Université Libre de Bruxelles

CoDE - SMG

# Approximating PROMETHEE II's net flow scores by piecewise linear value functions

## CoDE-SMG – Technical Report Series

Stefan Eppe, Yves De Smet

**CoDE-SMG – Technical Report Series**

Technical Report No.
TR/SMG/2012-004

October 2012

**CoDE-SMG – Technical Report Series**
ISSN 2030-6296

Published by:
CoDE-SMG, CP 210/01
Université Libre de Bruxelles
Bvd du Triomphe
1050 Ixelles, Belgium

Technical report number TR/SMG/2012-004

# Approximating PROMETHEE II's net flow scores by piecewise linear value functions

## CoDE-SMG – Technical Report Series

Stefan Eppe stefan.eppe@ulb.ac.be
Yves De Smet yves.de.smet@ulb.ac.be
CoDE-SMG, Université Libre de Bruxelles, Brussels, Belgium

October 2012

# Approximating PROMETHEE II's net flow scores by piecewise linear value functions

Stefan Eppe* and Yves De Smet

Technical Report -- October 2012

**Abstract**

PROMETHEE II is a prominent method for multicriteria decision aid (MCDA) that builds a complete ranking on a set of potential actions by assigning each of them a so-called *net flow* score. However, to calculate these scores, each pair of actions has to be compared, causing the computational load to increase quadratically with the number of actions, eventually leading to prohibitive execution times for large decision problems. For some problems, however, a trade-off between the ranking's accuracy and the required evaluation time may be acceptable. Therefore, we propose a piecewise linear model that approximates PROMETHEE II's net flow scores and reduces the computational complexity (with respect to the number of actions) from quadratic to linear at the cost of some misranked actions. Simulations on artificial problem instances allow us to quantify this time/quality trade-off and to provide probabilistic bounds on the problem size above which our model satisfyingly approximates PROMETHEE II's rankings. They show, for instance, that for decision problems as small as 10 actions evaluated on 3 criteria, our model ranks 9 actions accurately with a probability of 90%. Beyond its immediate applicability on large decision problems, our model also provides some insight into how (far) PROMETHEE II's outranking method is different from a much simpler weighted sum.

## 1 Introduction

Methods for multicriteria decision aid (MCDA) are grossly divided into two main families (Roy, 2005; Siskos et al., 1984): multiattribute utility theory (MAUT) related methods and outranking-based methods. Members of the first family can usually be formalized as an aggregation of unicriterion utility or value functions that provide a (binary) weak preference relation, resulting in a complete and transitive ranking over a set of potential actions (Dyer, 2005).

Outranking methods, on the other hand, are based on the pairwise comparison of actions. A sub-set of these methods – of which ELECTRE (Figueira et al., 2005b) and PROMETHEE II (Brans and Mareschal, 2005) are among the most prominent ones – also provide a complete and transitive ranking. However, despite the many applications reported in the literature (Behzadian et al., 2010; Figueira et al., 2005b), do outranking methods suffer from a certain lack of scalability. This issue arises particularly in situations where large data sets must be handled fast, e.g., for geographical information analysis (Marinoni, 2006) or preference querying (Pivert and Smits, 2012), because providing the decision maker with an outranking-based evaluation on $n$ actions comes at a cost of $O(n^2)$ pairwise action comparisons. The computation time increases accordingly, eventually reaching the limit of evaluation time that is considered as acceptable. But then, what would be the maximum size of a decision problem that could be handled by a given outranking method? Although answering this question would, of course, be of great practical interest to the practioner, this is not possible in a general way because of the many subjective and problem specific factors that intervene (time required to compute an evaluation, quality of implementation, time granted for the overall evaluation, etc.).

Based on these considerations, we have chosen to address the above mentioned limitations of outranking methods from another point of view. The approach we propose in this paper has been triggered by simulations on the PROMETHEE II model. For this method, the overall score of each action can be formulated as a

*Corresponding author at Computer & Decision Engineering department (CoDE), Polytechnic School of Brussels, Université Libre de Bruxelles, Belgium. E-mail: stefan.eppe@ulb.ac.be

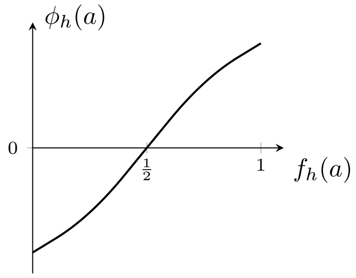



Figure 1: Observed general shape of the unicriterion net flow score $\phi_h(a)$ for a maximisation problem.

weighted sum of unicriterion scores, and our observations suggest that, for sufficiently large problem instances, these unicriterion action scores behave as MAUT-like value functions. As such value functions generally have a linear complexity $O(n)$ with respect to the problem size, we wanted to investigate the possibility of approximating an action's unicriterion PROMETHEE II score by such a function. Indeed, it seemed reasonable to assume that, at least for some types of applications, a trade-off between the evaluation speed and the accuracy of the resulting ranking would be acceptable. Our goal was thus to develop a value function with linear time complexity that would only depend on an action's evaluations and would lead to the best possible approximation of the ranking that we would obtain with PROMETHEE II (labeled `P2` in the following).

When plotting the unicriterion score of an action with respect to its rank (named "rank-score plot" in the sequel), we obtain a monotonically increasing, slightly $s$-shaped curve that crosses the $x$-axis approximatively at abscissa 0.5 (Figure 1). Its general shape suggests starting with a third order polynomial regression model (`P3R`) to build an approximation function of `P2`'s unicriterion scores. Comparing the produced `P3R`-ranking with `P2`'s reference ranking, we indeed observe a high rank concordance (properly defined in Section 3), even for instances that count as few as 10 actions. However, as this model is computed *ex post*, i.e., on the basis of a complete `P2` evaluation, it cannot be used for our purpose of reducing the time complexity; at best can it serve as a quality benchmark the *ex ante* approximation models we want to build should be compared with.

Since we also observed that the rank concordance between `P2` and `P3R` rankings increases with the size of the decision problem, we have developed two *ex ante* approximation value functions based on the continuous extension of `P2`'s discrete formulation, considering an infinite number of actions: the first model is a piecewise *polynomial* one (`PPA`), while the second one provides a piecewise *linear* approximation function (`PLA`).

We have extensively tested both *ex ante* models through an experimental study (Section 4) that aims at validating our approach, as well as providing an estimation of the rank concordance that could at least be reached by each approximation model. Results for different problem sizes are presented and discussed in Section 5 and, although they do not directly answer our initial question about a general "PROMETHEE II applicability limit", they do provide very related, quantitative information about: 1) how to approximate the original model when it becomes computationnally too expensive; and 2) what trade-off has to be taken into account when speeding up the evaluation process through a quadratic to linear complexity reduction.

To conclude this introduction, we would like to stress that the potentially polemical statement, that PROMETHEE II's outranking method could be approximated by MAUT-like value functions even for relatively small decision problems, should not be taken by the reader as a judgement of value, leading to prefer one method over the other. Indeed, selecting the most adapted method to tackle a given decision problem remains a complex topic (Guitouni and Martel, 1998) that we do not address here. Our more modest aim is to contribute to a better understanding of the PROMETHEE II outranking method through a formal and quantitative comparison with the other big family of MCDA methods.

## 2 The Promethee II method

We start by providing a brief description of the PROMETHEE II outranking method. For a more detailed introduction the interested reader may refer to Brans and Mareschal (2005).

Let $A = \{a_1, \dots, a_n\}$ be a set of $n$ actions. Each action $a_i$, with $i \in I = \{1, \dots, n\}$, is characterised by means of $m$ evaluations $f_h(a_i)$, $\forall h \in H = \{1, \dots, m\}$. To compare any pair of actions $(a, b) \in A \times A$, a

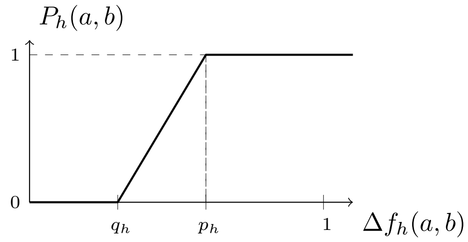



Figure 2: PROMETHEE's "V-shaped" preference function on criterion $h$, for an action pair $(a,b) \in A \times A$. $\Delta f_h(a,b) = f_h(a) - f_h(b)$ in the case of a maximisation problem. For each criterion, an indifference threshold $q_h$ and a preference threshold $p_h$ have to be provided by the decision maker.

so-called preference function $P_h(a,b)$ is introduced for each criterion $h$ to express the preference degree, on that criterion, of one action over the other. In this paper, we consider the widely used "*V-shaped*" preference function, with indifference and preference thresholds that are respectively denoted by $q_h$ and $p_h$ (Figure 2):

$$P_h(a,b) = \begin{cases} 0 & \text{, if } \Delta f_h(a,b) \leq q_h \\ \frac{\Delta f_h(a,b) - q_h}{p_h - q_h} & \text{, if } q_h < \Delta f_h(a,b) \leq p_h \\ 1 & \text{, if } p_h < \Delta f_h(a,b) \end{cases}$$

where $\Delta f_h(a,b) = f_h(a) - f_h(b)$. The pairwise action comparisons are aggregated for each action and provide the unicriterion net flow score $\phi_h(a)$ on criterion $h$:

$$\phi_h(a) = \tfrac{1}{n-1} \sum_{b \in A} [\ P_h(a,b) - P_h(b,a)\ ]\,.$$

Finally, the unicriterion scores are aggregated over all criteria through a weighted sum to yield that action's net flow score:

$$\phi(a) = \sum_{h \in H} w_h \phi_h(a), \tag{1}$$

where $w = \{w_1, \ldots, w_m\}$ is a vector of the criteria's relative importance, with $w_h \geq 0, \forall h \in H$, and $\sum_{h \in H} w_h = 1$.

Without loss of generality, we will consider a maximisation problem and assume that evaluations lie in the interval $f_h(a_i) \in [0,1]$, $\forall (i,h) \in I \times H$.

# 3 A value function to approximate the unicriterion score

Our goal, in this paper, is to determine an approximation of an action's net flow score $\phi(a)$ that acts like a value function, only depending on its evaluations $f_h(a)$, $\forall h \in H$, and on the preference parameters. It should not depend on any pairwise action comparison. With this aim in mind, and given the mathematical form of (1), we first focus on the unicriterion terms of the weighted sum. Indeed, if we manage to determine an *ex ante* approximation of each unicriterion score $\phi_h(a)$, the global approximation would immediately be determined by (1), boiling down to a simple weighted averaging approach.

For the sake of simplicity, we will assume, in this section, that the actions are sorted in increasing order of their evaluation for the considered criterion: $f_h(a_i) \leq f_h(a_j)$, $\forall i < j$. We can thus rewrite the unicriterion net flow score of action $a_i$ as

$$\phi_h(a_i) = \tfrac{1}{n-1} \sum_{j=1}^{i-1} P(a_i, a_j) \ - \ \tfrac{1}{n-1} \sum_{j=i+1}^{n} P(a_j, a_i)$$

In this form, however, the unicriterion score cannot be handled easily. As we want to consider large action sets, we choose to extend the formulation above to the case of an infinite set of actions. Exploring the

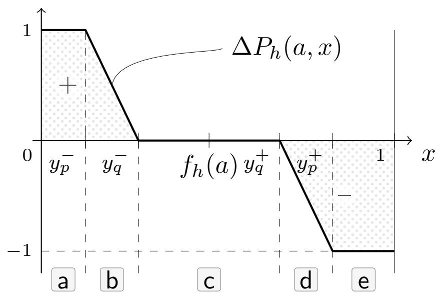



Figure 3: Integration domains of $\phi_h(a)$.

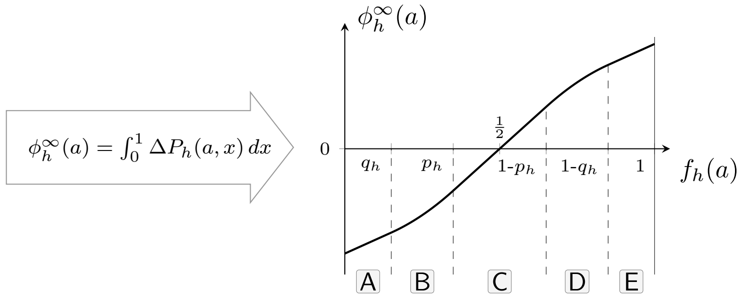



Figure 4: Shape of the unicriterion net flow score.

possible meanings of this continuous extension lays beyond the scope of this work; we will only consider it as a mathematical mean that could provide us some insight into the asymptotic behaviour of unicriterion net flow scores. Let us mention that the extension to an infinite set of actions has already been proposed as the PROMETHEE IV method (Brans et al., 1986), but, to the best of our knowledge, was never further developed nor applied. Recently, a continuous extension of PROMETHEE II has also been mentioned (De Smet et al., 2009) in the context of combinatorial multi-objective optimization problems (where the number of considered solutions is high), but not actually used.

Still assuming that the actions are sorted in ascending order of their evaluations on criterion $h$, we choose to identify each action through a real number $a \in [0,1]$: $a = 0$ is the worst ranked action, while $a = 1$ is the best ranked one. If we further consider that the continuous distribution of actions along the evaluation axis $x$ is given by the density function $\rho_h(x)$ (chosen such that $\int_0^1 \rho_h(x)dx = 1$), the evaluation of each action $a$ is given by $f_h(a) = \int_0^a \rho_h(x)dx$. The unicretrion net flow can thus be written as follows in the continuous case:

$$\phi_h^\infty(a) = \int_0^a P_h(a,x)\rho_h(x)dx - \int_a^1 P_h(x,a)\rho_h(x)dx \tag{2}$$

In the sequel of this section, we will consider the particular case of a uniform distribution of actions: $\rho_h(x) = 1$, and hence, $f_h(a) = a$. This will lead us to replce $f_h(a)$ by $a$ in the subsequent equations.

Introducing the following help variables (that depend on $a$):

$$\begin{cases} y_q^- &= \max\{\,0\,;\ a - q_h\,\} \\ y_p^- &= \max\{\,0\,;\ a - p_h\,\} \\ y_q^+ &= \min\{\,1\,;\ a + q_h\,\} \\ y_p^+ &= \min\{\,1\,;\ a + p_h\,\} \end{cases}$$

the integration range of (2) can be sliced into five segments, denoted by $\boxed{\mathsf{a}}$,. . .,$\boxed{\mathsf{e}}$ (Figure 3).

Once integrated, we obtain the following formulation:

$$\phi_h^\infty(a) = \underbrace{y_p^-}_{\boxed{\mathsf{a}}} + \underbrace{\tfrac{(y_q^- - y_p^-)^2}{2(p_h - q_h)}}_{\boxed{\mathsf{b}}} + \underbrace{0}_{\boxed{\mathsf{c}}} - \underbrace{\tfrac{(y_p^+ - y_q^+)^2}{2(p_h - q_h)}}_{\boxed{\mathsf{d}}} - \underbrace{\left(1 - y_p^+\right)}_{\boxed{\mathsf{e}}} \tag{3}$$

The unicriterion net flow $\phi_h^\infty(a)$ is composed of three terms: $\boxed{\mathsf{a}}$, $\boxed{\mathsf{c}}$, and $\boxed{\mathsf{e}}$, of linearly increasing values, separated by two intervals: $\boxed{\mathsf{b}}$ and $\boxed{\mathsf{d}}$, with quadratic terms. The extent and contribution of each of them to the unicriterion net flow depends on both thresholds $q_h$ and $p_h$, and also on action's $a$ evaluation $f_h(a)$. In the general case, five different ranges can again be distinguished in $\phi_h^\infty$'s domain (Figure 4 and Table 1): $\boxed{\mathsf{A}}$ – $\boxed{\mathsf{E}}$. Some characteristics of $\phi_h^\infty(a)$'s formulation are worth noting:

Table 1: For each segment $\boxed{\text{A}}$ to $\boxed{\text{E}}$ of $\phi_h^\infty(a)$, the integration segments $\boxed{\text{a}}$ to $\boxed{\text{e}}$ contribute in a specific way: not at all (0), with a constant value (*cst*), or linearly/quadratically increasing ( ↗ , ⤴ ), resp. decreasing ( ↘ , ⤵ ).

| $\phi_h^\infty(a)$ segment | $a \in$ | Integration segment $\boxed{\text{a}}$ | $\boxed{\text{b}}$ | $\boxed{\text{c}}$ | $\boxed{\text{d}}$ | $\boxed{\text{e}}$ |
|---|---|---|---|---|---|---|
| $\boxed{\text{A}}$ | $[\,0, q_h\,]$ | 0 | 0 | 0 | *cst* | ↗ |
| $\boxed{\text{B}}$ | $[\,q_h, p_h\,]$ | 0 | ⤴ | 0 | *cst* | ↗ |
| $\boxed{\text{C}}$ | $[\,p_h, 1\text{-}p_h\,]$ | ↗ | *cst* | 0 | *cst* | ↗ |
| $\boxed{\text{D}}$ | $[\,1\text{-}p_h, 1\text{-}q_h\,]$ | ↗ | *cst* | 0 | ⤵ | 0 |
| $\boxed{\text{E}}$ | $[\,1\text{-}q_h, 1\,]$ | ↗ | *cst* | 0 | 0 | 0 |

- Inside segments $\boxed{\text{A}}$ and $\boxed{\text{E}}$, the slope is constant: $\frac{d\phi_h^\infty}{da} = 1$;
- Inside segment $\boxed{\text{C}}$ the slope is constant too: $\frac{d\phi_h^\infty}{da} = 2$;
- The extreme values of the function are: $\phi_h^\infty(0) = \lambda_h - 1$ and $\phi_h^\infty(1) = 1 - \lambda_h$. Thus, the value range of the function is $2(1 - \lambda_h)$;
- There is a central symmetry with respect to the coordinate $(\frac{1}{2}, 0)$.

Note that this general interpretation yields for the case where $p_h < \frac{1}{2}$. For higher values of $p_h$, the shape changes slightly, but can be determined analytically in a similar way. Due to its mathematical formulation, we rename the model as the *piecewise polynomial approximation* (`PPA`) model, and denote so in the following: $\phi_h^{\text{PPA}}(a) = \phi_h^\infty(a)$.

Depending on the values of $q_h$ and $p_h$, some ranges may be reduced to an empty range. For instance, if $q_h = p_h$, there are no quadratic terms and $\phi_h^{\text{PPA}}(a)$ reduces to a piecewise linear function. This form is particularly appealing for its simplicity and, as a further simplification, we approximate (3) by a piecewise linear function $\phi_h^{\text{PLA}}(a)$ composed of three segments and defined by one single threshold-related parameter $\lambda_h = \frac{1}{2}(q_h + p_h)$:

$$\phi_h^{\text{PLA}}(a) \quad = \quad \begin{cases} a + \lambda_h - 1 & \text{if } a < \lambda_h \\ 2a - 1 & \text{if } \lambda_h \leq a < 1 - \lambda_h \\ a - \lambda_h & \text{if } a > 1 - \lambda_h \end{cases} \tag{4}$$

This particular piecewise linear approximation (`PLA`) model is built in order have the same four functional features as those previously noted for $\phi_h^\infty(a)$. As an addition resulting feature, the linear segments of our `PLA` model intersect at symmetric coordinates $(\lambda_h, 2\lambda_h - 1)$ and $(1 - \lambda_h, 1 - 2\lambda_h)$.

Let us stress again that, in this formulation, we only use one single "threshold" parameter, $\lambda_h$, that is the mean value of $q_h$ and $p_h$. This remarkable property questions (at least for bigger problem instances) the usefulness of requiring two parameters to determine a preference function. It also tends to show that the effects of indifference and preference parameters on an action's ranking do compensate each other in some way. This could shed a new light on the difficulty of eliciting these parameters (Eppe et al., 2011): experiments that would only try to elicit the relative weight and parameter $\lambda_h$ for each criterion could be run to verify this conjecture.

Finally, having defined value functions that approximate the unicriterion net flow score of an action on each criterion, we aggregate them through a weighted sum, just like for the original method, and we obtain an approximation of each action's net flow score

$$\phi^{\text{PLA}}(a) = \sum_{h=1}^{m} w_h \phi_h^{\text{PLA}}(a), \tag{5}$$

**Algorithm 1**: Standard experimental process that outputs a result vector $\kappa$ of $N_{\text{trials}}$ runs.

**Input**: $n$, $m$, $N_{\text{trials}}$
**for** $i = 1 \ldots N_{\text{trials}}$ **do**
  $A =$ `randEvals`$(n, m)$;
  $(w, q, p) =$ `randPrefParams`$(m)$;
  $R =$ `computeNetFlowRanking`$(A, w, q, p)$;
  $R^{PLA} =$ `computePLARanking`$(A, w, q, p)$;
  $\kappa_i =$ `computeCRatio`$(R, R^{PL})$;

The approximated net flow scores induce an "approximated" ranking over $A$. We denote $R^{\text{PLA}}(a)$ the rank of action $a$ based on our approximated model, and hope it to be as close as possible to action $a$'s reference rank $R(a)$ obtained with the classical Promethee II method.

# 4 Experimental setup

From an artificial continuous formulation, we have deduced a piecewise linear approximation (PLA) that we hope to be applicable to finite action sets. We are now going to put our model to the test, comparing the rankings it generates with the reference ranking produced by the original Promethee II method. Beyond the mere validation of our model, our main aim is to provide an empirical bound on the instance size above which Promethee II's net flow scores are reasonably well approximated by our *ex ante* parametrized PLA-function.

The experimental approach proposed in this paper consists (Algorithm 1) in generating a random instance of $n$ actions over $m$ criteria, as well as preference parameters (weights and thresholds) for each criterion. Therefrom, the rankings of the generated set of actions following respectively Promethee II's original model ($R^{\text{P2}}$) and our piecewise linear model ($R^{\text{PLA}}$) are computed and compared. We use the resulting similarity measure to

1. validate the approach by showing that for reasonably sized instances, our PLA-model satisfyingly approximates the Promethee II ranking;

2. produce a table that provides an experimental numerical bound for the instance size, as from which the approximation quality reaches a required level.

To make things more concrete, we now provide some practical details about different aspects of the experimental setup:

**Quality measure** We define a *rank concordance ratio* $\kappa$, which is the ratio of the number of concordant action pairs, i.e., pairs that have the same relative rank order in both rankings, over the total number of pairwise action comparisons:

$$\kappa = \tfrac{1}{n(n-1)} \sum_{a,b \in A} c(a, b).$$

The concordance

$$c(a,b) = \begin{cases} 1 & \text{, if } [\phi(a) \geq \phi(b) \wedge \phi^{\text{PLA}}(a) > \phi^{\text{PLA}}(b)] \\ & \quad \vee [\phi(a) > \phi(b) \wedge \phi^{\text{PLA}}(a) \geq \phi^{\text{PLA}}(b)] \\ 0 & \text{, otherwise} \end{cases}$$

indicates whether or not a rank difference between a pair of actions following respectively rankings $R^{\text{P2}}$ and $R^{\text{PLA}}$ is concordant. Although this measure is closely related to Kendall's $\tau$ rank correlation coefficient (Kendall and Gibbons, 1990), we prefer the former because it allows taking possible ties into account in a "MCDA consistant" way.

**Randomly generated instances** We generate instances of $n$ actions, evaluated on $m$ criteria. For each generated instance, one type of distribution (Figure 5) is uniform randomly associated to each criterion.[1]

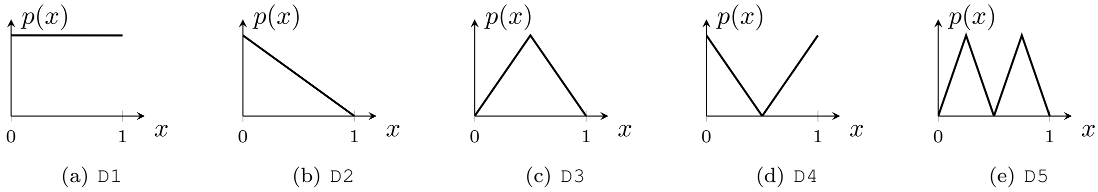



Figure 5: Shape of the probability density functions $p(x)$ used to generate random instances.

The evaluations on each criterion are then randomly generated for all actions following the corresponding, assigned distribution. By doing so, we aim at producing results that are not (too strongly) biased by the features of one specific distribution. Note that the PLA-model expressed by (3) assumes a uniform distribution. We will have to verify that mixed distributions do not affect the approximation's quality too much.

# 5 Results & discussion

Before delving into the empirical exploration of our model, we start this section by providing a first analysis of the statement (Section 3) that P2-rankings may depend on only one threshold-like parameter $\lambda_h$ per criterion. We then proceed with several qualitative and quantitative investigations to validate the PLA-model. Finally, we provide the results that we initially aimed for and that contribute to answering our question: "*As from what instance size is it possible to satisfyingly approximate an action's net flow score by our piecewise linear model?*".

## 5.1 Compensating effect of PROMETHEE II's threshold values

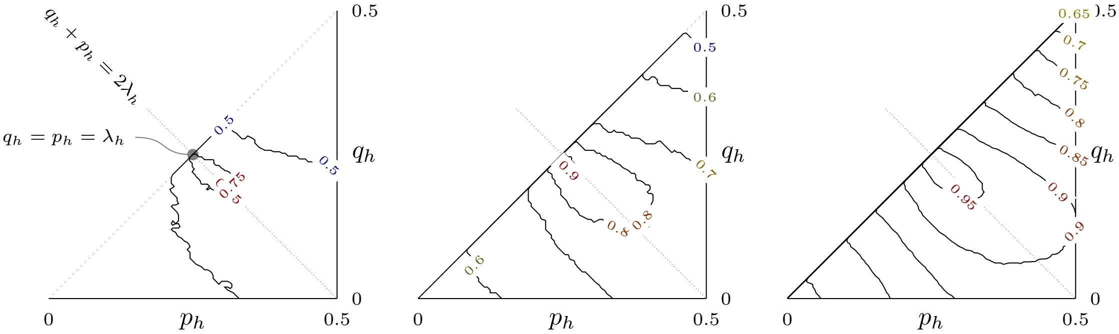
The piecewise linear approximation proposed in Section 3 only depends on one parameter, $\lambda_h = \frac{1}{2}(q_h + p_h)$. As already noted, this could suggest some sort of compensating effect between indifference and preference thresholds $q_h$ and $p_h$ in the PROMETHEE II preference model. To verify this, we observe, on the classical P2 formulation, how the ranking of an action set $A$ changes when the threshold values are altered (the weights remaining unchanged). Practically, we take the ranking $R_\lambda$, induced by the threshold values $q_h = p_h = \lambda$, as a reference and compare it with the ranking $R'$ induced by another pair of threshold values $(q'_h, p'_h)$. The comparison is done through the rank concordance ratio $\kappa(R_\lambda, R')$. In particular, we present results for the case where $\lambda_h = 0.25$. At each run: 1) a random set of actions is generated as described in Section 4, as well as a random weight vector; 2) the concordance $\kappa$ between $R_{\lambda=0.25}$ and $R'$ is computed, the latter being induced by threshold values $q_h$ and $p_h$, where $p_h \in [0, 0.5]$ and $q_h \in [0, p_h]$. We finally compute the 5% quantile of $\kappa(R_\lambda, R')$'s distribution for a series of 1000 runs, i.e., an approximation of the minimum concordance ratio reached with a probability of 95%.

The results (Figure 6) show, for all tested instance sizes, a symmetry with respect to the bisecting line $q_h + p_h = 2\lambda_h$. This tends to confirm the compensating role of $q_h$ and $p_h$: the influence of their average value $\lambda_h$ on PROMETHEE II's ranking is higher than their individual values. As the instance size increases, the isolines become more and more parallel to this bisecting line, thereby confirming our impression. On the other hand, however, the similarity of induced rankings with the reference ranking $R_\lambda$ descreases when the threshold pair tends to $(q_h, p_h) \rightarrow (0, 0.5)$. It is obviously the highest, i.e., $\kappa = 1$, for $q_h = p_h = \lambda$. The latter observation is particularly visible on smaller instances. The underlying reasons for this decrease should be investigated in a future work.

## 5.2 Empirical validation of our model

In the sequel of this section we provide a selection of the experimental results that highlight the observations made on a larger set of configurations (Table 2). Results corresponding to configurations that are not

[1]The attribution of a distribution function is independent for each criterion. Hence, the same distribution may be related to several criteria of the same instance.




(a) $n = 5$ actions (b) $n = 10$ actions (c) $n = 50$ actions

Figure 6: Isolines of 5%-quantile of rank concordance $\kappa$ measured between the reference ranking $R_\lambda$ corresponding to threshold values $q_h = p_h = \lambda_h$ and the ranking $R'$ induced threshold given by the coordinate $(q_h, p_h)$, with $p_h \in [0, 0.5]$ and $q_h \in [0, p_h]$. The statistics are computed on 1000 randomly generated instances (with "mixed distributions") of respectively 5, 10, and 50 actions, evaluated on 3 criteria. The thresold values are the same for all criteria.

Table 2: Parameters used for the experimental investigation. Values in bold represent the most often used combinations provided in the results section.

| Parameter | | Value(s) |
|---|---|---|
| Number of actions | $n$ | **5,10,50**,500,1000 |
| Number of criteria | $m$ | 2,**3**,5,**7**,10 |
| Evaluation distribution | | D1, D2, D3, D4, D5, **MX** |
| *Ex post* approximation models | | LiR, **P3R** |
| *Ex ante* approximation models | | **PLA**, PPA |
| Runs per instance config. | $N_{\text{trials}}$ | 1000 |

explicitely presented here show very similar behaviour and confirm our observations.

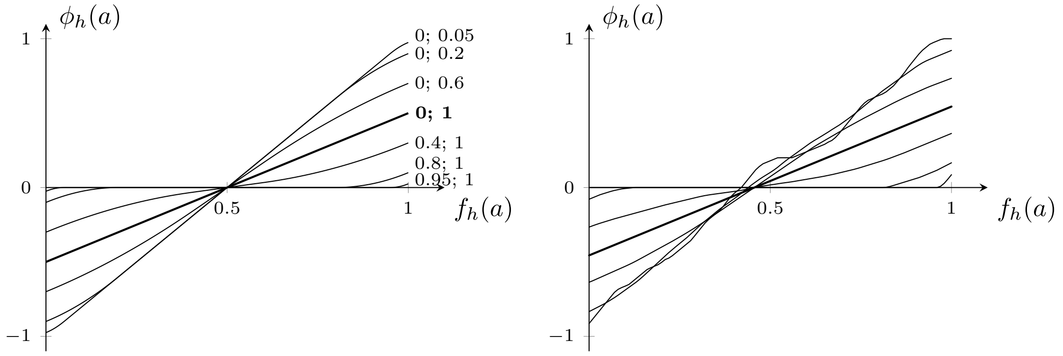
As a first validation, we visually compare both plots of Figure 7. It shows that, although the experimental results displayed in (b) are based on a relatively small instance of $n = 20$ randomly generated evaluations (with a uniform distribution D1), these results are close to the "theoretical" continuous results (a). This suggests that general features deduced from the theoretical model could also satisfyingly yield for practical instances. In a further step, we will investigate how the differences between the model and the practical results can be quantified.

As a further validation of our approximation models, we compare their quality, measured by the rank concordance ratio $\kappa$, with two other models: 1) *ex post* linear regression of P2-ranking (LIR); 2) *ex post* 3rd degree polynomial regression of P2-ranking (P3R). The comparison (Figure 8) shows that:

1. The *ex post* approximation models outperform the *ex ante* models. However, on an absolute basis, the latter still give good results: for the series of experiments, the PLA model misranks at most respectively one out of ten or three out of fifty actions.

2. Both *ex ante* models, PPA and PLA, give very comparable results. Since PLA has a simpler formulation, we will mainly report results with respect to this model in the following.

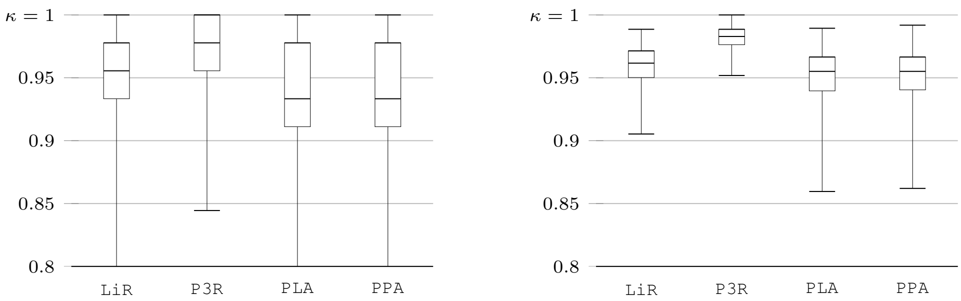
The use of "mixed-distribution" (MX) random problem instances, i.e., instances where the distribution for each criterion is uniform randomly selected from D1,. . .,D5, provides significantly worse results than any "single-distribution" random instance (Figure 9). This suggests that results for such MX distribution instances could




(a) Continuous model for $n \to \infty$ actions.

(b) Random instance of $n = 20$ actions.

Figure 7: The visual comparison of the theoretical continuous model (assuming $n \to \infty$) with a discrete instance of $n = 20$ actions tends to confirm the validity of the continuous approximation model. Indeed, the "theoretical" continuous functions (a) are very similar to the results obtained for a randomly generated discrete set of $n = 20$ actions (b). The pair of values attached to each plot of (a) are respectively the corresponding indifference and preference thresholds: "$q_h$; $p_h$". The same parameter values are used and appear in the same order for (b).




(a) $n = 10$ actions

(b) $n = 50$ actions

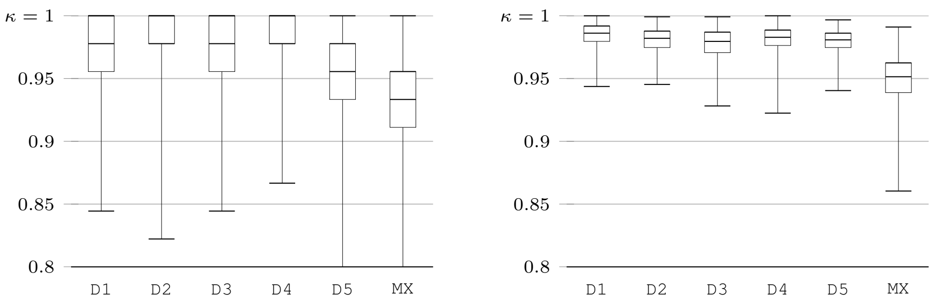
Figure 8: The boxplots compare four types of net flow score approximation models: *ex post* linear regression (`LiR`); *ex post* 3-rd degree polynomial regression (`P3R`); *ex ante* piecewise linear approximation (`PLA`); and *ex ante* piecewise polynomial approximation (`PPA`). The results are shown for 1000 runs over mixed-distribution randomly generated action sets of 10 (resp. 50) actions and 7 criteria.

be taken as a "pessimistic" point of view and used as a lower bound for most problem instances. A more detailed investigation of the distribution's influence on the results could be the topic of further investigations.

## 5.3 Empirical bounds for the use of our model

Figure 10 shows, for different instance sizes, the complement to 1 of the cumulated density function (CDF) of $\kappa$, for a series of 1000 runs. Concretely, the plots give the approximated probability of reaching at least a given similarity, measured by the rank concordance ratio $\kappa$. In the following, we will often refer to this ratio as a measure of quality: the higher $\kappa$, the better the approximation of PROMETHEE II's rankings by our piecewise linear model. Several observations can be done on the basis of these plots:

1. As could be expected: a higher number of actions increases the approximation's quality.
2. The quality curve converges to an "extreme curve" (approximated by the plots for $n = 1000$), which indicates that there exists an upper bound for the approximation quality. In other words, whatever the instance size, it will not in general be possible for our `PLA`-model to produce the same ranking as




(a) $n = 10$ actions

(b) $n = 50$ actions

Figure 9: The five distribution types `D1`-`D5` used for generating random instances yield comparable results for the `PLA` approximation when only one type is used for all criteria. For mixed instances (`MX`), i.e., randomly chosen distribution type for each criterion, are significantly worse. Results are given for 1000 repetitions and $m = 5$ criteria.

Table 3: Probability $P(\kappa > x)$ to reach a rank concordance ratio $\kappa$ that is at least as high than $x$ for mixed-distribution randomly generated instances of different sizes.

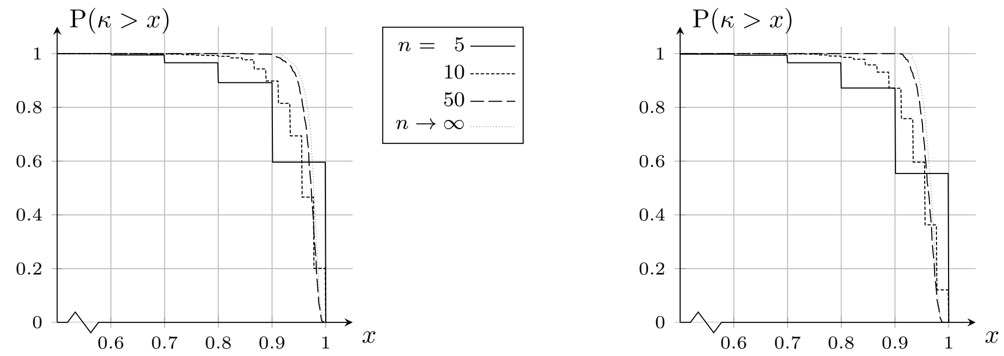
| | $n = 5$ | | | | | $n = 10$ | | | | | $n = 50$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x \setminus m$ | 2 | 3 | 5 | 7 | 10 | 2 | 3 | 5 | 7 | 10 | 2 | 3 | 5 | 7 | 10 |
| 0.70 | 0.98 | 0.97 | 0.97 | 0.97 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.80 | 0.93 | 0.89 | 0.89 | 0.87 | 0.89 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.90 | 0.93 | 0.89 | 0.89 | 0.87 | 0.89 | 0.92 | 0.90 | 0.88 | 0.87 | 0.86 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.95 | 0.67 | 0.60 | 0.59 | 0.55 | 0.55 | 0.75 | 0.69 | 0.61 | 0.60 | 0.58 | 0.89 | 0.88 | 0.73 | 0.75 | 0.69 |
| 0.99 | 0.67 | 0.60 | 0.60 | 0.55 | 0.55 | 0.24 | 0.20 | 0.13 | 0.12 | 0.13 | 0.07 | 0.04 | 0.00 | 0.00 | 0.00 |

PROMETHEE II.

3. Taking the opposite point of view, the results also show that a satisfying approximation (depending on a chosen quality level) can be reached, even for relatively small instance sizes that are frequently encountered in actual MCDA problems. *Example*: For instances of $n = 10$ actions and $m = 7$ criteria, a concordance of 90% can be reached with a probability of 87%. More numerical results are presented in Table 3.

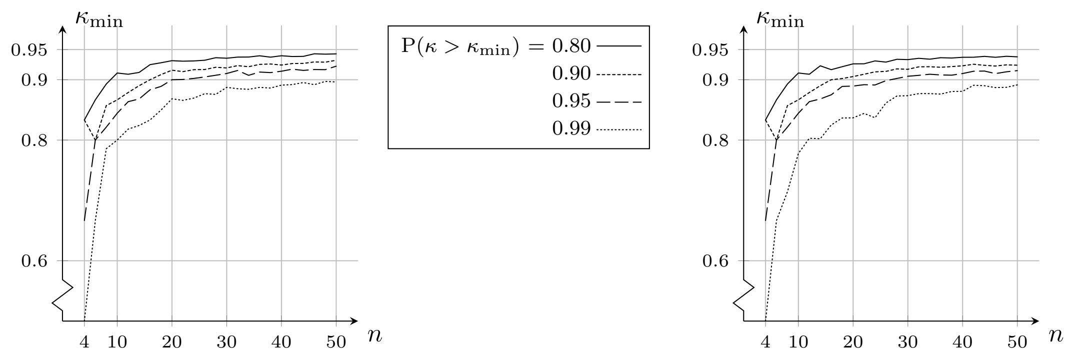
Figure 11 shows the same results from another perspective. For a given probability $P$ that measures some sort of required accuracy level of the approximation quality, the plots represent the minimum quality $\kappa$ that is reached as a function of the number $n$ of actions.

The question that naturally arises when using an approximation model, is to locate, if possible, regions where its performance is relatively better or worse. For our concern, we search for the actions that are not ranked appropriately with respect to the original `P2`-ranking. The lower plot of Figure 12 represents the distribution of misranked actions with respect to the original rank. The result can be quantitatively explained by the shape of the plot that represents the average net flow score (over 1000 runs) for each rank position (Figure 12, upper plot). The results show that the approximation should be even more satisfying than the rank concordance ratio indicates, since the actions ranked among the first or the last few are often considered with more attention. We could think of an additional metric that takes this into consideration, by taking rank concordance of well and badly ranked actions more into consideration as the middle-ranked ones. This could, for instance be done by adapting the generalized Kendall's rank correlation $\tau$ (Kumar and Vassilvitskii, 2010) to our needs.




(a) $m = 3$ criteria (b) $m = 7$ criteria

Figure 10: The probability $P(\kappa > x)$ to reach a rank concordance ratio $\kappa$ that is at least as high than $x$ for mixed-distribution randomly generated instances evolves as the number $n$ of actions and $m$ of criteria increases.




(a) $m = 3$ criteria (b) $m = 7$ criteria

Figure 11: For different probabilities $\mathrm{P}(\kappa > \kappa_{\min})$, the plots show the minimum quality $\kappa_{\min}$ that can be achieved with respect to the number of actions $n$ and depending also on the number $m$ of criteria.

# 6 Conclusion

The Promethee II method uses pairwise action comparisons to build a complete ranking over the set of considered alternatives. However, building this ranking is computationally demanding, and represents a significant drawback for Promethee II when tackling MCDA ranking problems that are very large and/or have to be computed very often. Being able to "switch" to an approximated model with linear complexity when a compromise between ranking accuracy and computation speed is affordable would therefore be of great practical interest.

In this paper, we propose such a model. It is based on a piecewise linear approximation of the Promethee II net flow score. Taking the usual Promethee preference parameters, i.e., weights and indifference/preference thresholds, an approximation of an action's net flow score is provided by a function that only depends on its evaluations. The approximated scores are then used to determine a complete ranking over the set of considered actions.

An experimental study has provided us with quantitative evaluations of the approximation's quality, yielding the minimum quality level that can be reached with a given probability. This has been done for a variety of instance sizes. Practically, these results provide empirical bounds on the instance size as from which the ranking may be approximated by our model.

On a more theoretical level, we have seen that piecewise linear formulation of our model only depends on the threshold's average value, $\lambda_h = \frac{1}{2}(q_h + p_h)$, for each criterion. This suggests that the ranking may, to

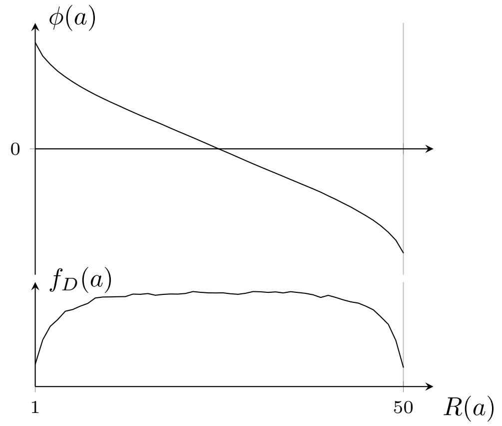



Figure 12: The distribution of misranked actions $f_D$ with respect to the original P2-ranking $R(a)$ shows that extreme ranks, i.e., the best and worst few, are relatively stable when compared to centrally-ranked actions. This is due to the shape of the net flow score, shown above: $\phi(a)$ is more discriminant in the extreme rank regions. Results are plotted for 1000 repetitions of the PLA-model on randomly generated instances (with mixed distributions) of 50 actions with 3 criteria.

a certain extent, only depend on that unique parameter, instead of a set of two threshold parameters. This feature, that has been confirmed qualitatively by a simple set of experiments, may provide a deeper insight into Promethee II's internal structure. In addition, it could also be usefully applied in eliciting procedures. Indeed, to the best of our knowledge, eliciting methods for Promethee II (Eppe et al., 2011; Eppe and De Smet, 2012; Frikha et al., 2010) have been limited to finding a representative weight vector, leaving the eliciting of threshold parameters aside, mostly because of unconclusive results. Replacing the eliciting of the threshold parameters by one single parameter could offer a promising perspective. Both the theoretical relation between threshold parameters as well as its practical implications should be studied in more depth in the future.

# References

Behzadian, M., Kazemzadeh, R., Albadvi, A., Aghdasi, M., 2010. PROMETHEE: A comprehensive literature review on methodologies and applications. European Journal of Operational Research 200 (1), 198–215.

Brans, J.-P., Mareschal, B., 2005. PROMETHEE methods. In: Figueira et al. (2005a), Ch. 5, pp. 163–195.

Brans, J.-P., Vincke, P., Mareschal, B., 1986. How to select and how to rank projects: The PROMETHEE method. European Journal of Operational Research 24 (2), 228–238.

De Smet, Y., Mareschal, B., Verly, C., 2009. Extending the PROMETHEE II method to continuous and combinatorial multi-objective optimization problems: a first model. In: Industrial Engineering and Engineering Management, IEEE 2009 International Conference, IEEM2009. IEEE Press, Piscataway, NJ, pp. 1608–1611.

Dyer, J. S., 2005. MAUT – multiattribute utility theory. In: Figueira et al. (2005a), Ch. 7, pp. 265–295.

Eppe, S., De Smet, Y., 2012. Studying the impact of information structure in the PROMETHEE II preference elicitation process: A simulation based approach. In: Greco et al. (2012), pp. 383–392.

Eppe, S., De Smet, Y., Stützle, T., 2011. A bi-objective optimization model to eliciting decision maker's preferences for the PROMETHEE II method. In: Brafman, R. I., Roberts, F., Tsoukiàs, A. (Eds.), Algorithmic Decision Theory, Third International Conference, ADT 2011. Vol. 6992 of Lecture Notes in Computer Science. Springer, Heidelberg, Germany, pp. 56–66.

Figueira, J. R., Greco, S., Ehrgott, M. (Eds.), 2005a. Multiple Criteria Decision Analysis, State of the Art Surveys. Springer.

Figueira, J. R., Mousseau, V., Roy, B., 2005b. ELECTRE methods. In: Figueira et al. (2005a), Ch. 4, pp. 133–162.

Frikha, H., Chabchoub, H., Martel, J.-M., 2010. Inferring criteria's relative importance coefficients in PROMETHEE II. International Journal of Operational Research 7 (2), 257–275.

Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R. R. (Eds.), 2012. 14th International Conference on Information Processing and Management of Uncertainty, IPMU 2012, Catania, Italy, July 9 - 13, 2012. Proceedings. Vol. 300 of Communications in Computer and Information Science. Berlin / Heidelberg.

Guitouni, A., Martel, J.-M., 1998. Tentative guidelines to help choosing an appropriate MCDA method. European Journal of Operational Research 109 (2), 501–521.

Kendall, M., Gibbons, J. D., 1990. Rank Correlation Methods. Edward Arnold.

Kumar, R., Vassilvitskii, S., 2010. Generalized distances between rankings. In: Rappa, M., Jones, P., Freire, J., Chakrabarti, S. (Eds.), Proceedings of the 19th International Conference on World Wide Web, WWW 2010. ACM press, New York, NY.

Marinoni, O., 2006. A discussion on the computational limitations of outranking methods for land-use suitability assessment. International Journal of Geographical Information Science 20 (1), 69–87.

Pivert, O., Smits, G., 2012. Towards an efficient processing of outranking-based preference queries. In: Greco et al. (2012), pp. 471–480.

Roy, B., 2005. Paradigms and challenges. In: Figueira et al. (2005a), Ch. 1, pp. 3–24.

Siskos, Y., Wäscher, G., Winkels, H.-M., 1984. Outranking approaches versus MAUT in MCDM. European Journal of Operational Research 16 (2), 270–271.