



**HAL**  
open science

# Probabilistic models for permutations and dependence

Arthur Fétiveau, Gilles Durrieu, Emmanuel Frénod

► **To cite this version:**

Arthur Fétiveau, Gilles Durrieu, Emmanuel Frénod. Probabilistic models for permutations and dependence. 2025. hal-04925265

**HAL Id: hal-04925265**

**<https://hal.science/hal-04925265v1>**

Preprint submitted on 1 Feb 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Probabilistic models for permutations and dependence

Arthur Fétiveau,  
arthur.fetiveau@gmail.com<sup>1</sup>, Gilles Durrieu,  
gilles.durrieu@univ-ubs.fr<sup>1</sup>, and Emmanuel Frénod,  
emmanuel.frenod@univ-ubs.fr<sup>1</sup>

<sup>1</sup>Univ Bretagne Sud, CNRS UMR 6205, LMBA, F-56000 Vannes, France

February 1, 2025

## Abstract

In order to improve the financial performance of a company, we must classify all their commercialized products according to their interests using financial criteria. The criteria and the final classification can be modeled using permutations. We consider Mallows' models defined in the space of permutations. We are particularly interested in the question of dependence in Mallows' models. Here, we introduce new machine learning approaches based on a cost function minimizing the impact of the dependence between criteria. In the multi criteria aggregation model, we consider some of the criteria as a permutation generated using a Mallows' model, in which the modal permutation is the permutation we want to find, while some other criteria are simulated using other criteria. Finally, the methodology is illustrated with a simulation study that compares the performances of the approaches.

Keywords : Computational statistics, Kendall distance, Machine learning, Mallows' models, Permutations.

## 1 Introduction

Imagine we want to rank bikes in an interest order based on different criteria. Then, we use as criterion the color as a ranking based on preferences of the user. Except this ranking, we have no clue on how much favorable a color is against another. We can also use as criterion the number of gear ratios, the saddle size, the stem shape, the wheel resilience, the whole weight of the bicycle, the price and many other criteria. Some variables are quantitatives as the price or the weight and some others are qualitatives as the color or the number of gear ratios.

All of those criteria can help to lead to the best quality bike, but some of those variables are correlated. We know that for sure the weight depends on the weight of all the components of the bike, including some information we already have thanks to

some other criteria. For example, a higher number of gear ratios will inevitably lead to a heavier bike, as well as a bigger saddle or an higher wheel resilience. Another example is the price of the bike which is correlated with a lower weight, the color, the stem shape and all the elements of the bike. However, this variable, also includes some unknown information. We can not ignore this information brought by those variables. That is why we cannot remove them from the analysis.

We can also imagine to use the difference between the real price and the expected price from the customers. The expected price will probably be correlated to the real price of the bike, but we do not really know how the costumer values each component of the bike.

Another problem to not be robust on dependent data is when somebody uses twice the same criterion. In this case, as we evaluate the parameter in total independence between each criterion, when we will determine the consensus, we will count twice the same criteria. This will lead to an unjust favor to this criterion.

We can also see that as the price and the weight of the bike both depend on the number of gear ratios, they also have a dependence link between them.

The first example of dependence, the real link between two criteria due to their formulas is what we will call the redundancy dependence. We can, by changing the link between initial variable and the built variable, for example the number of gear ratios and the price of the bike, see what happen to the built one.

There is a lot of multi-criteria aggregation models, such as the use of the Choquet integral [11], [5] or the Shapley value [14], the methods based on the comparison of elements by pairs as ELECTRE, [8], [18] or PROMETHEE [4], [16], [3] and the methods based on the utility theory [19], [7], among others. In this paper, we have chosen the permutations based models. We are particularly focused on the Mallows' model [15], but we can also think about the Coset Permutation Space model (CPS) [17].

In [10], we developed a new model based on the Mallows' model to take into account the real value of the indicator meaning that for some criterion the gap between their real values had an impact on the result. However, we did not take into account the fact that some variables are, by construction, dependent. We also saw that with the Mallows' model, the maximum likelihood estimator, compared to the Borda method [2] and a quadratic estimator, less likely to have a dictatorship, is the most robust when we add some low informative criteria. It even gained some precision looking for the best consensus.

The Mallows' model is described by the probability of a permutation to be at some distance of a modal permutation  $\pi$ . The decay rate of the probability based on the distance to the modal permutation is modeled by the parameter  $\theta$ .

In this paper, we are interested to study the effect of dependence in the Mallows' model. Then, we propose an alternative approach to be more robust in the presence of dependent data. The objective here is to construct an easy way to adjust the estimations of the parameters in a case where some of the criteria are correlated.

This paper is organized as follows. In Section 2, the Mallows' model is described and the Mallows' model simulation using the Kendall distance is given. The way to simulate

dependent data for our approach is described. In Section 3, the process to construct new variables to explain the redundancy dependence between our criteria is described. In Section 4 the correction applied on the parameters in order to correct the effect of the dependence on the estimates are given along with a method to estimate it directly. Finally, in Section 5 we compare the results on simulations studies. Lastly, concluding remarks are given in Section 6.

## 2 Data Simulation

### 2.1 Simulation of the Mallows' model

The Mallows' model [15] is defined as the probability to select randomly a permutation  $\sigma$ , in the set of all the permutations  $\mathcal{S}_n$ , based on the permutation  $\pi$  with a dispersion parameter  $\theta$ . We have

$$P_\theta(\sigma) = \frac{\exp(-\theta d(\sigma, \pi))}{Z(\theta)} \quad (1)$$

with  $\theta \geq 0$ ,  $d(., .)$  a right invariant distance between two permutations of  $\mathcal{S}_n$  and

$$Z(\theta) = \sum_{\sigma \in \mathcal{S}_n} \exp(-\theta d(\sigma, \pi)), \quad (2)$$

the normalization term which is not influenced by  $\pi$ , thanks to the right invariant property. A distance  $d(., .)$  is right invariant [6] if  $d(\pi, \sigma) = d(\pi\tau, \sigma\tau)$  for every  $\pi$  and  $\sigma$  in  $\mathcal{S}_n$ . The normalization term  $Z(\theta)$  only varies with  $\theta$ .

The Mallows' model is based on the distances between the permutations of a set of  $n$  items to order. A permutations is a one to one application from  $\{1, \dots, n\}$  to itself. The set of all the permutations of  $n$  items is denoted by  $\mathcal{S}_n$ . We write  $\pi(i)$  the position associated to the item  $i$  in the permutation named  $\pi$  for  $i \in \{1, \dots, n\}$  and  $\pi^{-1}(i)$  the item associated to the  $i^{\text{th}}$  position of the permutation named  $\pi$  for  $i \in \{1, \dots, n\}$ . The cardinal of  $\mathcal{S}_n$  is equal to  $n!$ .

Using the Mallows' model, or the generalized version, we can simulate an order, knowing the real value of  $\pi$  and  $\theta$ .

In this framework, we consider the permutation  $\pi$  as the perfect theoretical ranking. This permutation  $\pi$  is the one we want to determine.

We recall that the definition of the Kendall distance given in [13].

**Definition 1.** *The Kendall distance represents the minimal number of adjacent transpositions to transform a permutation  $\sigma$  into a permutation  $\pi$ , both from the set of permutations  $\mathcal{S}_n$ .*

$$d_k(\pi, \sigma) = \sum_{i=1}^{n-1} \sum_{j>i} \mathbb{1}((\pi(i) > \pi(j) \cap \sigma(i) < \sigma(j)) \cup (\pi(i) < \pi(j) \cap \sigma(i) > \sigma(j))) \quad (3)$$

where  $\mathbb{1}(\mathcal{P})$  takes the value 1 or 0 depending on whether the condition  $\mathcal{P}$  is satisfied or not. This distance is right invariant.

Considering the Kendall distance (3) in the Mallows' model and the generalized Mallows' model, [9] and [10], we generate permutations using the product of probabilities to move each item of  $\pi$  with their followings, we can find a R package in [12] based on ideas in [1].

**Theorem 1.** *For each item of  $\pi$ , the movement of this item with a number of following elements is independent of the previously moved items. We denote by  $\zeta(\pi^{-1}(i))$  the number of displacements of the  $i^{\text{th}}$  item of the permutation  $\pi$  with the following items. The probability to move an item of a number  $x \leq (n - i)$  of following items is given by*

$$\mathbb{P}_\theta (\zeta(\pi^{-1}(i)) = x) = \frac{\exp(-\theta x)}{\sum_{j=0}^{n-i} \exp(-\theta j)}. \quad (4)$$

*Proof.* Rewriting the Kendall distance (3) using  $\pi^{-1}$  and  $\sigma^{-1}$ , we obtain

$$d_k(\pi, \sigma) = \sum_{i=1}^{n-1} \sum_{j>i} \mathbb{1}(\sigma(\pi^{-1}(i)) > \sigma(\pi^{-1}(j))). \quad (5)$$

We can split the equation over each element of  $\pi$ . Also, we can consider a permutation  $\sigma \in \mathcal{S}_n$  as a sequence of transposition of each element  $i$  of  $\pi$ . Then the distance between  $\pi$  and  $\sigma$  can be expressed as the sum of this sequence.

So, we have

$$\begin{aligned} \exp(-\theta d(\pi, \sigma)) &= \exp\left(-\theta \sum_{i=1}^{n-1} \zeta(\pi^{-1}(i))\right) \\ &= \prod_{i=1}^{n-1} \exp(-\theta \zeta(\pi^{-1}(i))), \end{aligned} \quad (6)$$

where  $\zeta(\pi^{-1}(i))$  is the number of following items the  $i^{\text{th}}$  item of  $\pi$  needs to be transposed with. And  $\zeta(\pi^{-1}(i))$  is an integer and takes its possibilities in  $[0, 1, \dots, n - i]$ .

If we want to sum all the possibles  $\sigma \in \mathcal{S}_n$  then we can reduce it to a product of sum. By recurrence, we have for  $n = 2$ ,

$$Z(\theta) = \exp(-\theta \times 0) + \exp(-\theta \times 1), \quad (7)$$

and for  $n$  items, we have

$$Z(\theta) = (\exp(-\theta \times 0) + \exp(-\theta \times 1)) \times \dots \times (\exp(-\theta \times 0) + \dots + \exp(-\theta \times (n - 1))). \quad (8)$$

Then for  $n+1$  elements, for all of the possibilities of the  $n$  last elements of  $\pi$  we need to multiply every possibility of movement of the first element, leading to

$$\begin{aligned} Z(\theta) &= ((\exp(-\theta \times 0) + \exp(-\theta \times 1)) \times \dots \times (\exp(-\theta \times 0) + \dots + \exp(-\theta \times (n - 1)))) \\ &\quad \times (\exp(-\theta \times 0) + \dots + \exp(-\theta \times n)). \end{aligned} \quad (9)$$

So finally, we can write

$$Z(\theta) = \prod_{i=1}^{n-1} \sum_{j=0}^{n-i} \exp(-\theta j). \quad (10)$$

Then, the probability of moving the  $i^{\text{th}}$  item independently of any other movement, is given by

$$\mathbb{P}_\theta(\zeta(\pi^{-1}(i)) = x) = \frac{\exp(-\theta x) \times \prod_{k=1, k \neq i}^{n-1} \sum_{j=0}^{n-k} \exp(-\theta j)}{\prod_{k=1}^{n-1} \sum_{j=0}^{n-k} \exp(-\theta j)}. \quad (11)$$

We deduce

$$\mathbb{P}_\theta(\zeta(\pi^{-1}(i)) = x) = \frac{\exp(-\theta x)}{\sum_{j=0}^{n-i} \exp(-\theta j)}, \quad (12)$$

where the denominator can be rewritten using the property of the sum of a geometric suite.  $\square$

Using the probability given in (4), we can represent the number of transposition with the following elements of each element in the permutation  $\pi$ . Then, we build the unique permutation associated with those movements.

As an example in Figure 1, we represent an heatmap in the space of permutations  $\mathcal{S}_5$  of the simulation of 10,000 permutations with a parameter  $\theta = 0.5$  around the permutation  $\pi$  the identity permutation. In this representation, each point represent a permutation. The numbers [1 2 3 4 0] mean that the first element, denoted by 0 is in fifth position while the second element denoted by 1 is in first position. It is a representation of  $\pi^{-1}$ .

We recall that the generalized Mallows' model for the Kendall distance is given by

$$P_\theta(\sigma) = \frac{1}{Z(\theta)} \exp \left( \sum_{k=1}^{n-1} -\theta_k \zeta_\sigma(\pi^{-1}(k)) \right), \quad (13)$$

where

$$Z(\theta) = \sum_{\sigma \in \mathcal{S}_n} \exp \left( \sum_{k=1}^{n-1} -\theta_k \zeta_\sigma(\pi^{-1}(k)) \right) \quad (14)$$

corresponds to the normalization term not influenced by  $\pi$  and  $\zeta_\sigma(\pi^{-1}(k))$  is the number of following items the  $i^{\text{th}}$  item of  $\pi$  needs to be transposed with to match the  $\sigma$  permutation. Finally,  $\theta$  is the vector of parameters  $\theta_k$  for  $k = 1, \dots, n-1$ . The Mallows' model is a particular case of the generalized Mallows' model with  $\theta = \theta_k$ , for  $k = 1, \dots, n-1$  when all  $\theta_k$  are the same.

The second example in Figure 2 represents an heatmap in the space of permutation  $\mathcal{S}_5$  of the simulation of 10,000 permutations with a generalized Mallows' model (13) with

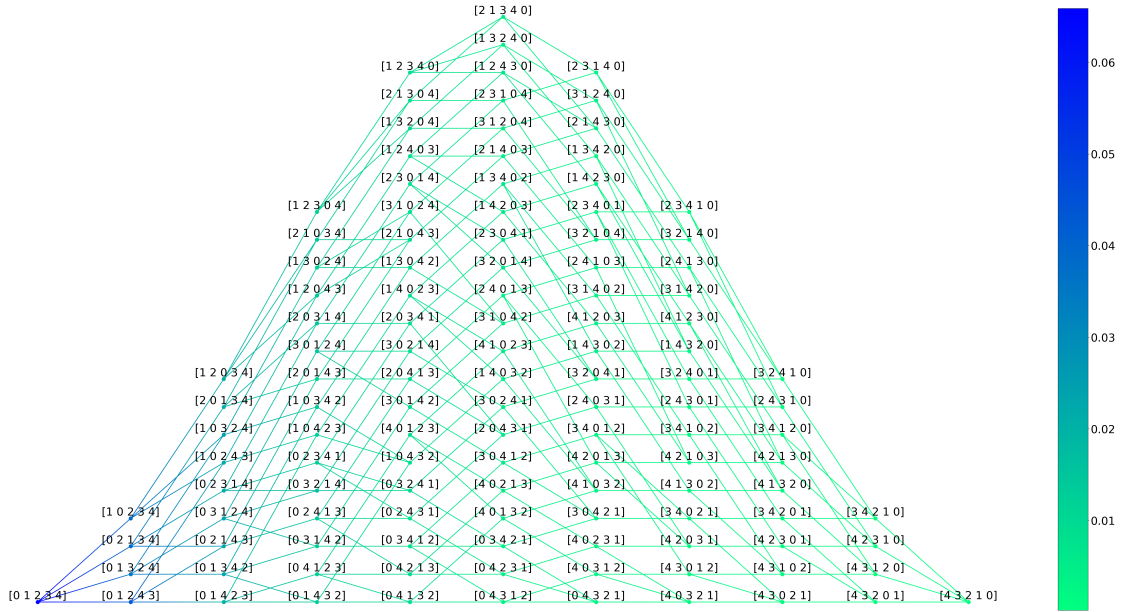


Figure 1: Simulation of 10,000 permutations using the Mallows' model. Each link indicates a value of the Kendall distance of 1 between 2 permutations. The gradient of color is between green and dark blue. The green color represents the lowest probabilities while the dark blue represents the highest probabilities.

a parameter  $\theta = [0.1, 0.5, 1, 1]$  around the permutation  $\pi$  the modal permutation. With those values of  $\theta$ , we observe that the first element is easy to move backward, the second slightly less and the lasts are really difficult to move backward.

## 2.2 Simulating chained Mallows' models

We define the chain Mallows' models as a succession of Mallows' models. We consider the model (1) given by

$$P_{\theta}(\sigma) = \frac{\exp(-\theta d(\sigma, \pi))}{Z(\theta)}$$

where the parameter  $\pi$  is given by the permutation  $\sigma$  obtained using the same equation (1) at the previous step. This process is repeated. Consequently, the chained Mallows' model is a repetition of Mallows' models centered on the result of the previous link of the chain. The chained Mallows' model of two Mallows' models is given by

$$P_{\theta_0, \theta_1}(\sigma_1) = \sum_{\sigma_0 \in S_n} \frac{\exp(-\theta_1 d(\sigma_1, \sigma_0))}{Z(\theta_1)} \times \frac{\exp(-\theta_0 d(\sigma_0, \pi))}{Z(\theta_0)}. \quad (15)$$

We are interested in chains of any length. Here we represent graphically what happens in the case of a chained Mallows' model and the influence of the parameters  $\theta$  at each

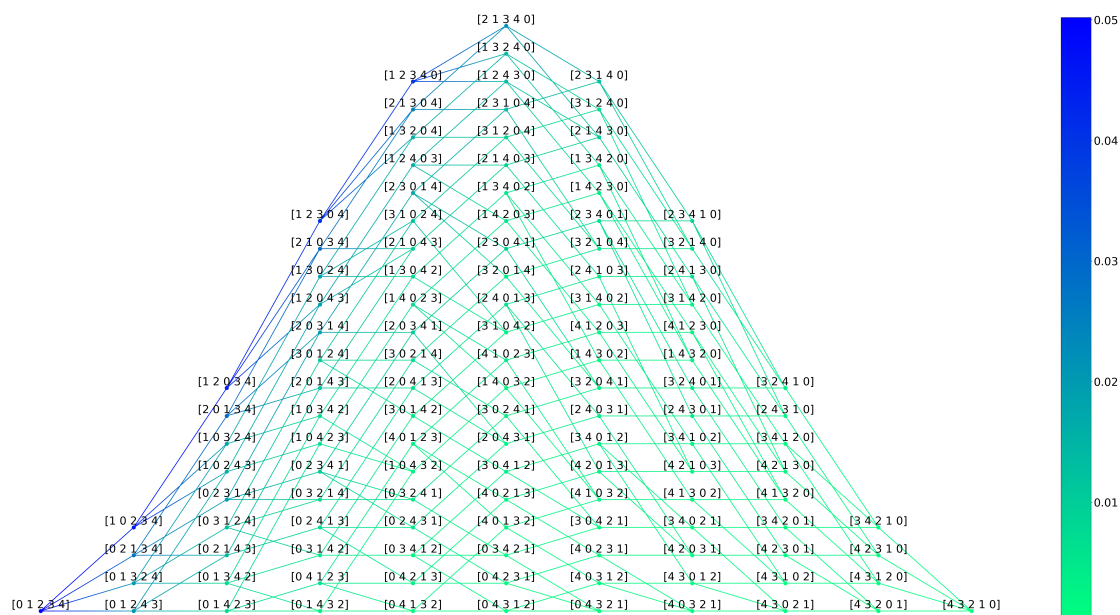


Figure 2: Simulation of 10,000 permutations using the generalized Mallows' model. Each link indicates a value of the Kendall distance of 1 between 2 permutations. The gradient of color is between green and dark blue. The green color represents the lowest probabilities while the dark blue represents the highest probabilities.

step of the chain. For example, with a dispersion parameter  $\theta$  at any link chain set to 0, the influence of the first  $\pi$ , the modal permutation will have no importance at all. Meaning that the last simulated  $\sigma$  could be any permutation with the same probability. However, if any dispersion parameter  $\theta$  at any link of the chain goes to infinity, then the link chain associated would be completely useless in the chain. Now, the question remains of the importance of every  $\theta$  that are not a special case, the importance of their value, their position in the chain and obviously, the length of the chain. It seems that the longer the chain, the more random the final permutation  $\sigma$ .

By calculation of each possible permutation's probability for a chained Mallows of two successive Mallows, the chained Mallows does not lead to a Mallows' model. We compute the chained Mallows for each possibility when  $n = 3$  with different values of  $\theta$  for each step of the chain. As the Kendall distance is right invariant, the initial modal permutation  $\pi$  of the first step of the chain does not matter, so we will use the identity permutation  $[0 1 2]$ .

Using Table 1, we can find all the distances to go from a permutation  $\pi$  to another permutation  $\sigma$ . When we only have one intermediate state, we have for  $n = 3$  six path to go to each possible outcome. Indeed, the intermediate permutation can be any of the six permutations of length  $n = 3$ . For example, to go from the identity permutation  $\pi [0 1 2]$  to the final permutation  $\sigma [1 2 0]$ , we can go through the intermediate permutation  $[1 0 2]$



Permutations	[0 1 2]	[0 2 1]	[1 0 2]	[1 2 0]	[2 0 1]	[2 1 0]
[0 1 2]	0	1	1	2	2	3
[0 2 1]	1	0	2	3	1	2
[1 0 2]	1	2	0	1	3	2
[1 2 0]	2	3	1	0	2	1
[2 0 1]	2	1	3	2	0	1
[2 1 0]	3	2	2	1	1	0

Table 1: Kendall distances between permutations.

obtained with probability

$$P_{\theta_0}([1 0 2]) = \frac{\exp(-\theta_0)}{Z(\theta_0)}$$

where  $Z(\theta_0) = \sum_{\sigma \in \mathcal{S}_n} \exp(-\theta_0 d(\sigma, [1 0 2]))$ . Then, the probability of the permutation  $\sigma [1 2 0]$ , knowing the intermediate state  $[1 0 2]$ , is given by

$$P_{\theta_1}([1 2 0]) = \frac{\exp(-2\theta_1)}{Z(\theta_1)}$$

where  $Z(\theta_1) = \sum_{\sigma \in \mathcal{S}_n} \exp(-\theta_1 d(\sigma, [1 0 2]))$ .

0	1	2	3
(0, 0)	(0, 1)	(0, 2)	(0, 3)
(1, 1)	(1, 0)	(2, 0)	(3, 0)
(1, 1)	(1, 2)	(1, 1)	(1, 2)
(2, 2)	(2, 1)	(2, 2)	(1, 2)
(2, 2)	(2, 3)	(3, 1)	(2, 1)
(3, 3)	(3, 2)	(1, 3)	(2, 1)

Table 2: List of Kendall distances by step for each possible distance outcome given in first row. Each row represents a different intermediate permutation.

For  $n = 3$  and only one intermediate step, we have in Table 2 the exhaustive list of distances for each possible outcome distance.

Finally, the probabilities of each permutation after a chain of only one intermediate step and permutations of length  $n = 3$  are given by

$$P_{\theta_0, \theta_1}([0 1 2]) = \frac{1}{Z} (1 + 2 \exp(-\theta_0 - \theta_1) + 2 \exp(-2\theta_0 - 2\theta_1) + \exp(-3\theta_0 - 3\theta_1)), \quad (16)$$

$$P_{\theta_0, \theta_1}([0 2 1]) = P_{\theta_0, \theta_1}([1 0 2]) = \frac{1}{Z} (\exp(-\theta_0) + \exp(-\theta_1) + \exp(-2\theta_0 - \theta_1) + \exp(-\theta_0 - 2\theta_1) + \exp(-2\theta_0 - 3\theta_1) + \exp(-3\theta_0 - 2\theta_1)), \quad (17)$$

$$P_{\theta_0, \theta_1}([1\ 2\ 0]) = P([2\ 0\ 1])_{\theta_0, \theta_1} = \frac{1}{Z} (\exp(-2\theta_0) + \exp(-2\theta_1) + \exp(-\theta_0 - \theta_1) + \exp(-2\theta_0 - 2\theta_1) + \exp(-\theta_0 - 3\theta_1) + \exp(-3\theta_0 - \theta_1)) \quad (18)$$

and

$$P_{\theta_0, \theta_1}([2\ 1\ 0]) = \frac{1}{Z} (\exp(-3\theta_0) + \exp(-3\theta_1) + 2\exp(-2\theta_0 - \theta_1) + 2\exp(-\theta_0 - 2\theta_1)), \quad (19)$$

where

$$Z = \sum_{\pi_1 \in \mathcal{S}_n} \sum_{\sigma \in \mathcal{S}_n} \exp(-\theta_0 d(\pi, \pi_1)) \exp(-\theta_1 d(\pi_1, \sigma)) \quad (20)$$

is the sum of all the numerators of the different possibilities of  $\sigma$ .

The probabilities given in (16), (17), (18) and (19) do not depend on the order of  $\theta_0$  and  $\theta_1$ . We have the same property as the Mallows' model, we can invert  $\pi$  and  $\sigma$  and the probability will still be the same.

It appears that two successive Mallows' models do not lead to a Mallows' model. However, simulation study of a chain Mallows' model leads us to approximate the results by a Mallows' model. Using this approximation, we study the parameter of dispersion of the permutations probabilities, given by the parameters  $\theta_0$  and  $\theta_1$  in every step of the chained Mallows' model.

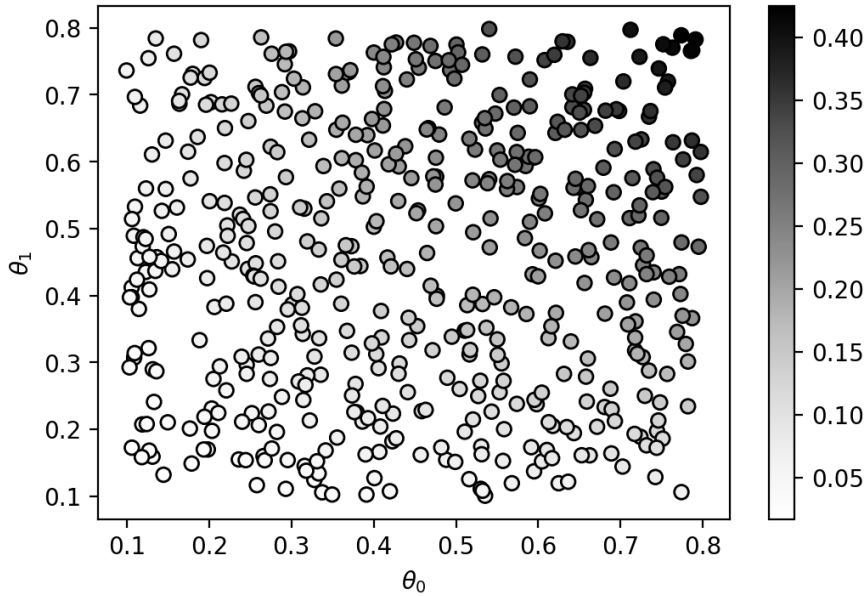


Figure 3: Estimation of  $\theta$  considering the Mallows' model on  $\sigma$  simulated by a chained Mallows' model with different values of  $\theta_0$  and  $\theta_1$  on  $[0.1; 0.8]$ . The gradient of color is between white and black. The white color represents the lowest values of  $\hat{\theta}$  while the black represents the highest values of  $\hat{\theta}$ .

In Figure 3, we represent the values of estimator  $\hat{\theta}$  of  $\theta$  in the case of an approximation of a chained Mallows' model using a Mallows' model. The values of  $\theta_0$  and  $\theta_1$  used to simulate the chained Mallows are given in the two axis. We observe in this Figure 3 that the decrease of the estimated  $\hat{\theta}$  is not linear with the original values of  $\theta$ 's. In fact, the  $\hat{\theta}$  value converges quickly to 0 with a chained Mallows. We conclude that the result tends to be equiprobable for each possible permutation after only a few iterations of Mallows' models.

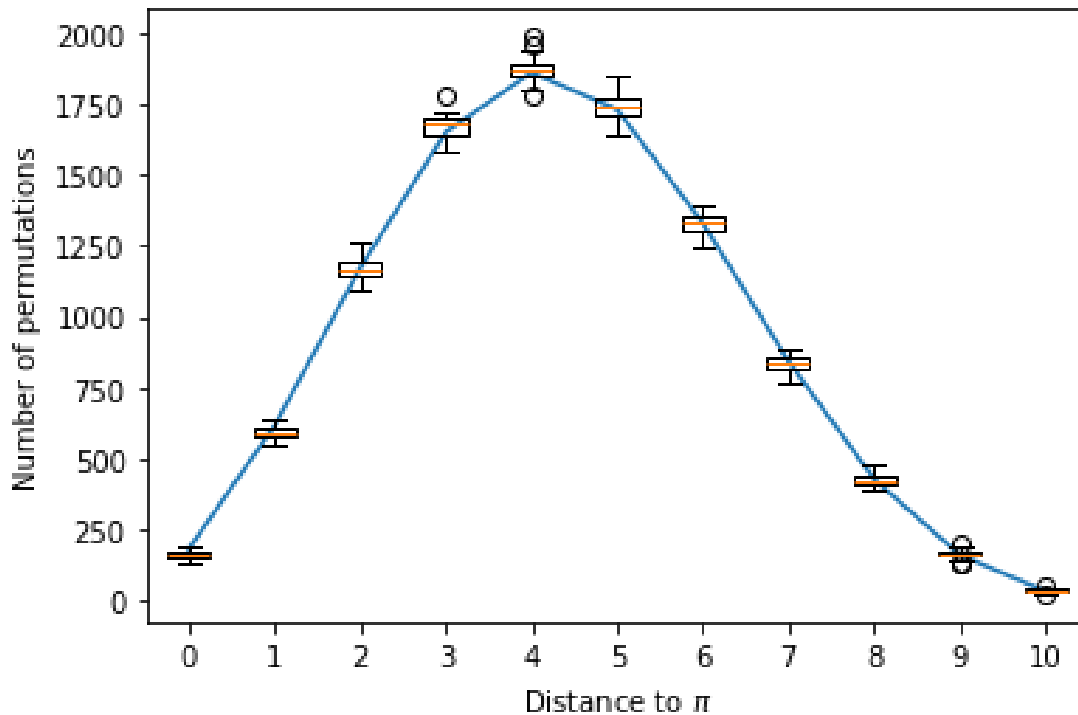


Figure 4: Goodness-of-fit of a Mallows' model on simulated chained Mallows' data.

We simulate 500 different replicates of parameters  $\theta_0$  and  $\theta_1$  for the chained Mallows' models. The parameters  $\theta_0$  and  $\theta_1$  are each uniformly distributed on  $[0.1; 0.8]$ . The chain is composed of two consecutive Mallows' models. For each set of parameters, we carry out 50 iterations of 10,000 permutations of size  $n = 5$ . We do not reject the null hypothesis " $H_0 : \sigma$  follows Mallows' model of parameters  $\theta$  and  $\pi$ " (p-value  $> 5\%$ ,  $\chi^2$  goodness-of-fit test). For instance in Figure 4, the p-value is equal to  $0.96 > 0.05$ . In this Figure, the blue curve represents the theoretical number of permutations at each distance of  $\pi$  in the case of a Mallows' model of dispersion parameter  $\theta$  adjusted to the values we have simulated. The boxplot represents the real values obtained for the considered iteration.

### 2.3 Simulation of criteria

To simulate the value of a secondary criterion using a set of the primary criteria simulated earlier, we use a known function  $f$

$$X_d = f(X_1, \dots, X_N), \quad (21)$$

where  $X_d$  is the value of the secondary criterion simulated and  $N$  is the number of primary criteria. For example, we can consider  $f(X_1, \dots, X_N) = X_1 + X_2 + X_N$  or  $f(X_1, \dots, X_N) = X_1 \times X_2$ . We can also consider  $f(X_1, \dots, X_N) = -X_2$  which will give the exact inverse order.

The values of the primary criteria  $X_1, \dots, X_N$ , are simulated according to the lognormal distribution. The simulated values are set to match the order given by the Mallows' model.

### 2.4 Choice of the length of the permutations

We evaluate the differences between bootstrapping  $m$  in  $n$  elements and simulating  $n$  elements. As a constraint of our data, we only have one permutation of each criteria  $\sigma_j$ ,  $j$  in  $1, \dots, J$  each month. To bypass the problem and add some variations in the results, we made a bootstrap of 5 values in the  $n$  values of  $\sigma_j$ . The normalized Kendall distance of 5 randoms elements out of  $n$ , between two rankings, tends to be the same as the normalized distance of the  $n$  elements. This distance however, is not the same as the distance obtained when we simulate permutations of 5 elements using the Mallows' model with the same value of parameter  $\theta$ . In the Figure 5, we represent the normalized Kendall distances of the 3 possibilities. We simulate with the Mallows' model  $M = 100$  permutations  $\sigma_m$ , for  $m = 1, \dots, M$ , using a parameter  $\theta = 0.1$  and  $\pi$  the identity permutation for  $n = 100$  elements. In black, we plot the average values of the normalized Kendall distance between  $\sigma_m$  and  $\pi$  for  $m$  in  $\{1, \dots, M\}$ . In those permutations  $\sigma_m$ , we chose randomly  $n$  elements  $k = 100$  times and compute the Kendall distance between the identity permutation of those  $n$  elements and the permutation  $\sigma_m$  with only those  $n$  elements. In red dashed, we plot the average values of the normalized Kendall distance for a selection of  $n$  elements. Finally, in blue dashed and dotted, we simulate for  $n \in \{5, \dots, 100\}$   $M = 100$  permutations  $\sigma_m$ , for  $m = 1, \dots, M$ , using a parameter  $\theta = 0.1$  and  $\pi$  the identity permutation, and plot the average values of the normalized Kendall distances between  $\pi$  and  $\sigma_m$ .

The Figure 5 implies that, while using the bootstrap of a permutation, we must keep in mind the real value of  $n$  to keep the same  $\theta$ .

## 3 Indicators of Proximity

As we want to adjust our Mallows' model for useless dependent criteria, we need to determine how to define the proximity between two criteria on the permutation space. To this end, as we know the construction of the secondary criterion value  $X_d$  based on

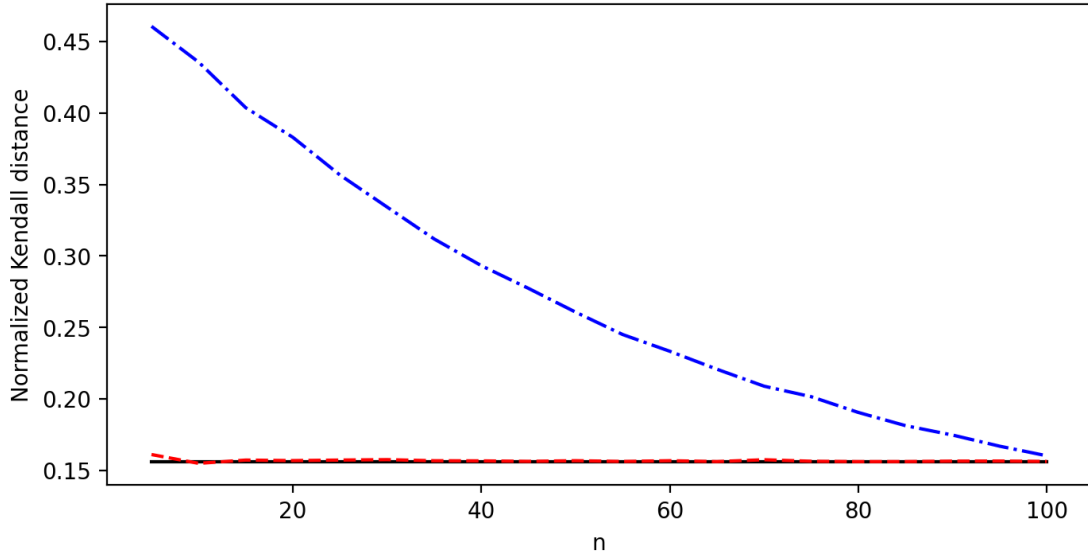


Figure 5: Normalized Kendall distance for  $\theta = 0.1$ .

the values of the primary criteria  $X_i$  for  $i \in \{1, \dots, N\}$  using (21), we made the values  $X_i$  change by a factor, and determine in which way the two permutations will differ one from the other. Our method also allows the comparison of two secondary criteria linked by the same primary criterion. By altering the primary criterion, the two derived criteria will change. They can change differently so we collect data on how they differ.

Theoretically for  $i \in \{1, \dots, N\}$  and for the comparison between a primary and a secondary criterion, if the factor applied to the primary criterion value  $X_i$  is strong enough, then, the permutation associated to the secondary criterion value  $X_d$  becomes the same as the permutation associated to the primary criterion value  $X_i$ . By opposition, if this factor is weak enough, and  $X_d$  is not identical to  $X_i$  then the permutation associated to  $X_d$  should differ from the permutation associated to  $X_i$ .

We want to describe the effects between two dependent criteria by construction when we apply a change to the primary criterion which link them. To this end, we collect some information. We choose to use 11 factors to apply to our primary criterion value  $X_i$ .

For each of the 11 factor, from  $10^{-5}$  to  $10^5$ , applied to the primary criterion, we compute for a bootstrap of sub-sample of 5 elements the distance between the original permutation of both criteria and the permutation modified by the alteration of the primary criterion and the new distance between them. We also compute the movement of each element with its following from the beginning and the ending of the permutation. As we compute the movement of the elements with their followings, we have, for 5 elements, 4 variables. Starting from the beginning and the ending with both of the criteria, this count as 16 variables for each sample.

Then we consolidates the results by calculating the mean and standard deviations

of all those distances. We also compute, over the permutations found for the movement of both secondary criteria to themselves and the one found for the movement between them, the modal permutation of those. With this modal permutation, we compute its distance to the identity permutation to see if it goes somewhere else or if it stays around the same position. The most likely of the 3 to go somewhere else is the movement between the two criteria. We also compute the mean and standard deviation of the distance between the modal permutation found and each permutation found for every iteration. Those are the consolidation of the last 3 variables given in the last paragraph.

We do not forget that we have only a few iterations of the rankings but with  $n$  really large. That's why we compute the results on a bootstrap of 5 elements, to have some diversity.

With all those information, we have created for bootstrap of  $n = 5$  elements  $4 \times 8 + 3 \times 2 + 3 = 41$  indicators on the dependence for each of the 11 factor applied. Meaning, for our simulation, we have a total of 451 variable to understand the modification to apply to our parameters to get rid of the dependence. This is by far too much. In our modelisation, we use the mean distance and standard deviation of the 3 compared movement, but, about the movement of each element, we chose to only keep the mean distance and to not use the standard deviation. Moreover, we reduce for each of the 16 distances from 11 observations to 4 parameters to model those observations by a logit modelisation based on the modification of the primary criterion.

With those reduction, we keep only, if we can say so, 130 variables to describe the dependence of our data. With those computed variables, we want to produce only  $J$  value, a redundancy coefficient to apply to the maximum likelihood estimator of  $\theta_j$  with  $j \in 1, \dots, J$ ,  $J$  the number of criteria, computed as if we were on an independent modelisation. In the future, we will call these 130 variables : proximity indicators, PI.

## 4 Methods to adjust dispersion parameters with dependent permutations

For the Kendall distance (3), we showed in [9], [10] that the maximum likelihood estimator  $\hat{\theta}_j$  of  $\theta_j$  is solution in  $\theta_j$  of the following equation

$$\frac{n \exp(-\theta_j)}{1 - \exp(-\theta_j)} - \sum_{k=1}^n \frac{k \exp(-k\theta_j)}{1 - \exp(-k\theta_j)} = \frac{1}{T} \sum_{t=1}^T d(\sigma_{jt}, \pi_t), \quad (22)$$

with  $j \in \{1, \dots, J\}$  a criterion and  $t \in \{1, \dots, T\}$  our observations of the criteria and the real modal permutation  $\pi_t$ . As Kendall distance is right invariant, we do not need to have the same  $\pi_t$  for each  $t$ .

Then, the estimator  $\hat{\pi}_t$  of  $\pi_t$  is solution of the maximization problem

$$\hat{\pi}_t = \operatorname{argmax}_{\pi \in \mathcal{S}_n} \sum_{j=1}^J -\hat{\theta}_j d(\sigma_{jt}, \pi) \quad (23)$$

where  $\mathcal{S}_n$  is the set of all possible permutations of  $n$  elements, see [10], Section 3.1, page 6 for a way to compute  $\hat{\pi}_t$ .

#### 4.1 Correction procedure

We propose a correction procedure based on the maximum likelihood estimation. The new estimator  $\tilde{\theta}_j$  of  $\theta_j$  is

$$\tilde{\theta}_j = \frac{\hat{\theta}_j}{1 + \sum_{k=1}^J c(X_j, X_k)} \quad (24)$$

where  $\hat{\theta}_j$  is solution of equation (22) and  $c(X_j, X_k)$  is the redundancy coefficient between the criterion  $j$  and the criterion  $k$ . This coefficient takes its values in  $[0; 1]$ . If the two criteria are identical, then  $c(X_j, X_k) = 1$ . If they are independent then  $c(X_j, X_k) = 0$ .

When one of the two criteria is a primary criteria, the computation of  $c(X_j, X_k)$  is quite straightforward as we have only one source of variables describing the dependence. When both the criteria are secondary, we need to take the information coming from all the primary sources used to determine those criteria.

In the case of multiple primary criteria, the redundancy coefficient can be written

$$c(X_j, X_k) = 1 - \prod_{p \in \mathcal{P}} \hat{c}_p(X_j, X_k) \quad (25)$$

where  $\mathcal{P}$  is the set of primary criteria and  $\hat{c}_p(X_j, X_k)$  is the estimated value of the coefficient due to the proximity coming from the primary criterion value  $X_p$ . Here,  $\hat{c}_p(X_j, X_k)$  is equal to 1 if the two criteria are independent and 0 if they are the same.

The estimation  $\hat{c}_p(X_j, X_k)$  of  $c_p(X_j, X_k)$  can be modeled by

$$\hat{c}_p(X_j, X_k) = \kappa(\text{PI}) \quad (26)$$

where PI are the proximity indicators. To learn this model  $\kappa$ , we need to have *a priori* good estimations of  $\hat{c}_p(X_j, X_k)$ . Here we use a random forest model for  $\kappa$ .

To find those *a priori*, the objective is to minimize for all  $j, k \in \{1, \dots, J\}$  and  $p \in \mathcal{P}$ , the following cost function

$$\Psi(\hat{c}_p(X_j, X_k)) = \sum_{t=1}^T d(\hat{\pi}_t, \pi_t) \quad (27)$$

where  $\hat{\pi}_t$  is the maximum likelihood estimator given in (23) replacing  $\hat{\theta}_j$  by  $\tilde{\theta}_j$ .

It is not possible to evaluate  $\Psi$  by a gradient descent method as it is a stair function. The cost remains constant until the consensus permutation  $\hat{\pi}_t$  changes. We then propose to improve the learning by giving directions for the minimization. This creates an artificial slope on the stairs. For each time  $t$ , our new cost is based on the likelihood function used to estimate  $\hat{\pi}_t$ .

We know that  $\hat{\pi}_t$  is a permutation maximizing  $\sum_{j=1}^J \tilde{\theta}_j d(\sigma_{jt}, \hat{\pi}_t)$  so minimizing the weighted distance to each  $\sigma_{jt}$ . There is no necessity for  $\pi_t$  to maximize this sum. Thus, we add to our cost function  $\Psi$  the difference between  $\sum_{j=1}^J \tilde{\theta}_j d(\sigma_{jt}, \hat{\pi}_t)$  and  $\sum_{j=1}^J \tilde{\theta}_j d(\sigma_{jt}, \pi_t)$ . This has as a consequence to force the weighted distance of the theoretical permutation  $\pi_t$  with  $\sigma_{jt}$  to be closer to the weighted distance of the estimated permutation  $\hat{\pi}_t$  with  $\sigma_{jt}$ . We consider for all  $j, k \in \{1, \dots, J\}^2$  and  $p \in \mathcal{P}$  the cost function

$$\tilde{\Psi}(\hat{c}_p(X_j, X_k)) = \sum_{t=1}^T \left( d(\hat{\pi}_t, \pi_t) + \left( \sum_{j=1}^J -\tilde{\theta}_j d(\sigma_{jt}, \hat{\pi}_t) - \sum_{j=1}^J -\tilde{\theta}_j d(\sigma_{jt}, \pi_t) \right) \right). \quad (28)$$

For the computation, the gradient descent method can be used to optimize the cost function  $\tilde{\Psi}$ .

With the best  $\hat{c}_p(X_j, X_k)$  computed for a bunch of simulated data, we learn the link  $\kappa$  between it and the proximity indicators PI. The aim is to have a model  $\kappa$  that we do not need to recompute for different kinds of link between the criteria. This model has to be the more general possible. Also, as we can not compute easily  $\hat{c}_p(X_j, X_k)$  with real data, in contrast to the proximity indicators. It is then really interesting for us to be able to compute  $\kappa$  on simulated data and to use it on real data.

## 4.2 Dispersion parameters corrections on estimators

To have a point of comparison, we also learn all the estimators  $\check{\theta}_j$  of the parameters  $\theta_j$  for  $j \in \{1, \dots, J\}$  where  $J$  is the number of criteria, at the same time. To achieve this, we use the cost given in (28) without the correction on the  $\theta_j$ . Instead, we learn directly  $\check{\theta}_j$ . We can rewrite the expression (28) as

$$\tilde{\Psi}(\check{\theta}_1, \dots, \check{\theta}_J) = \sum_{t=1}^T \left( d(\hat{\pi}_t, \pi_t) + \left( \sum_{j=1}^J \check{\theta}_j d(\sigma_{jt}, \hat{\pi}_t) - \sum_{j=1}^J \check{\theta}_j d(\sigma_{jt}, \pi_t) \right) \right). \quad (29)$$

For  $j \in \{1, \dots, J\}$ , this new expression offers a more general framework for each estimator  $\check{\theta}_j$  of  $\theta_j$  but we need to use a lot of information from the past. With the method described in Section 4.1, we use the Proximity's Indicators PI to escape the difficult learning phase with a smaller model to connect those to the corrections to apply.

With this method however, we are able to compute more accurate estimations of  $(\theta_1, \dots, \theta_J)$  the vector of dispersion parameter of all criteria. At the end, this method must lead to better results. The method with the proximity indicators PI goes one step further trying to compute a simpler model. This method assumes that we have independence between the primary criteria, but those primary variables could be dependent. This is in particular the case when we measure the primary criteria. For the sake of example, we can think of a company that have as a criterion the revenue of a product in January and February. Those two criteria could be considered as primary variables



as we can only measure them. However, we can say that those two values are similarly impacted by the product and differently impacted by its seasonality. In this case, we could say that the only primary variables are the product and the seasonality.

## 5 Comparisons by simulation

In order to evaluate the performance of our methods correcting the dependence in the Mallows' model, we generate some data. We want first to generate independent criteria, called primary criteria, the only criteria related to  $\pi$ , the consensus we want to find. All the dependent criteria, the secondary criteria, are then determined on those criteria related to  $\pi$ .

For the simulation, we generate 12 criteria. Among those criteria, we simulate only 3 primary criteria, denoted by  $s1$ ,  $s2$  and  $s3$ , using a Mallows' model. We generate respectively according to the criteria the dispersion parameters  $\theta_1$ ,  $\theta_2$  and  $\theta_3$ , using an uniform law in the interval  $[0.1; 0.8]$ . We generate also an useless criteria  $s0$ . This useless criteria is just a random permutation in the set  $\mathcal{S}_n$ .

We simulate 5 of the secondary criteria using a chained Mallows' model, as seen in Section 2.2. For those criteria, we generate a dispersion parameter  $\theta_4$  to  $\theta_8$  using an uniform law in the interval  $[0.8; 1.5]$ . Among those 5 criteria, 2 are simulated using the primary criteria  $s1$  as intermediate modal permutation and  $\theta_4$  and  $\theta_5$  as dispersion parameters. We call them  $s1c4$  and  $s1c5$ . The 3 other criteria are simulated using the primary criteria  $s2$  as intermediate modal permutation and  $\theta_6$ ,  $\theta_7$  and  $\theta_8$  as dispersion parameters. We call them  $s2c6$ ,  $s2c7$  and  $s2c8$ . Then, we simulate the 3 last secondary criteria. Those are a combination of the other criteria. We then have  $s12$  that is the permutation resulting of the sum of the latent variables of  $s1$  and  $s2$ ,  $s2_4$  that is the permutation resulting of the difference of the latent variables of  $s2$  and  $s1c4$  and  $s2_4a$  which is the permutation resulting of the absolute value of the latent variable of  $s2_4$ .

The latent variables of each of the primary and chained Mallows' criteria are simulated using a lognormal distribution of parameters generated uniformly in  $[0.5; 3.5]$  for the mean parameter and  $[0.5; 1.5]$  for the standard deviation parameter.

We generate 500 sets of parameters. For each set, we generate a total of  $T = 500$  replicates of 5 items.

As we have seen in Section 2.2, the sequence of Mallows' models tends to over disperse the permutations in  $\mathcal{S}_n$ . That is why, we have chosen to use quite big values for the dispersion parameters of the chained Mallows' criteria.

In this Section, we evaluate the different possibilities introduced in Section 4. To this end, we compare the results when we assume every criteria is independent from each other with the results we found when we apply our corrections. We also compare them to the results we found when we use only the primary criteria.

As a comparison between the different models, we use the Kendall distance between  $\hat{\pi}_t$  the estimation of the modal permutation and  $\pi_t$  the real value of this modal permutation with  $t \in \{1, \dots, T\}$ .

Those models only estimate the values of the dispersion parameters. However, we

want to estimate their value on the localization parameter. To this end, we consider the estimators

$$\hat{\pi}_t = \operatorname{argmax}_{\pi_t \in \mathcal{S}_n} \sum_{j=1}^J -\theta_j d(\sigma_{jt}, \pi_t) \quad (30)$$

where  $\theta_j$  is replaced by the estimator  $\hat{\theta}_j$ ,  $\tilde{\theta}_j$  or  $\check{\theta}_j$  and  $\mathcal{S}_n$  is the set of all permutations of length  $n$  and  $J$  is the number of criteria.

For the model with the redundancy coefficients, we use 360 of the parameters sets to learn the function  $\kappa$  of equation (26). This function is estimated using a random forest. Then, for the 140 remaining parameters sets, we use  $T_{Tr} = 300$  replicates of 5 items to learn the independent values of  $\theta_j$  for  $j \in \{1, \dots, J\}$ . Finally, we test on the  $T_{Te} = 200$  remaining replicates of 5 items the differences between  $\hat{\pi}_t$  and  $\pi_t$ .

We evaluate the model with the direct estimation  $\check{\theta}$  of  $\theta$  on 360 parameters sets. For each parameter set, we consider  $T = 500$  replicates of 5 items. Among those sets, we use  $T_{Tr} = 300$  replicates for the training and  $T_{Te} = 200$  replicates for the tests.

To recapitulate, we represent in Figure 6 the links between the criteria in this simulation. In blue and first column, we have the primary criteria, in red and second column, we have the secondary criteria obtained by a chained Mallows' model and in purple and third and fourth columns, we have the secondary criteria obtained as a combination of the latent variables of other criteria.

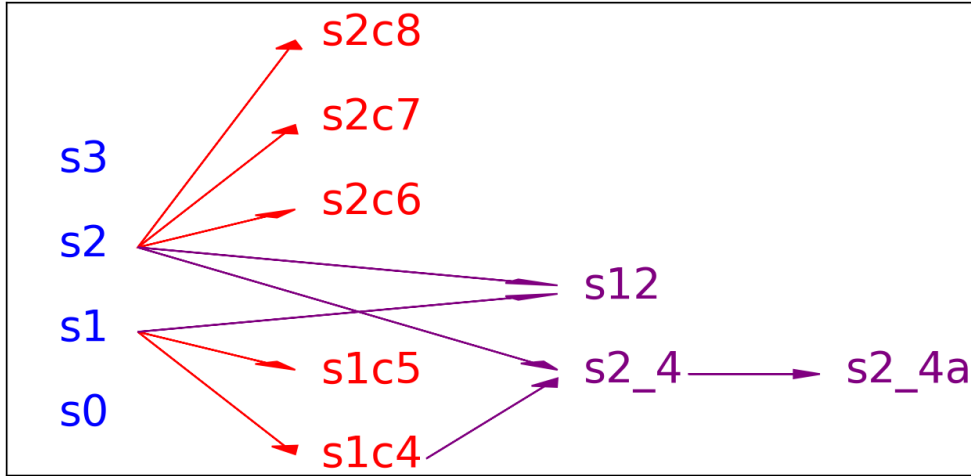


Figure 6: Hierarchy of criteria.

## 5.1 Comparison using the Proximity Indicators

In Figure 7, we represent the 3 possibilities to compare in terms of the Kendall distance between  $\hat{\pi}_t$  and  $\pi_t$  on the test set. To make the Figure more readable, we have sorted the

elements by an increasing Kendall distance in the first model, the case where we consider every criteria as independent. This case is represented in blue square. The case where we only use the 3 primary criteria is represented in green circle and the interest case, when we use our proximity indicators, is represented in orange diamond. We observe that even if the model using only the primary criteria (green circle) is clearly better than any other model, our model can be slightly better than the independent case.

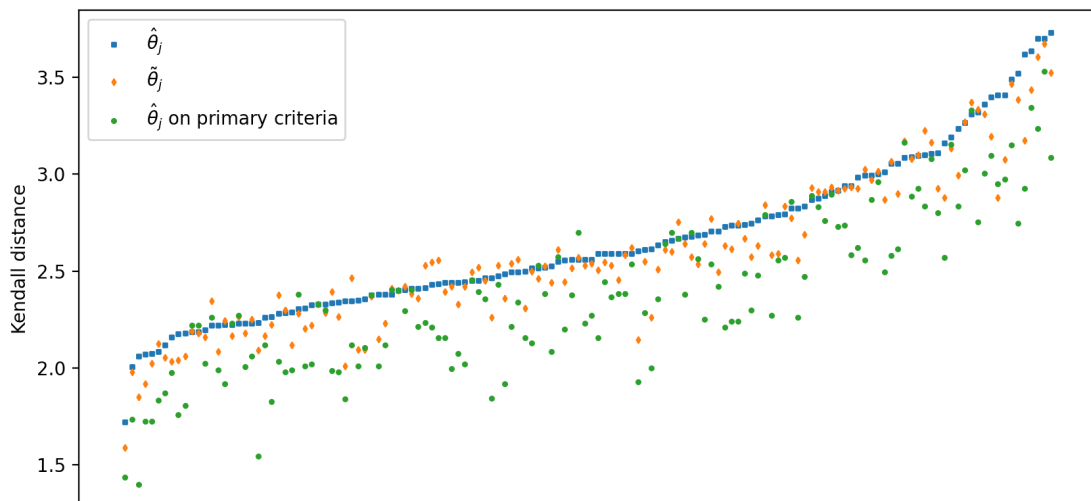


Figure 7: Comparisons of Kendall distances between  $\hat{\pi}_t$  and  $\pi_t$ .

Model	M1	M2	M3	M4	M5	M6
Mean	2.652	2.591	2.578	2.585	2.579	2.395
Variance	0.151	0.137	0.144	0.146	0.149	0.167
First quartile	2.364	2.335	2.319	2.321	2.315	2.093
Median	2.593	2.519	2.485	2.506	2.482	2.345
Third quartile	2.891	2.853	2.842	2.838	2.819	2.664
Minimum	1.780	1.716	1.634	1.722	1.560	1.402
Maximum	3.684	3.698	3.648	3.608	3.594	3.546

Table 3: Comparisons of Kendall distances between  $\hat{\pi}_t$  and  $\pi_t$ . In the model  $M1$  we consider every  $\theta_j$  as independent. In the model  $M2$  we consider every redundancy coefficient as 1. In the model  $M3$  we learn each redundancy coefficient as a function of the 130 Proximity indicators. In the model  $M4$  we set every redundancy coefficient randomly between 0 and 1. In the model  $M5$  we consider each  $\theta_j$  independent but we are limited to the primary and the secondary created with the chained Mallows' criteria. In the model  $M6$ , we consider each  $\theta_j$  independent but we are limited to the primary criteria.

In Table 3, we compare 6 different possibilities. Among them, the model  $M1$  using

all criteria and assuming the  $\theta_j$  independent is the worst while the model  $M6$  using only the primary criteria is clearly the best. Using the testing set, the Kruskal-Wallis test between the the models  $M1$ , assuming every  $\theta_j$  is independent and  $M3$ , our interest model, can not reject the null hypothesis, with a p-value of  $0.20 > 0.05$ , saying that the median are equal. We can't conclude that the model  $M4$  when we set each redundancy coefficient randomly is different than our interest model  $M3$  either as the p-value is  $0.98 > 0.05$ .

Using the training set however, we have a significant difference between  $M1$  and  $M3$  with a p-value of  $0.036 < 0.05$ . But we still cannot say that the model of interest  $M3$  is different to the random values  $M4$  with the p-value equal to  $0.83 > 0.05$ .

The method is interesting in the aim to find a general coefficient to apply to the parameters of dispersion independently computed. This coefficient can help to have a lower computing time. The computational time gain is due to the lower number of computations of the consensus. Its computation time is exponential with the number of items with the algorithm developed in [10] as it is done in this article, but linear considering Borda approach [2].

## 5.2 Comparison with a direct estimation of $\theta$

In this Section, we want to estimate the value of  $\check{\theta}_j$ , for  $j \in \{1, \dots, J\}$ , without using the estimation of  $\hat{\theta}_j$ .

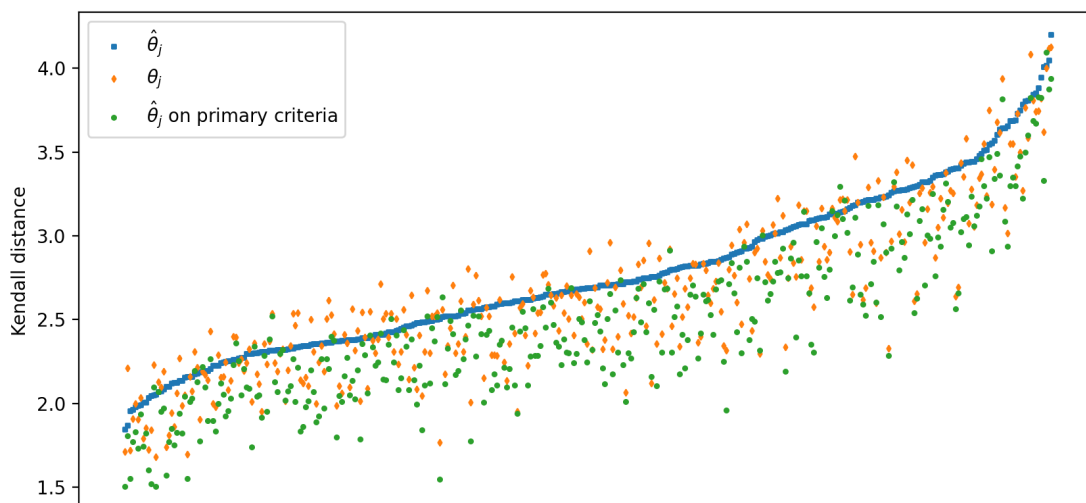


Figure 8: Comparisons of Kendall distances between  $\hat{\pi}_t$  and  $\pi_t$ .

In Figure 8, we compare the distance between  $\hat{\pi}_t$  and  $\pi_t$  for the model with every criteria assumed independent in blue square, the model with only the primary criteria in green circle and in orange diamond our new estimations  $\check{\theta}_j$  of  $\theta_j$  optimized with the equation given in (29).

In Figure 8, we notice that in the test, even though the estimation  $\check{\theta}_j$  of  $\theta_j$  looks better than the original model with every criteria assumed independent, it may be worse for some cases, which is not the case of the model containing only the primary criteria.

Model	M1	M7	M8	M5	M6
Mean	2.766	2.676	2.639	2.693	2.494
Variance	0.221	0.237	0.233	0.220	0.236
First quartile	2.384	2.325	2.310	2.345	2.140
Median	2.690	2.600	2.573	2.610	2.413
Third quartile	3.106	3.006	2.928	3.016	2.760
Minimum	1.845	1.640	1.680	1.645	1.505
Maximum	4.200	4.135	4.125	4.015	4.095

Table 4: Comparisons of Kendall distances between  $\hat{\pi}_t$  and  $\pi_t$ . In the model *M1* we consider every  $\theta_j$  as independent. In the model *M7* we consider an optimization of the  $\theta_j$  with as initialization, the  $\theta_j$  obtained with the model *M1*. In the model *M8* we consider an optimization of the  $\theta_j$  with as initialization, a random value between 0 and 1. In the model *M5* we consider each  $\theta_j$  independent but we are limited to the primary and the secondary created with the chained Mallows' criteria. In the model *M6*, we consider each  $\theta_j$  independent but we are limited to the primary criteria.

In Table 4, we compare the 2 new models *M7* and *M8* and 3 previous models described in Section 5.1, especially in Table 3. Here, we explore the possibility to strengthen the learning of the dispersion parameters  $\theta_j$  by learning all of them directly. We especially consider 2 possible initializations of the learning algorithm. First, in model *M7*, we initialize with the values of the parameters obtained as if all the criteria were independent  $\theta_j$ . Then, in model *M8*, we launch 10 random initializations and keep the best one during the training.

If we compare the results globally on the testing set, with Table 4 we notice that both models *M7* and *M8* are significantly better in terms of the median than the initial model assuming every criteria independent (*M1*). We have for the model *M7* a p-value for the Kruskal-Wallis test at  $0.01 < 0.05$  and for the model *M8*, a p-value of  $3.20 \times 10^{-4} < 0.05$ . We conclude that they are both better than the model *M1*. The model *M8* looks better than the model *M7*, however the p-value of  $0.30 > 0.05$  does not let us conclude to a significant difference. Considering the model *M5* in which we only use the primary criteria and the secondary criteria obtained using a chained Mallows' model, we cannot say that the models *M7* nor *M8* are significantly better due to a p-value of respectively  $0.61 > 0.05$  and  $0.12 > 0.05$ . In the training set however, we can say that models *M7* and *M8* are significantly better than model *M5* with respectively a p-value of  $0.01 < 0.05$  and  $3.45 \times 10^{-5} < 0.05$ . The model *M6*, using only the primary criteria is better than any other.

The Figure 9 shows the values of the  $\theta_j$  used to simulate the primary criteria. In blue square, we have  $\theta_1$  used to simulate the criterion of permutation *s1*. This criterion is directly associated to the criteria *s1c4*, *s1c5* and *s12*. It is also included in the criteria

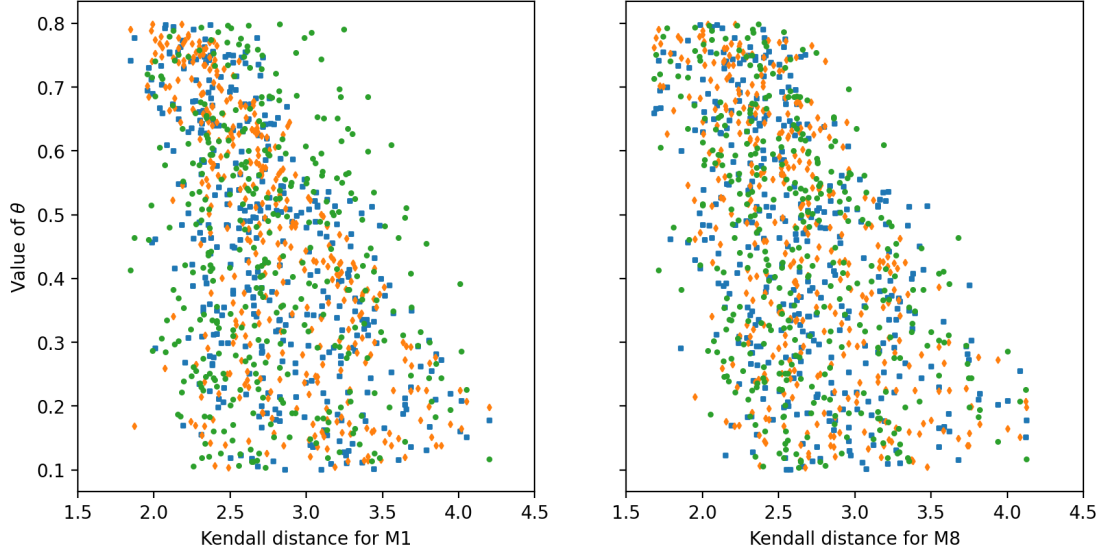


Figure 9: Values of  $\theta$  for each of the primary criteria by Kendall distances between  $\hat{\pi}_t$  and  $\pi_t$  for the model  $M1$ , assuming every criteria independent, and the model  $M8$ ,  $\theta$  optimized by (29). In blue square,  $\theta_1$  applied to the primary criteria  $s1$ , in orange diamond,  $\theta_2$  applied to the primary criteria  $s2$  and in green circle,  $\theta_3$  applied to the primary criteria  $s3$ .

$s2.4$  and  $s2.4a$  via the criterion  $s1c4$ . In orange diamond, we have the value of  $\theta_2$ , the parameter used to simulate the criteria of permutation  $s2$ . This criterion is associated directly to the criteria  $s2c6$ ,  $s2c7$ ,  $s2c8$ ,  $s12$ ,  $s2.4$  and  $s2.4a$ . Finally, we have in green circle the value of  $\theta_3$ , used to simulate the criterion of permutation  $s3$ .

This Figure 9, reflects the fact that the criterion independent  $s3$  does not participate enough in the choice of  $\hat{\pi}_t$  when we assume every criteria is independent. This is particularly clear on the left figure using the model  $M1$ . In this figure, we observe green circle points associated to a high value of  $\theta_3$  also associated to a high distance between  $\hat{\pi}_t$  and  $\pi_t$ . We do not experience this with  $\theta_1$  and  $\theta_2$  because they are over represented due to their secondary criteria associated. In the right graphic, we observe that with our function to estimate the better values  $\check{\theta}_j$  of  $\theta_j$ , none of the parameters associated to the primary criteria are under or over represented.

A high value of  $\theta_j$  in the Mallows' model means that the simulated permutations are closer to the modal permutation  $\pi_t$ . Then a high value is more reliable than 10 times a low value. For example in the model  $M1$  if we imagine  $\theta_2 = 0.25$  and the dispersion parameters of the chained Mallows associated with  $s2$  at 3.5, the maximum value, then we can easily imagine to have 4 times the permutation  $s2$ . In this case, we will have 4 times the same permutation associated with a  $\theta_j$  around 0.25. In the estimation of  $\hat{\pi}_t$ , we will have the cost associated to the distance to the permutation  $s2_t$  equal to  $0.25 \times 4 \times d_K(s2_t, \hat{\pi}_t)$ . At the same time, we have  $\theta_3 = 0.8$ . As this criterion does not

have any associated criterion and is totally independent to the others, we have as cost of divergence between  $\hat{\pi}_t$  and  $s3_t$ ,  $0.8 \times d_K(s3_t, \hat{\pi}_t)$ . Finally, the permutation  $s2_t$  will weight more on the estimation of  $\hat{\pi}_t$  than the permutation  $s3_t$ . This is an issue but is reduced thanks to the learning of all  $\theta_j$  simultaneously with the cost function we introduced.

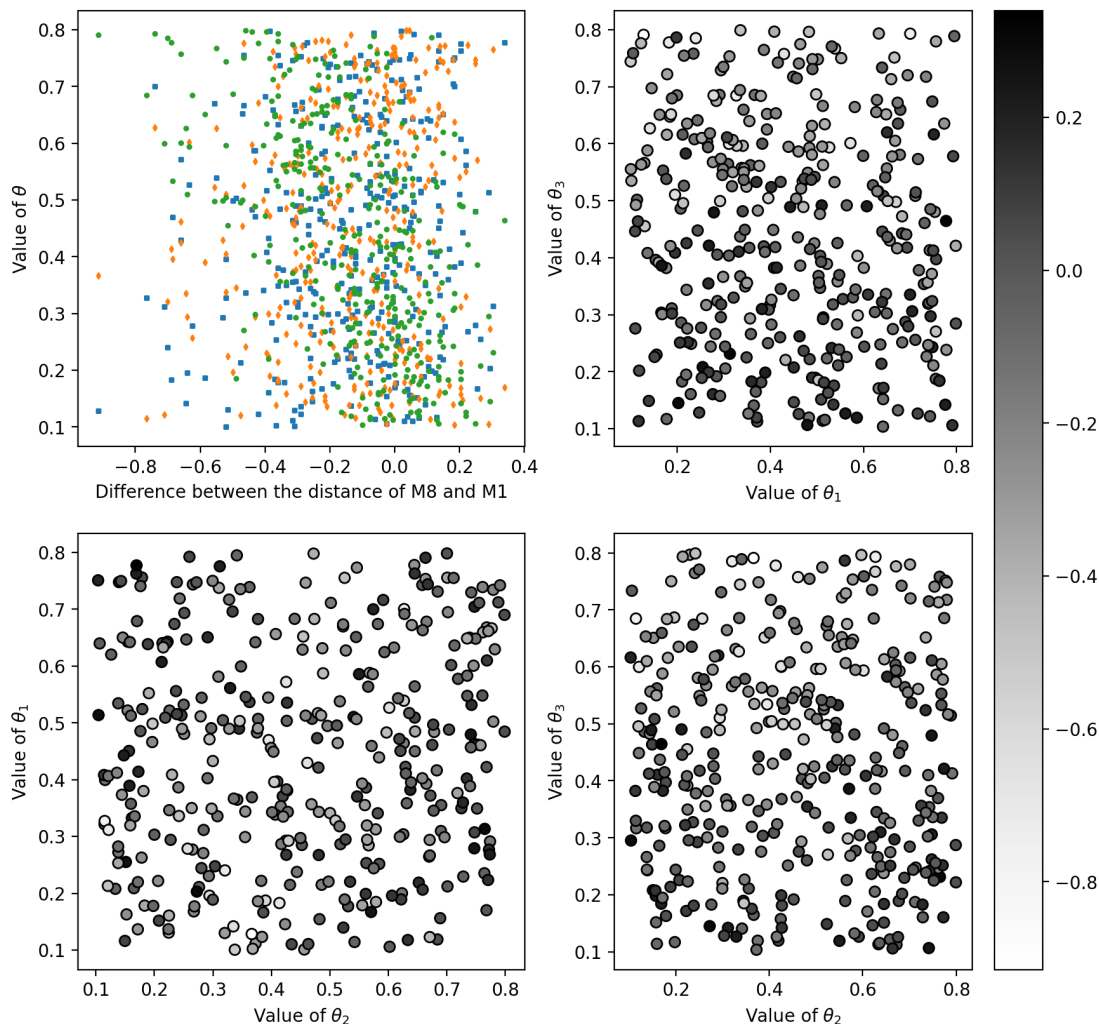


Figure 10: Differences between the Kendall distances between  $\hat{\pi}_t$  and  $\pi_t$  for the new estimation of the  $\theta_j$  and the case we assume all of the  $\theta_j$  independent.

Finally, with the Figure 10, we compare the difference of Kendall distance between  $\hat{\pi}_t$  and  $\pi_t$  for the model *M8* in which we applied our estimation of the dispersion parameters  $\theta_j$  and the model *M1* in which we assume each criteria is independent. Then, a negative difference means that *M8* performs better and a positive difference means that the model *M1* performs better. In the top left quarter of Figure 10, we show the values of the dispersion parameters of the 3 primary criteria by the value of the difference.

The blue square dots represent the parameter  $\theta_1$  associated to  $s_1$ , the orange diamond dots represent the parameter  $\theta_2$ , associated to  $s_2$  and the green circle dots represent the parameter  $\theta_3$  associated to  $s_3$ . The left part of the graph shows the sets of parameters where the new model  $M_8$  performs highly better. This side is notably composed of high values of  $\theta_3$  and low or intermediate values of  $\theta_1$  and  $\theta_2$ . At the same time, we see on the right part of the graphic the sets of parameters where the new model performs worse. This side is notably composed of low values of  $\theta_3$  and higher values of  $\theta_1$  and  $\theta_2$ .

Then, still in Figure 10, we have the 3 other graphics that shows the interactions between the values of the dispersion parameters of the primary criteria. If the interaction between  $\theta_1$  and  $\theta_2$ , in the bottom left quarter of the Figure 10, does not lead to any information, we remark in the top right quarter, with the interaction between  $\theta_1$  and  $\theta_3$ , and in the bottom right quarter, with the interaction between  $\theta_2$  and  $\theta_3$ , that high values of  $\theta_3$  leads to a better benefice of the model  $M_8$  and it's even more noticeable when the values of  $\theta_1$  and  $\theta_2$  are low. For those 3 sub-figures, the black value is quite bad and the best for us is the white which correspond to a great improvement in the distance between  $\hat{\pi}_t$  and  $\pi_t$ .

Coefficient	Estimate	Std. Error	p-value
Intercept	0.3596	0.1440	0.0130
$\theta_1$	-0.5224	0.2941	0.0766
$\theta_2$	-0.2093	0.2874	0.4671
$\theta_3$	-1.9905	0.3044	$2.17 \times 10^{-10}$
$\theta_1 : \theta_2$	0.5301	0.5697	0.3528
$\theta_1 : \theta_3$	2.9714	0.6230	$2.71 \times 10^{-6}$
$\theta_2 : \theta_3$	2.4150	0.6037	$7.72 \times 10^{-5}$
$\theta_1 : \theta_2 : \theta_3$	-5.0436	1.2064	$3.67 \times 10^{-5}$

Table 5: Results of the linear regression to predict the difference between the models  $M_8$  and  $M_1$ .

After running a linear regression to predict the difference with the values of those 3  $\theta_j$ , we obtain the results seen in Table 5. With this linear regression, we can confirm that the impact on distance between  $\hat{\pi}_t$  and  $\pi_t$  of our new method to evaluate the parameters  $\theta_j$  in the dependent case and particularly in this simulation is strongly impacted by the value of  $\theta_3$  and its interaction with the smallest possible values of  $\theta_1$  and  $\theta_2$ .

With this simulation, we are not able to find a configuration of parameters where we clearly see that our models of interest with the proximity indicators nor with the direct estimation of the  $\theta_j$  is better than the estimation using only the primary criteria. However, the direct estimation of  $\theta_j$  is the better to approximate the best model for this simulation.



## 6 Concluding remarks

We can ask ourselves that if the primary criteria alone can lead to better results, why do we try to compensate for the dependence ? The first point is that those proximity indicators can help to understand how the different secondary criteria are correlated between them. The second point is about the meaning of the secondary indicators. Their definitions can help to understand what's important in the consensus. At the end, we do not want a black box. We want to be able to say why an item is above another in the consensus. Then we also have the problematic with the chained Mallows that we do not know which criterion is the primary. For example, in the case of the revenues of a set of items, we can understand that the criterion is correlated to the criterion of the expected revenue of this set of items, but it is difficult to know which one is the primary.

In this paper, we were particularly interested in the degradation of the results due to secondary criteria only derived from the primary criteria. However, we do not try to understand what happens when those secondary criteria also include information. This may lead to gains against the model based on the primary criteria only.

Another point of future investigation concern the generalized Mallows' models. Here, we were limited by the Mallows' models but its generalized version may have some links to our proximity indicator and let the models based on them have some better results. We may also try to change the model used to predict the redundancy coefficients.

In [10], we used an approximation and the generalized Mallows' model to estimate the dispersion parameter. A future work could be to apply the direct estimation of  $\theta$  to estimate this parameter without the approximation.

Finally, we can also compare the parametric Mallows' model with non parametric modeling approaches by simulations studies.

## References

- [1] Jörg Arndt. *Matters Computational: ideas, algorithms, source code*. Springer Science & Business Media, 2010.
- [2] J-C de Borda. Mémoire sur les élections au scrutin: Histoire de l'académie royale des sciences. *Paris, France*, 12, 1781.
- [3] Jean-Pierre Brans and Yves De Smet. Promethee methods. In *Multiple criteria decision analysis*, pages 187–219. Springer, 2016.
- [4] Jean-Pierre Brans, R Nadeau, and M Landry. L'ingénierie de la décision. *Elaboration d'instruments d'aide à la décision. La méthode PROMETHEE*. In *l'Aide à la Décision: Nature, Instruments et Perspectives d'Avenir*, pages 183–213, 1982.
- [5] Gustave Choquet. Theory of capacities. In *Annales de l'institut Fourier*, volume 5, pages 131–295, 1954.

- [6] Persi Diaconis. Group representations in probability and statistics. *Lecture Notes-Monograph Series*, 11:i–192, 1988.
- [7] James S Dyer, Peter C Fishburn, Ralph E Steuer, Jyrki Wallenius, and Stanley Zionts. Multiple criteria decision making, multiattribute utility theory: the next ten years. *Management science*, 38(5):645–654, 1992.
- [8] José Rui Figueira, Vincent Mousseau, and Bernard Roy. Electre methods. In *Multiple criteria decision analysis*, pages 155–185. Springer, 2016.
- [9] Michael A Fligner and Joseph S Verducci. Distance based ranking models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(3):359–369, 1986.
- [10] Arthur Fétiveau, Gilles Durrieu, Emmanuel Frénod, Claude-Henri Meledo, and Benoît Prat. Permutations based model for business performance. *Discrete and Continuous Dynamical Systems - S*, 2024.
- [11] Michel Grabisch. L’utilisation de l’intégrale de choquet en aide multicritère à la décision. *Newsletter of the European Working Group” Multicriteria Aid for Decisions*, 3(14):5–10, 2006.
- [12] Ekhine Irurozki, Borja Calvo, and Jose A. Lozano. Permallows: An r package for mallows and generalized mallows models. *Journal of Statistical Software*, 71(12):1–30, 2016.
- [13] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.
- [14] Christophe Labreuche and Simon Fossier. Explaining multi-criteria decision aiding models with an extended shapley value. In *IJCAI*, pages 331–339, 2018.
- [15] C. L. Mallows. Non-null ranking models. i. *Biometrika*, 44(1/2):114–130, 1957.
- [16] Bertrand Mareschal and Jean-Pierre Brans. Bank adviser: un système interactif multicritère pour l’évaluation financière des entreprises à l’aide des méthodes promethee. *L’Actualité économique*, 69(1):191–205, 1993.
- [17] Tao Qin, Xiubo Geng, and Tie-Yan Liu. A new probabilistic model for rank aggregation. *Advances in neural information processing systems*, 23, 2010.
- [18] Bernard Roy. Classement et choix en présence de points de vue multiples. *Revue française d’informatique et de recherche opérationnelle*, 2(8):57–75, 1968.
- [19] Jyrki Wallenius, James S Dyer, Peter C Fishburn, Ralph E Steuer, Stanley Zionts, and Kalyanmoy Deb. Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management science*, 54(7):1336–1349, 2008.