:

Vig_A > Vol_A
Frs_B > Sca_2
Fros_2 > Pin_1
Sca_2 > Pis_A
Pin_5 > Rol_1
Crc_A > Bur_B
Sca_1 > Sca_3
Air_3 > Non_1
Air_4 > Non_A
Frs_A > Pin_A
Pin_A > Bur_A
Air_B > Caf_A





By the 3rd Reference Point

We can differentiate the pairwise comparisons below:

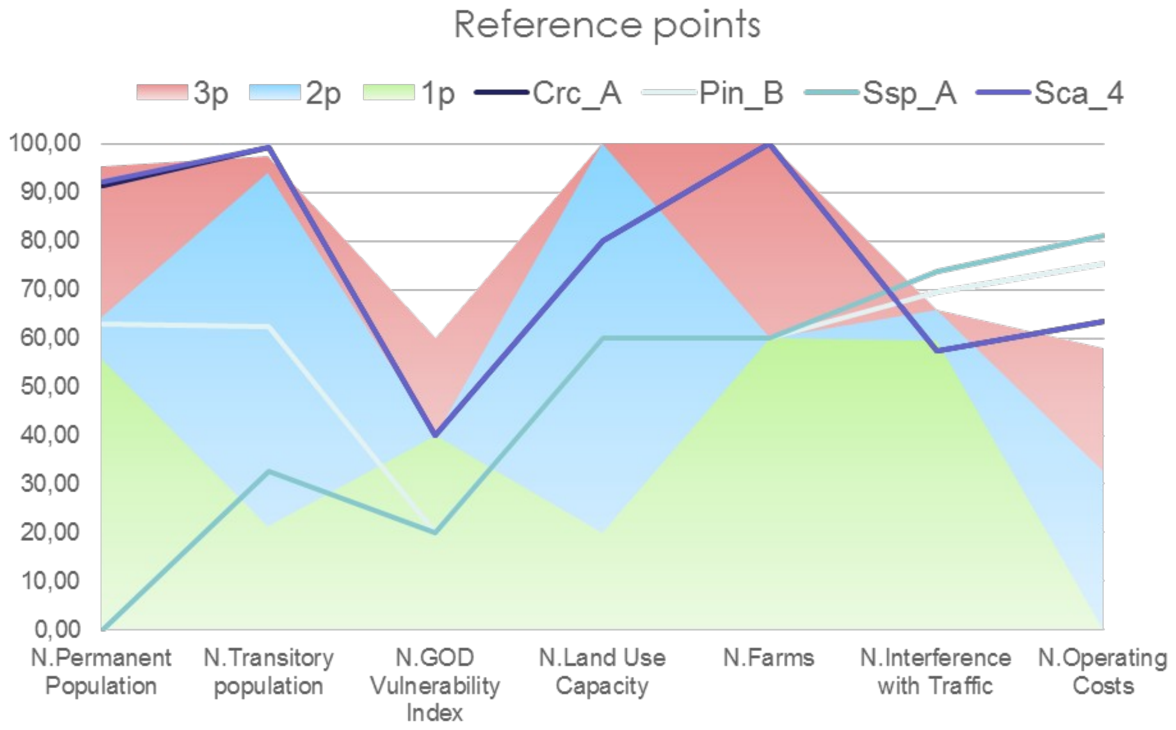
Pin_B > Ssp_A

Rol_1 > Sca_4

Non_1 > Cav_A

Bur_A > Cum_B

Sca_4 and Crc_A are indifferent.



Question No.3

- Q: How many reference points do we need to restored as many pairwise comparisons as possible?

  - How we measure
    "as many pairwise comparisons as possible"?

✓ The ranking accuracy (including indifferences) – R.A. (%)

✓ The strict ranking accuracy (only preferences) – Strict R.A. (%)

$$100\% \geq \text{R.A.} \geq \text{Strict R.A.} \geq 0\%$$

✓ The number of equivalence classes

$$\text{Num. Pairwise Comparisons} \geq \text{Num. Eq. Classes}$$

With 1 Reference Point

	Num. Ref. Pts	1
MIP algorithm	R.A.	**78.95%**
	Strict R.A.	**71.05%**
	Num. Eq. Classes	**19**
Metaheuristic algorithm	Max. R.A. 500	**78.95%**
	Max. Strict R.A. 500	**60.53%**
	Num. Eq. Classes	**20**

- Max. XXX 500 : The maximum value of the 500 repeated trials

Remark: Impossible to reach 100% with 1 reference point!

With 2 Reference Point

	Num. Ref. Pts	1	2
MIP algorithm	R.A.	**78.95%**	**84.21%**
	Strict R.A.	**71.05%**	**78.95%**
	Num. Eq. Classes	**19**	**29**
Metaheuristic algorithm	Max. R.A. 500	**78.95%**	**81.58%**
	Max. Strict R.A. 500	**60.53%**	**68.42%**
	Num. Eq. Classes	**20**	**21**

Remark: Impossible to reach 100% with 2 reference point, neither!!

With 3 Reference Point

	Num. Ref. Pts	1	2	3
MIP algorithm	R.A.	**78.95%**	**84.21%**	**86.84%**
	Strict R.A.	**71.05%**	**78.95%**	**86.84%**
	Num. Eq. Classes	**19**	**29**	**31**
Metaheuristic algorithm	Max. R.A. 500	**78.95%**	**81.58%**	**81.58%**
	Max. Strict R.A. 500	**60.53%**	**68.42%**	**73.68%**
	Num. Eq. Classes	**20**	**21**	**23**

Remark: Never reach 100% with 3 reference point!!!

With More Reference Points

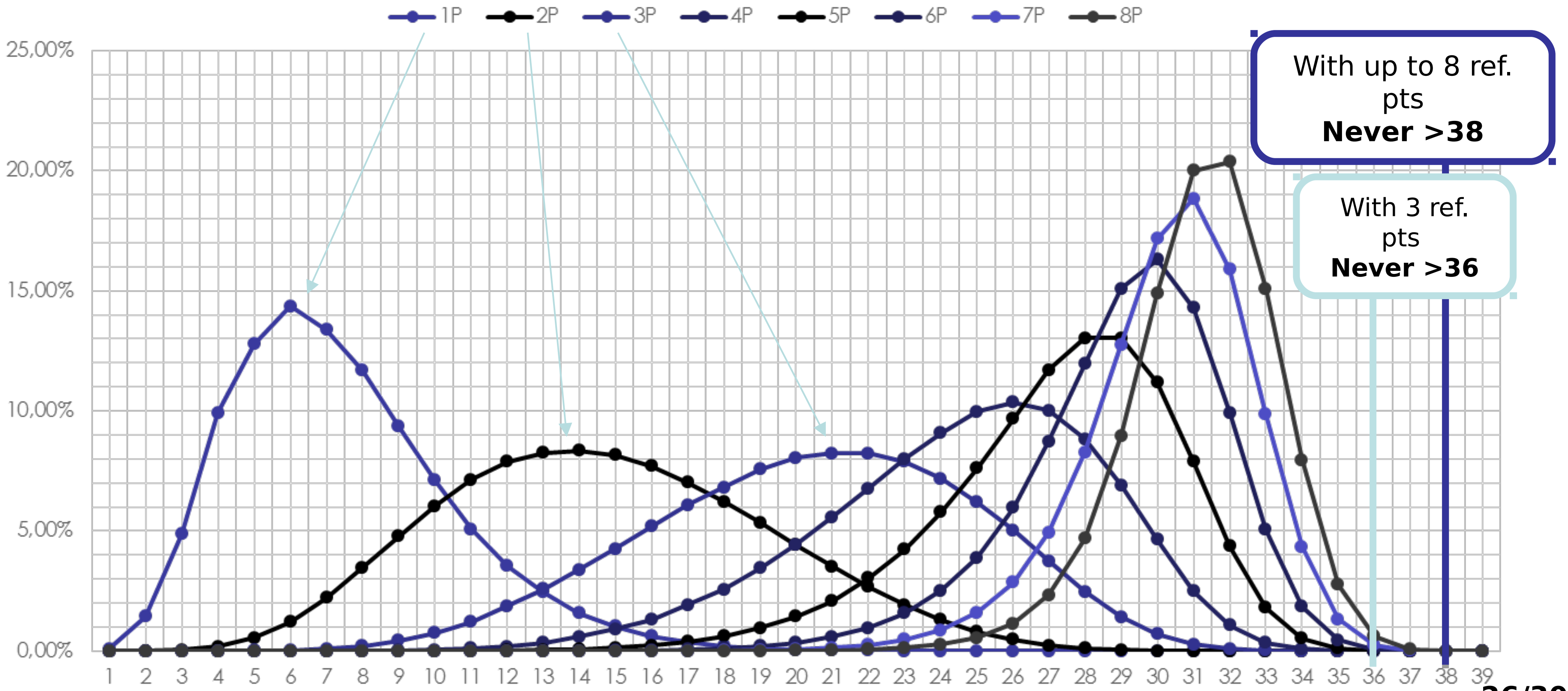
- Q: Why the strict rank. acc. never reaches 100%?
- Q: How many equivalence classes could be reproduced by an S-RMP model?

- Method Monte Carlo:
  - 7 criteria with the same evaluating scales in the case.
  - Randomized and normalized weights (Butler et al., 1997)
  - From 1 to 8 reference points randomly generated
  - 1 000 000 repeated trials for each num. of ref. pts selected

Distribution of the num. of equivalence classes

by the 1 000 000 randomized S-RMP models for each num. of ref. pts

- Apply the randomly generated S-RMP models to the 39 sites in the case



Validation of the metaheuristic

- MIP algorithms

  - The exact algorithm without considering inconsistency (Zheng et al., 2012)
  - The adapted exact algorithm taking account of the inconsistency (Liu et al., 2013)

No Solution!

- Main Performance Index

  - Ranking Accuracy and Strict Ranking Accuracy
  - Number of equivalence classes
  - Computation time

Solution Quality

Existence of multiple solutions

	Num. Ref. Pts	1	2	3
MIP algorithm	R.A.	**78.95%**	**84.21%**	**86.84%**
	Strict R.A.	**71.05%**	**78.95%**	**86.84%**
	Num. Eq. Classes	**19**	**29**	**31**
	Num. Solutions	**1**	**1**	**1**
Metaheuristic algorithm	Max. R.A. 500	**78.95%**	**81.58%**	**81.58%**
	Max. Strict R.A. 500	**60.53%**	**68.42%**	**73.68%**
	Num. Eq. Classes	**20**	**21**	**23**
	Num. Solutions	**1/500**	**4/500**	**1/500**

Thanks to the multitude of solutions

- The decision makers could choose the one which is closer to their thinking from the solution pool.

Computation time

High efficiency of the metaheuristic algorithm

	Num. Ref. Pts	1	2	3
MIP algorithm	Num. cores	**12**	**12**	**12**
	Main frequency	**2.66 GHz**	**2.66 GHz**	**2.66 GHz**
	CPU type	**Intel Xeon X5650**		
	Cluster type	**Altix ICE 8400 LX**		
	Elapsed time	**2.09 h**	**> 22.16 h**	**> 75.15 h**
Metaheuristic algorithm	Num. cores	**4**	**4**	**4**
	Memory	**8 MB**	**8 MB**	**8 MB**
	Main frequency	**2.3 GHz**	**2.3 GHz**	**2.3 GHz**
	CPU type	**Intel Core i7**		
	Elapsed time per trial	**11.67 s**	**17.97 s**	**23.30 s**
	Elapsed time 500 trials	**1.62 h**	**2.50 h**	**3.24 h**

Conclusions

Advantages

- ✓ This is the first time we applied S-RMP model on a real application
- ✓ From this application we know how we can improve the whole decision aiding process involving S-RMP models
- ✓ Through this application we gained a novel understanding of the capacity of S-RMP model
- ✓ The participants in the focus group grasped very well the potential of the tool

Future work

- ✓ Improvement of the metaheuristic algorithm
- ✓ Test the model on other real word case studies
- ✓ Test the model with robustness analysis
- ✓ Further study the S-RMP models' capacity