Multiple-criteria performance ranking based on profile distributions:

An application to university research evaluations

Pierre L. Kunsch, Corresponding Author, Vrije Universiteit Brussel (VUB), BUTO Dept., Pleinlaan 2, Brussels, Belgium, email: pkunsch@vub.ac.be

Alessio Ishizaka, Portsmouth Business School, Centre for Operations Research and Logistics, University of Portsmouth, Portsmouth, PO1 3DE, United Kingdom,
email: alessio.ishizaka@port.ac.uk

Abstract

This article addresses a category of multi-criteria ranking problems in which performance evaluations of the ‘objects’ or ‘alternatives’ to be ranked are not given by unique numbers but by performance profile distributions running over an ordered set of performance score levels. The numerical values to be given to score levels are not specified *a priori* using cardinal scales. A weighted sum approach is developed based on order statistics to combine the individual profile distributions. In this way, a global ranking indicator is obtained, considering not only mean distribution values but also standard deviations. As a test of feasibility, the resulting Profile Ranking with Order Statistics Evaluations (PROSE) approach has been applied to the performance profile distributions provided by the UK Research Excellence Framework 2014, evaluating the quality of UK research.

Keywords: Multi-criteria ranking; Performance profiles; Order statistics; Ranking indicator; 2014 Research Excellence Framework

1. Introduction

This article addresses a new category of multi-criteria ranking problems in which evaluations of the ‘objects’ or ‘alternatives’ are not given by unique numbers but by performance profile distributions normalised to one. Such profile distributions have properties analogous to probability distributions; they run over a range of ordered ordinal score levels, e.g. for five score levels: *0 (not classified), 1 (marginal), 2 (mediocre), 3 (rather good), 4 (good)*. This type of problems is more and more common in practice because it is often difficult to give a crisp score. However, they have not been yet considered in the academic literature.

The question then arises if and how such ordinal score levels may be combined by simple arithmetic through a weighted sum to provide a global score, as is usual in many multi-criteria aggregation methods. Taking the weighted sum approach as given, we shall discuss in this paper the way in which the profile distributions are to be combined in a rigorous way: without anticipating the results, we conclude that a weighted sum mean value is not sufficient for ranking the ‘objects’ or ‘alternatives’ and thus a more elaborate ranking indicator must be considered. In addition, order statistics are introduced to value the ordinal score levels.

The applicability of the approach, called Profile Ranking with Order Statistics Evaluations (PROSE), is tested using a real-world case study: the ranking of UK research units based on data from the 2014 UK Research Excellence Framework (2014 REF) exercise [1]. The data resulting from the 2014 REF have exactly the format described, which to our knowledge is a unique case in university research evaluations. Before analysing this important case further, the problem can be made more concrete by means of a simple didactic example to present the ranking approach in PROSE. The most common and well-known example of this type is found in school evaluations. If the school teacher gives a pupil a fractional grade, e.g. 2.7, for a course on the ordinal 0–4 score scale, the thinking is: the pupil is well above the 'mediocre' grade 2, but has not entirely achieved the objectives of the 'rather good' grade 3 and therefore my subjective judgement is that the performance is at the level 2.7, i.e. 30% lower than 3, but 70% above 2. In this simple example, the normalised distribution is a binary one (i.e. 0% at 0, 0% at 1, 30% at 2, 70% at 3, 0% at 4); it is thus easy to verify that the mean grade value is indeed 2.7.

Distributions could have more than two non-vanishing components, for instance if grades are given by several school teachers for the same discipline. Considering several disciplines, the common – and probably unique school practice giving no room in general to second thoughts about its validity – is to combine the fractional grades between two ordered ordinal scores by discipline (i.e. criteria) using a weighted sum. Most of the time, all weights are taken to be equal, unless other directives are imposed by the school authorities: for instance, that more weight should be given to mathematics than to music. This does not call into question the weighted sum approach. But the weighted sum approach used as such in the specified setting, in addition to its lack of rigour in combining arithmetically ordinal level scores, which is not directly permissible, has the well-known drawback that only one value results in the form of an average final grade: therefore, compensation effects are possible between good and less good grades. Another pupil with the same average grade may be more 'regular' in the sense that the dispersion of the global grade distribution would be smaller, but this is generally not considered by school teachers. This article presents a way of avoiding this shortcoming by considering the spread in performance.

The article is structured as follows: section 2 presents stepwise the proposed PROSE approach, using simple school examples which, as set out, have a comparable input data structure. A useful ranking indicator is introduced. Section 3 is dedicated to discussing the 2014 REF. First, a detailed introduction to how the profile data were obtained is provided. Profile distributions and weights are applied in the second part to test the PROSE approach developed in section 2. Section 4 concludes the work and presents future research developments regarding multi-criteria ranking problems with a similar structure.

Two UK research rankings resulting from the 2014 REF data are given in Appendices A and B.

2. Ranking approach: PROSE

Mathematical notations are defined (section 2.1); we address the basic weighted sum approach in multi-criteria problems (section 2.2); we show by means of a school example how this method can be modified when the measures of performance are profile distributions rather than single values (section 2.3.1). We extend the weighted sum to define the ordinal performance score levels as order statistics distributions (section 2.3.2).

2.1. Mathematical notations and glossary

A bar above a variable indicates the mean value.
PROSE = **P**rofile **R**anking with **O**rder **S**tatistics **E**valuations

$k = 1, 2, ..., K$: running index of K criteria in decreasing order of importance
$m = 0, 1, 2, ..., M$: running index of M performance levels
$D = (d_{km})$: $K \times (M+1)$ matrix of performance profiles
$G = (g_m)$: $1 \times (M+1)$ row vector of the global performance distribution
$\Omega = (\omega_1 = 1, \omega_2, ..., \omega_K)$: vector of ordered-weight ratios in decreasing order
$\varphi(\cdot)$: distribution law of a random variable
$S = (s_m)$: $1 \times (M+1)$ row vector of performance scores
$\sigma_S = (\sigma_m) : 1 \times (M+1)$ row vector of performance-score standard deviations
v : mean global score with distribution G and discrete values in $\bar{S}$
σ_v : standard deviation of v
V : mean global score with distribution G and continuous values in S
σ_V : standard deviation of V
V_C : ranking indicator combining the mean score $\bar{V}$ and σ_V
$x_{m:M-1}$: m^{th} variable in a set of $M-1$ increasingly ordered random variables (order statistics)
$G \bullet S$: scalar product of G and S

2.2. Basic weighted sum

Let us recall the well-known and most common weighted sum approach to multi-criteria ranking problems. Let us assume $k = 1, 2, ..., K$ criteria , where $K > 1$ [21]. Defining a positive weight for each criterion, representing its relative importance with respect to the most important criterion placed in the first position, gives the ordered weight vector $\Omega = (\omega_1 = 1, \omega_2, ..., \omega_K)$. Eliciting for each alternative for all criteria a performance score on a common dimensionless scale gives the multi-criteria score vector $(v_1, v_2, ..., v_K)$ for some object or alternative. The weighted sum in (1) aggregates this information to give the global multi-criteria score v for each object or alternative. The resulting v's are then used for ranking:

$$v = \sum_{k=1}^{K} \left(\omega_k \Big/ \sum_{l=1}^{K} \omega_l \right) v_k \qquad (1)$$

In the next section, (1) is adapted to consider profile distributions rather than a unique performance value by criterion and to give values from distributions to the scores on the dimensionless scale.

2.3. Weighted sum with profile and score distributions

2.3.1. Performance profile distributions

To illustrate the extension of the weighted sum approach to performance profiles, i.e. PROSE, let us again use the school example mentioned in the introduction: a school pupil, say John, is evaluated on his language ability in Spanish. Three criteria are used: 'speaking', 'reading' and 'grammar', ordered by importance as follows: 'speaking' > 'reading' > 'grammar'. Assume that the school teacher, following her own personal subjective view on the relative importance of the criteria compatible with this ranking, gives not-normalised weight values on a ratio scale, resp.: 1, 0.8 and 0.6.; 0.8, resp. 0.6 represent the importance ratios between reading and grammar with respect to 'speaking', the most important criterion. To calculate the global profile in Table 1 the weights must be normalised, so that we obtain $1/(1+0.8+0.6)=41.7\%$ for speaking, $0.6/(1+0.8+0.6)=25\%$ for reading, and $0.6/(1+0.8+0.6)=25\%$ for grammar. We assume further that the evaluation is made with reference to five levels ordered from 0 to 4, where 0 denotes no performance at all and 4 is the best performance level. A discrete real-valued score s_m is associated with each level $m=0,\ 1,\ldots,\ 4$. The evaluation of the language teacher is generally not a single value, but a binary profile, analogous to a probability distribution because the teacher hesitates between two neighbouring levels. In the example given in Table 1, John attains a grade of 2.7 for 'speaking' to be interpreted as a binary distribution: he has achieved 70% of the speaking objective corresponding to level 3. He is therefore 30% below performance level 3 and 70% above performance level 2. The mean value of the (30%, 70%) distribution for the (2, 3) levels is indeed 2*0.3+3*0.7=2.7.

Let us formalise the PROSE approach regarding this aggregation of profiles. The profiles for all criteria are entered a performance matrix, $D=(d_{km})$, in which the three criteria are labelled $k=1,\ 2,\ 3$ and the four scores levels $m=0,\ 1,\ 2,\ 3,\ 4$.

		Level scores					
Weight	Criteria	0	1	2	3	4	Grade
1	Speaking			30%	70%		2.7
0.8	Reading				20%	80%	3.8
0.6	Grammar		50%	50%			1.5
Global profile			12.5%	25%	35.8%	26.7%	$\bar{v}=2.77$

Table 1: Discrete profile distribution for the three criteria in the Spanish test example

According to the previous development, the normalised global profile, $G=\left(g_m\right)$, is given by (2) and the percentage values obtained are given in the last row of Table 1:

$$G=\left(g_m\right)=\frac{\sum_{k=1}^{K}\omega_k d_{km}}{\sum_{l=1}^{K}\omega_l} \tag{2}$$

The mean global score $\overline{v}$, also called the grade point average (GPA) [40], equal to 2.77 results from the weighted sum:

$$\overline{v}=\text{GPA}=\sum_{m=0}^{M} g_m s_m=\sum_{m=0}^{M}\left(\frac{\sum_{k=1}^{K}\omega_k d_{km}}{\sum_{l=1}^{K}\omega_l}\right)s_m=G\bullet S \tag{3}$$

where (see notations in 2.1):

ω_k = positive weight of criterion k

d_{km} = value of the criterion k at score level m

K = number of criteria

$G=\left(g_m\right)$ = global profile vector

$S=\left(s_m\right)$ = vector of discrete score values

M = number of positive performance-score levels

$G\bullet S$ = scalar product of G and S

However, using the sole global mean score $\overline{v}$ =GPA results in a loss of important information concerning the profile dispersion. To observe this, we consider two global profiles for the five score levels previously introduced (0%, 50%; 0%; 0%; 50%) and (0%, 0%; 50%; 50%; 0%), giving the same mean global score $\overline{v}=2.5$. The mean value of the first global profile hides the poor performance at score level 1 by the excellent performance obtained at score level 4, while the second profile is more balanced towards two 50% median performances at score levels 2 and 3; the standard deviation of the first profile is 1.5 compared to 0.5 for the second. The spread in performance becomes problematic, for example in terms of security or quality, as poor individual scores increase the risk of damaging consequences, although there is compensation with better scores. Therefore, the global mean score needs to be complemented with the measure of dispersion σ_v, the standard deviation of v given in (4):

$$\text{variance }(v) = \sum_{m=1}^{M} g_m \left(s_m - \bar{v}\right)^2 = \sum_{m=0}^{M} g_m s_m^2 - 2\,\bar{v} \sum_{m=1}^{M} g_m\, s_m + \bar{v}^2 = \sum_{m=0}^{M} g_m s_m^2 - \bar{v}^2$$
$$\rightarrow \sigma_v^2 = \overline{v^2} - \bar{v}^2 \qquad (4)$$

To obtain a safer ranking based not only on GPA, we combine both mean and spread in performance $\left(\bar{v}, \sigma_v\right)$ to the 'ranking indicator'. Our proposition is to consider the interval $\bar{v} \pm \sigma_v$ as an approximation to the range of values obtained for the global score, replacing the generally not well-behaved score distribution by a rectangular distribution centred on the mean value. Keeping only the lower value for being on the safe performance side, we define the combined ranking indicator as follows:

$$V_C = \bar{v} - \sigma_v \qquad (5)$$

The data in Table 1 and (5) give a value for the ranking indicator $V_C = 2.77 - 0.981 = 1.79$. The dispersion-corrected 1.79 of the ranking indicator is significantly smaller than GPA=2.77; the performance spread gives a more trustworthy measure of the pupil's actual performance.

2.3.2. Ordered score-level distributions in PROSE

In this sub-section, we examine how we may combine ordinal score levels in the weighted sum (3) using PROSE. This combining practice is commonly accepted without second thought when using weighted sums, for example in schools. However, combining ordinal-score levels arithmetically is not directly permitted, while we need cardinal values for the scores. Score surrogates is a well-known elicitation technique in decision theory when only ordinal information is available in the form of a ranking [10]; [17]; here the ranking of the score levels; the score surrogates are the centroids of the ranking simplex. To introduce those concepts let us consider now that the school teacher uses alphabetic grades instead of ordinal numbers for the performance levels, as is usual in some countries. For five levels, there is a one-to-one correspondence with the ordinal score levels in Table 1: grade E in Table 2 corresponds to the zero-score level 'no performance', whereas A corresponds to the highest level 4. However, no numerical value is given *a priori* to the levels and therefore no global mean score or standard deviations can be calculated using (3) and (4), as with the numerical score levels in Table 1. We thus now need to associate cardinal values to the score levels.

Let us call these cardinal values, some are not unique, s_m, $m = 0,1,2,3,4$, corresponding to the levels E, D, C, B and A respectively. The sole information at our disposal is the ranking

$s_0 < s_1 < s_2 < s_3 < s_4 = 4$. We fix $s_0 = 0$ because level E represents the 'no performance' situation. The upper A grade is the 'best performance' score, not allowed to take larger or smaller values than 4, therefore its value must be also fixed. The larger score value fixed at 4 is here freely chosen; any strictly positive value like 1, 10, 100 etc., would only induce a rescaling of the set of scores without consequences. Dividing the intermediate scores by $s_4 = 4$ we obtain the bounded ranking-simplex in $3D$:

$$0 < s_1/s_4 < s_2/s_4 < s_3/s_4 < 1 \tag{6}$$

		Level scores				
Weight	Criteria	E	D	C	B	A
1	Speaking			30%	70%	
0.8	Reading				20%	80%
0.6	Grammar		50%	50%		
Global profile			12.5%	25%	35.8%	26.7%

Table 2: Example of discrete profile distributions on three criteria with alphabetical level scores.

The ranking-simplex defines the feasibility domain of the intermediate scores. The three intermediate surrogate scores (divided by 4) are given by the centroid coordinates of the simplex, which is a $3D$ tetrahedron, see Fig.1. The centroid is here the same as the barycentre, or centre-of-gravity (COG) of the simplex, because all point values within its volume are homogeneously distributed. It is well known that the COG of a tetrahedron is the arithmetic averages of the four extreme points $(0,0,0);(0,0,1);(0,1,1);(1,1,1)$, i.e., $(1/4, 2/4, 3/4)$. Adding the fixed score values, the five surrogate scores are accordingly $(0,1,2,3,4)$. The values are at the same time the score means, thus $\left(\overline{s_m} = m, m = 0,1,2,3,4\right)$: this is trivially true for s_0, s_4, which are fixed, and the three centroid scores are the arithmetic averages of all scores in the simplex. With this property we can calculate all means and standard deviations by integration over the tetrahedron volume, for instance $\overline{s_3} = 3!\int_0^1 x_3 dx_3 \int_0^{x_3} dx_2 \int_0^{x2} dx_1 = 6\int_0^1 x_3^3/2\, dx_3 = 3/4$ (the integration volume is $\int_0^1 dx_3 \int_0^{x_3} dx_2 \int_0^{x2} dx_1 = \frac{1}{3!}$), but this is a lengthy exercise for standard deviations, and it is easier to use the theory of *order statistics* to obtain these values.

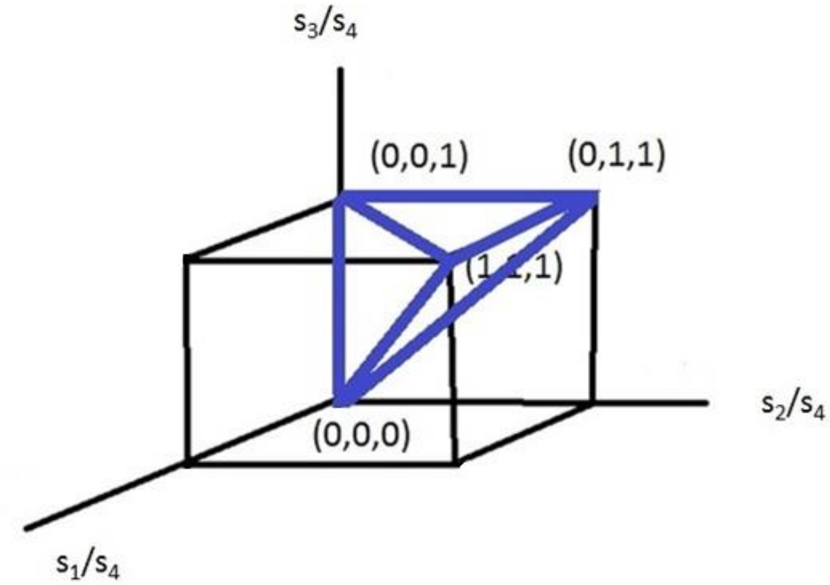

Figure 1: The ranking simplex containing the feasible values for the ranked score ratios, $s_m/s_4, m=1,2,3$ in (6) with $s_4=4$ (fixed value). The surrogate score ratios are the simplex-centroid coordinates, i.e., the arithmetic averages of the extreme points coordinates $(1/4, 2/4, 3/4)$ and of all point values in the tetrahedron.

For $M \geq 3$, consider $M-1$ random variables $x_i; i=1,\ldots,M-1$ with a uniform probability distribution in $[0,1]$; after ordering these variables, *called parent variables*, in increasing order we obtain $M-1$ new ordered random variables $x_{m:M-1}; m=1,\ldots,M-1$ in the simplex:

$$0 < x_{1:M-1} < x_{2:M-1} < \ldots < x_{M-1:M-1} < 1 \quad (7)$$

Order-statistics theory [7] gives the probability distributions of $x_{m:M-1}$'s. For uniformly distributed parent variables $\varphi(x_{m:M-1}) = Beta(m, M-m)\ 0 < m \leq M-1$. The mean and standard deviation of $x_{m:M-1}$ are respectively:

$$\overline{x_{m:M-1}} = \frac{m}{M} \quad (8)$$

$$\sigma^2_{m:M} = \frac{m(M-m)}{M^2(M+1)} \quad (9)$$

For $M=4$ and writing $s_m = 4x_{m:3}, m=1,2,3$, we obtain means and standard deviations of the five scores:

$$\overline{S} = (s_0=0, \overline{s}_1=1, \overline{s}_2=2, \overline{s}_3=3, s_4=4) \quad (10)$$

$$\sigma_m^2 = 16\frac{m(4-m)}{16\times 5} = 0.2\,m(4-m)$$
$$\rightarrow Variance(S) = \sigma_S^2 = (0, 0.6, 0.8, 0.6, 0) \tag{11}$$

In summary, the mean scores are identical to the discrete ordinal scores in Table 1; however, three scores $s_m, m = 1,2,3$ now take continuous values obeying order statistics: means and standard deviations are given by (10) & (11).

To obtain the ranking indicator (5) the mean and standard deviation of the global score, called V are needed. V is obtained with the random S scores, generalising v obtained with mean scores $\overline{S}$, see (3) and (4); we get with the global performance distribution G, and the notation $\bullet$ for the scalar product:

$$V = \sum_{m=0}^{M} g_m\, s_m = G \bullet S \tag{12}$$

The mean of a sum of random variables being equal to the sum of the means, even if these variables are not statistically independent, which is the case for the order statistics because of the ordering, we get from (12):

$$\overline{V} = G \bullet \overline{S} = \overline{v} \tag{13}$$

(13) shows that the global mean score obtained before with (3) is unchanged and equal to 2.77, as previously obtained in the school example for the GPA. This is because the mean score vector is given in (10): its components appear to be just the same as the set of the four ordered ordinal level scores. This gives validity to the common weighted sum school practice of using ordinal scores to compute global GPA grades. However, this is only partly adequate, as spreads in global grades are neglected in GPA. Indeed, an increased variance is expected, not only due to the global profile spread as in (4), but also the supplementary variance (11), stemming from the order statistics. More formally, we must now consider, in addition to the profile-induced correction, that the score vector S is represented by probability distributions (order statistics), and not by unique ordinal values: therefore $\overline{S^2} \neq \overline{S}^2 = (0,1,4,9,16)$. $Variance(V)$ is obtained by the usual variance formula [48] by summing up over the global profile G, and integrating over the order statistics, the squared differences of random s_m's with the mean value $\overline{V}$. The definition formally expressed in (15) below generalises (4), used for calculating the variance of v with the discrete global distribution G and crisp scores $\overline{s_m}$'s; it brings the additional contribution stemming from the integration over the order statistics. Using $\overline{V} = \overline{v} = \sum_{m=1}^{4} g_m\, \overline{s_m}$ from (3); (13) we obtain:

$$Variance(V) = \sigma_V^2 = \overline{\sum_{m=1}^{4} g_m \left(s_m - \overline{V}\right)^2} = \overline{\sum_{m=1}^{4} g_m \left(s_m^2 - 2\overline{V}\, s_m\right)} + \overline{V}^2$$
$$= \sum_{m=1}^{4} g_m \overline{s_m^2} - \overline{v}^2 = \sum_{m=1}^{4} g_m \left(\overline{s_m^2} - \overline{s_m}^2\right) + \left(\sum_{m=1}^{4} g_m \overline{s_m}^2 - \overline{v}^2\right) \quad (14)$$

here the long bar on top indicates the mean value obtained by integration over the s_m's order statistics. The second term of this expression is σ_v^2 from (4), and s_m's variances are $\sigma_m^2 = \left(\overline{s_m^2} - \overline{s_m}^2\right)$, so that (14) becomes:

$$\sigma_V^2 = Variance(v) + \sum_{m=1}^{4} g_m \sigma_m^2 = \sigma_v^2 + G \bullet \sigma_S^2 \quad (15)$$

In addition to the variance of v in (4), the new term $G \bullet \sigma_S^2$ in (15) results from the S variances (11). In the school example, the total variance becomes $\text{Variance}(V) = \sigma_V^2 = 1.21^2$ instead of $\text{Variance}(v) = \sigma_v^2 = 0.981^2$ in 2.3.1, with a substantial increase of 23% in the standard deviation.

In summary, the global score $\overline{V}$ is unchanged from $\overline{v} = 2.77$, but the standard deviation σ_V now increases to 1.21 by combining both spreading effects of the global profile G, and of the random-scores. The ranking indicator (5) becomes:

$$V_C = 2.77 - 1.21 = 1.56 \quad (16)$$

This value is significantly smaller than 1.79, obtained from the discrete global profile in 2.3.1, and much smaller than the uncorrected GPA=2.77. These results confirm that the corrected weighted sum formula leading to the ranking indicator (5) is badly needed. Thanks to this indicator in PROSE, a much more accurate picture of the actual global performance is obtained.

3. Case Study: The REF Framework

3.1. Review of university research evaluations

The evaluation of education by ranking universities has become increasingly popular over the past few years. Some examples in the UK are the *Times Higher Education*, *The Complete University Guide*, *The Guardian University Guide* and The *Sunday Times University Guide*. These rankings have a sizeable impact on universities as they may give some indication of prestige and exert a direct influence on the number and quality of applicants.

As an introductory remark, the evaluation of universities or departments is a complicated process as it is based on several conflicting criteria. Multi-criteria decision analysis (MCDA) methods are therefore appropriate for evaluation, as is data envelopment analysis (DEA), used to rank universities and departments according to their efficiency. A detailed discussion of the steps requiring the definition of several specific input specific data for actual evaluations using MCDA or DEA (such as lists of criteria and technical parameters, e.g. weights, trade-offs, thresholds, etc.) is beyond the scope of this article, which entirely dedicated to the ranking once the input data have been agreed upon.

DEA, which is the technique most commonly applied, has been used for British universities [22-26, 44, 46, 49, 51], as well as universities from other countries: Greece [29, 31], Japan [6], Italy [2-5], Australia [13], Spain [35], South Korea [53], the USA [12, 14, 47], Poland [37], Germany [18, 30], Bulgaria [50], Taiwan [27] and Turkey [11]. However, as the main objectives of DEA are to provide a measure of university performance with respect to an input/output-type analysis, it brings few useful methodological elements for the ranking purpose here, in contrast to MCDA techniques which have been used for both evaluation and ranking. VIKOR was used in isolation in [38, 39] and combined with AHP in [52]. ELECTRE III was used in [19] for the evaluation of UK universities. [34] used AHP to determine the weights of three criteria (research, education and environment) and a normalised weighted sum to evaluate a sample of 35 top universities forming the Academic Ranking of World Universities. None of these approaches used performance profiles in the evaluations.

Several countries have introduced national assessment exercises related to the research quality of universities [42] comparable to the 2014 REF. The REF is performed approximately every six years by the Higher Education Funding Council, on the last occasion in 2014. More details are provided in next section concerning the way in which data were obtained in the evaluation step. Let us only note here that the REF consists of a peer assessment of UK universities' research in all disciplines to meet the need for regular national evaluation of universities' research units, primarily for funding, but also to define organisational issues [9, 36, 41]. The assessment is performed by a panel of academic researchers and research users nominated by a wide range of academic constituencies. The panel agree on common data, such as the ordered score levels (0* to 4*) for evaluating research quality and the weights given to three independent criteria. We accepted all input and output data available from the 2014 REF without any further analysis of the methodology used for the evaluation steps.

No mean values of the global profiles appear in the published 2014 REF reports, although this would have been an immediate task in obtaining a first ranking. This is because the 2014 REF was not intended for producing a ranking and the profile distribution has been adopted to avoid having ranking. However, the results are used by the research authorities for financing purposes, directly employing the proportions of 3* and 4* publications in the global profiles of each research unit, obtained from a plain weighted sum powered by the number of entered staffs. Therefore, the ranking obtained in this article is an original contribution, beyond main objective of testing the feasibility of the PROSE approach, as no official 2014 REF ranking exists.

Because of the rather unique format, using performance profiles, none of the available MCDA ranking techniques found in the literature are directly applicable in the REF case. Moreover, because the purpose of PROSE is limited to the ranking step in the university evaluation, the upstream preference modelling of experts (decision makers) involved in the evaluation process found in most MCDA applications is beyond the scope of this study. A multi-criteria aggregating approach based on weighted sums – using without further analysis the criteria weights provided – remains the best and simplest way to proceed in obtaining a ranking.

3.2. Research Exercise Framework

This sub-section provides further detail on the data acquired from the 2014 REF, adopted as a performance-based research funding system. It was the seventh such formal assessment of research in UK universities and other research institutions. The exercise was mostly based on ex-post research evaluations through peer review judgement across 36 sub-panels. It graded 190,000 research outputs and nearly 7,000 impact case studies written by more than 52,000 full-time equivalent (FTE) staff at 154 higher education institutions. Participation is voluntary and institutions can decide which staff to include and under which discipline to include them.

As mentioned above, the evaluation of each research unit, considering the three criteria *outputs, impact* and *environment* detailed below, is given by percentage distribution profiles and not by single scores. Each performance profile considers five-star score levels: 0* (unclassified), 1* (recognised nationally), 2* (recognised internationally), 3* (internationally excellent), 4* (world-leading quality).

The three criteria are detailed as follows:

- Outputs: each person evaluated has to present their four best publications, which are evaluated according to their 'originality, significance and rigour';
- Impact: two impact case studies are needed for up to 14.99 FTE staff numbers submitted for evaluation. A further case study is needed for 15–24.99 FTE staff and so on for any further 10 FTE staff. The impact cases are assessed on their 'reach' and 'significance';
- Environment: the research environment is assessed in terms of its 'vitality and sustainability'.

As an example, the sub-panel 'Computer Science and Informatics' analysed the submission of this research unit at the University of Portsmouth, producing the following normalised 'output' profile with respect to the five-star score levels $(0\%, 4.8\%, 35.7\%, 40.5\%, 19\%)$ (see Table 3).

The weight agreed between all sub-panels for the criteria was 65% for outputs, 20% for impact and 15% for environment. FTE staff numbers were available and allowed the aggregation of unit profiles to global profiles per institution.

The star profiles have the same format for all research units; they are provided from the web site in an EXCEL® file. To apply the PROSE methodology to the 2014 REF data, we have developed a MATLAB® programme, importing and sorting the data from the Excel file to perform the computations per research unit and the aggregations of units by institution. Sorting by unit type and institutions can easily be undertaken to generate the ranking tables A.1 and B.1 in Appendices A and B.

Table 3 shows the profile data for the Computer Science and Informatics department of the University of Portsmouth (see sub-section 4.2 for the evaluation of the ranking indicator (7) in this example).

		Star scores				
Weights	Criteria	0*	1*	2*	3*	4*
65%	Output		4.8%	35.7%	40.5%	19%
20%	Impact				60%	40%
15%	Environment			65%%	35%	
Global profile (weighted sum)		**0%**	**3%**	**33%**	**44%**	**20%**

Table 3. 2014 REF profile data of the Computer Science & Informatics research unit of the University of Portsmouth

Before proceeding to evaluations with PROSE, for deriving rankings from the 2014 REF data, we reiterate clearly the limited scope of the article:

- The methodology used by the REF for the acquisition of data, the choice of star score scales, criteria and corresponding weights, the pertinence and independence of criteria, the research unit profiles obtained in this framework, etc., are not within the scope of this paper. Therefore, all data are accepted without modification for the ranking evaluations;
- For the same reason, any preference elicitation process – such as, for example, utility evaluations regarding the choice of weights or star score levels – upstream and downstream of the data acquisition was clearly not only beyond the scope of our study, but also beyond possibility due to the large number of expert sub-panels agreeing on the common framework data, which was not open to later discussion. This is the most obvious reason why MCDA techniques requiring interactive elicitation processes with the decision makers could not be considered; thus, a weighted sum approach using available data has been used;
- The rankings obtained are not interpreted here either; the scope does not go beyond developing the mathematically sound ranking approach in PROSE.

3.3. Ranking evaluations for the 'Computer Science & Informatics' sub-panel

Sub-panel 11, 'Computer Science & Informatics' (CS&I), received 89 submissions; it was one of the larger areas returned to the 2014 REF to be ranked now with PROSE. Let us use the

CS&I unit of the University of Portsmouth as a representative example; the calculation scheme in PROSE is identical for all other units across all institutions and is the same as in the school example provided in section 2.3.2. All research units are evaluated by profile, ranging from 0* to 4* levels on the three weighted criteria: outputs (O) 65%, impacts (I) 20%, and environment (E) 15%.

The matrix of the grading profile of CS&I Portsmouth is given in the performance matrix $D=\left(d_{km}\right)$ (17) (see Table 3), where the row index $k=1,2,3$ runs over the three criteria in the order *O, I, E* and the column index $m=0,1,2,3,4$ represents the number of stars in $S=\left(s_m\right)=\left(s_0=0,s_1,s_2,s_3,s_4=4\right)$:

$$D=\begin{pmatrix} 0\% & 4.8\% & 35.7\% & 40.5\% & 19\% \\ 0\% & 0\% & 0\% & 60.0\% & 40\% \\ 0\% & 0\% & 65.0\% & 35.0\% & 0\% \end{pmatrix} \qquad (17)$$

The criteria weights for *O, I, E* respectively were set by the expert panel to $\left(65\%,20\%,15\%\right)$ and can also be represented as ordered ratios $\Omega=\left(\omega_1=1,\omega_2=\frac{4}{13},\omega_3=\frac{3}{13}\right)$ with respect to the most important criterion O. Using the weighted sum (2), the global performance profile $G=(g_m)\quad m=0,\ 1,\ 2,\ 3,\ 4$ is obtained as shown in Table 3:

$$G=\left(g_m=\frac{d_{1m}+4/13d_{2m}+3/13d_{3m}}{20/13}=\frac{13d_{1m}+4d_{2m}+3d_{3m}}{20}\right)=\left(0\%,3\%,33\%,44\%,20\%\right) \qquad (18)$$

The global mean score $\overline{V}$ according to (13) is $\overline{V}=2.81$. The standard deviation including both the contributions of the profile and the score order statistics is, according to (15), $\sigma_V=1.08$; the ranking indicator (5) is $V_C=2.81-1.08=1.73$ a rather smaller value than the plain mean of 2.81.

The same type of computation is readily performed for all 89 CS&I units to obtain the corresponding ranking indicator values. The corresponding ranking is given in Table A.1, Appendix A. CS&I Portsmouth has the 40th rank.

3.4. Ranking evaluations for the whole REF panel

Going further, the ranking of all UK research institutions is obtained with PROSE in the same way. This exercise is readily performed using the available profile data per unit, given that an assumption is made regarding the weighting of each unit in each institution. The weighting used here comprises the FTE staff submitted for each unit in each institution [40]. In this way, a global institution star distribution is obtained by combining all global unit distributions using this weighting. The ordered global ranking indicators, V_C, of the 154 institutions are provided in Table B.1, Appendix B.

4. General conclusions and future work

The objective of this article has been to propose and test a ranking approach called Profile Ranking with Order Statistics Evaluations (PROSE) for an important class of multi-criteria ranking problems in which evaluation data are given as profile distributions running over a set of performance levels. A frequent case is the ranking based on school grades or more generally the evaluation of candidates or institutions by expert panels.

We chose the UK Research Excellence Framework exercise undertaken in 2014 [1] to test the feasibility and pertinence of PROSE as a representative real-world case study. 'The lure of league tables is hard to resist, despite their well-known limitations' (Pidd, 2012, chapter 10, cited in [40]) and therefore this article does not intend to propose one more ranking for the REF beyond those described in [40], all being variants of the GPA with different weighting of the individual criterion profiles. Rather, the purpose is to propose a mathematically sound weighted sum approach for this general class of problems.

The weighted sum approach in PROSE has been adapted to incorporate the use of performance profile aggregations and ordinal ordered scores. The ordinal scores representing the performance levels are modelled by means of order statistics and this probabilistic approach makes it possible to calculate global score indicators combining both the mean multi-criteria values and spreads represented by the standard deviations. Note here that the approach would also be useful in the case of one profile only corresponding to a unique criterion, i.e. a general preference judgment or vote on some alternatives, candidates, etc. In such cases, the resulting indicator used for the final ranking gives a much richer content than the sole median value used for instance in the 'majority judgment' approach in voting theory [8].

Future work will be dedicated to looking for multi-criteria ranking problems of the same category for applying PROSE. In the 2014 REF, profiles and criterion weights are already available, without the need to address the whole preparation procedure. In other ranking problems, profiles will have to be elaborated from scratch, so that procedures must be established to achieve these results. Protocols such as those in [20] or [43] could be developed to assist in the cognitive task of the decision makers in elaborating profiles.

PROSE could also be extended to the use of importance profiles for the criterion weights. In the 2014 REF, the expert panel fixed the weights, making further analyses on our part unnecessary. In future work addressing similar ranking problems, the weighting of criteria could be handled with weight importance profiles, i.e. considering importance score levels from 'no importance' to 'greatest importance' as in [16, 28], in analogy with the performance score levels in this article. In such approaches, importance values would again be modelled by means of order statistics. Interesting comparisons can be made with other probabilistic approaches, such as the rank-order centroid (ROC) weights used in the SMARTER method [17] and the SMAA centroid weights [32, 33]. In dealing with weights, new issues are raised, such as the independence and non-interactivity of criteria. This may require more research on possible network structures, as in the analytic network process [45] or the Choquet integral [15], taking into account dependencies between importance profiles by adding covariance terms in the ranking indicator calculations used in PROSE. Finally, different alternate ways to treat the problem of ranking profile distribution may be proposed and compared in the future.

Acknowledgement
The authors wish to thank the University of Portsmouth for financially supporting this research.

Appendix A

Table A.1: Ranking of Computer Science and Informatics research units

Ranks	INSTITUTION CS&I Dept.	V_C
1	University College London	2.82
2	University of Warwick	2.77
3	Imperial College London	2.70
4	University of Manchester	2.59
5	University of Sheffield	2.53
6	University of Liverpool	2.50
7	University of York	2.46
8	University of Cambridge	2.45
9	Newcastle University	2.42
10	Queen Mary University of London	2.39
11	Lancaster University	2.36
12	University of Oxford	2.31
13	University of Nottingham	2.31
14	King's College London	2.29
15	University of Bristol	2.23
16	University of Leeds	2.22
17	University of Glasgow	2.20
18	University of Birmingham	2.16
19	Swansea University	2.14
20	University of Edinburgh	2.14
21	University of Dundee	2.13
22	University of Southampton	2.11
23	University of East Anglia	2.11
24	University of Durham	2.06
25	Cardiff University	2.05

26	University of St Andrews	2.01
27	Royal Holloway, University of London	2.00
28	University of Plymouth	1.99
29	University of Kent	1.99
30	University of Bath	1.98
31	University of Aberdeen	1.96
32	University of Exeter	1.95
33	Heriot-Watt University	1.94
34	University of Sussex	1.92
35	Aberystwyth University	1.91
36	Open University	1.89
37	University of Essex	1.82
38	Queen's University Belfast	1.82
39	Goldsmiths' College	1.74
40	University of Portsmouth	1.73
41	Aston University	1.73
42	City University London	1.70
43	University of the West of England, Bristol	1.68
44	University of Leicester	1.67
45	Brunel University London	1.62
46	University of Ulster	1.62
47	Birkbeck College	1.61
48	University of Surrey	1.59
49	De Montfort University	1.58
50	Kingston University	1.57
51	University of Hull	1.56
52	University of Stirling	1.54
53	University of Strathclyde	1.48
54	Teesside University	1.48
55	Edinburgh Napier University	1.46
56	University of Brighton	1.45
57	Loughborough University	1.45
58	Coventry University	1.40
59	Liverpool John Moores University	1.34
60	Keele University	1.31
61	University of Hertfordshire	1.30
62	University of Salford	1.24
63	University of Northumbria at Newcastle	1.24
64	University of Bedfordshire	1.24
65	Manchester Metropolitan University	1.23
66	Oxford Brookes University	1.22
67	Glasgow Caledonian University	1.21
68	University of Huddersfield	1.19

69	University of South Wales	1.14
70	Bangor University	1.12
71	Middlesex University	1.09
72	Liverpool Hope University	1.01
73	University of the West of Scotland	1.00
74	Nottingham Trent University	0.98
75	University of Greenwich	0.96
76	Robert Gordon University	0.95
77	University of Westminster	0.93
78	University of East London	0.92
79	Glyndŵr University	0.92
80	University of Lincoln	0.85
81	University of Derby	0.74
82	Birmingham City University	0.74
83	Staffordshire University	0.71
84	University of Sunderland	0.62
85	Leeds Beckett University	0.51
86	The University of West London	0.46
87	London Metropolitan University	0.44
88	Edge Hill University	0.28
89	University of Chester	0.17

Appendix B

Table B.1: Ranking of all UK institutions across all research units

RANKS	INSTITUTION	V_C
1	Courtauld Institute of Art	2.64
2	Institute of Cancer Research	2.52
3	Imperial College London	2.47
4	University of Oxford	2.40
5	London School of Economics and Political Science	2.40
6	University of Cambridge	2.39
7	Cardiff University	2.33
8	University of Warwick	2.27
9	King's College London	2.26
10	Institute of Zoology	2.23
11	Queen Mary University of London	2.22
12	University of Sheffield	2.21
13	University of Bath	2.21
14	University College London	2.20
15	University of Bristol	2.19
16	London School of Hygiene & Tropical Medicine	2.19
17	University of York	2.19
18	University of Edinburgh	2.18
19	London Business School	2.17
20	University of Southampton	2.17
21	University of Manchester	2.16
22	University of Durham	2.15
23	University of St Andrews	2.15
24	University of Leeds	2.14
25	Lancaster University	2.14
26	University of the Arts, London	2.14
27	University of East Anglia	2.13
28	Liverpool School of Tropical Medicine	2.11
29	Cranfield University	2.10
30	University of Glasgow	2.09
31	Swansea University	2.09
32	Royal Holloway, University of London	2.08
33	Royal Veterinary College	2.08
34	University of Exeter	2.08
35	Newcastle University	2.08
36	University of Liverpool	2.07
37	Heriot-Watt University	2.06
38	University of Birmingham	2.06
39	University of Nottingham	2.06

40	SRUC	2.06
41	Aston University	2.03
42	University of Essex	2.03
43	Royal Northern College of Music	2.02
44	University of Reading	2.02
45	University of Strathclyde	2.01
46	Royal College of Art	1.99
47	Cardiff Metropolitan University	1.98
48	University of Surrey	1.98
49	University of Dundee	1.97
50	University of Sussex	1.96
51	Queen's University Belfast	1.96
52	Bangor University	1.95
53	St.George's, University of London	1.95
54	University of Aberdeen	1.93
55	University of Stirling	1.92
56	Loughborough University	1.90
57	Royal College of Music	1.89
58	University of Kent	1.89
59	Birkbeck College	1.89
60	University of Leicester	1.89
61	City University London	1.89
62	Royal Central School of Speech and Drama	1.89
63	University of Bradford	1.89
64	Open University	1.87
65	University of Ulster	1.83
66	Goldsmiths' College	1.82
67	Keele University	1.81
68	Aberystwyth University	1.74
69	Heythrop College	1.74
70	University of Brighton	1.73
71	Roehampton University	1.73
72	University of Wales	1.72
73	Royal Conservatoire of Scotland	1.70
74	Liverpool John Moores University	1.70
75	University of the Highlands and Islands	1.69
76	School of Oriental and African Studies	1.69
77	University of Portsmouth	1.67
78	Trinity Laban Conservatoire of Music and Dance	1.67
79	Manchester Metropolitan University	1.66
80	Sheffield Hallam University	1.66
81	University of Plymouth	1.64
82	Harper Adams University	1.63
83	Bournemouth University	1.62
84	Glasgow School of Art	1.61

85	University of Hull	1.61
86	University of Northumbria at Newcastle	1.61
87	Royal Academy of Music	1.60
88	Kingston University	1.60
89	University of Westminster	1.60
90	University of East London	1.59
91	University of the West of England, Bristol	1.59
92	Glasgow Caledonian University	1.58
93	University for the Creative Arts	1.58
94	Norwich University of the Arts	1.57
95	Coventry University	1.56
96	De Montfort University	1.55
97	Oxford Brookes University	1.55
98	Brunel University London	1.55
99	Queen Margaret University Edinburgh	1.52
100	University of Huddersfield	1.52
101	University of Hertfordshire	1.50
102	Birmingham City University	1.48
103	Teesside University	1.46
104	London South Bank University	1.43
105	Middlesex University	1.43
106	Nottingham Trent University	1.42
107	University of Central Lancashire	1.40
108	University of Salford	1.40
109	Edinburgh Napier University	1.38
110	University of Lincoln	1.37
111	University of Bedfordshire	1.35
112	University of South Wales	1.35
113	Bath Spa University	1.34
114	University of Chichester	1.32
115	London Metropolitan University	1.24
116	Canterbury Christ Church University	1.24
117	University of the West of Scotland	1.22
118	Anglia Ruskin University'	1.20
119	University of Gloucestershire	1.20
120	University of Wales Trinity Saint David	1.20
121	Robert Gordon University	1.19
122	University of Greenwich	1.14
123	University of Winchester	1.12
124	University of Wolverhampton	1.12
125	Arts University Bournemouth	1.12
126	Guildhall School of Music & Drama	1.06
127	Stranmillis University College	1.05
128	Rose Bruford College	1.03
129	Liverpool Hope University	1.01

130	Staffordshire University	1.01
131	Edge Hill University	1.00
132	Glyndŵr University	0.99
133	University of Abertay Dundee	0.98
134	University of Cumbria	0.97
135	Buckinghamshire New University	0.97
136	Newman University	0.96
137	Leeds Beckett University	0.93
138	University of Sunderland	0.92
139	University of Northampton	0.91
140	University of Derby	0.87
141	University of Chester	0.87
142	University of Worcester	0.87
143	Leeds Trinity University	0.81
144	Falmouth University	0.79
145	York St John University	0.79
146	University of Bolton	0.79
147	Bishop Grosseteste University	0.75
148	The University of West London	0.72
149	St Mary's University, Twickenham	0.67
150	Southampton Solent University	0.37
151	University of London Institute in Paris	0.30
152	Royal Agricultural University	0.22
153	St Mary's University College	0.13
154	Writtle College	0.02

References

[1] Results of REF 2014, Retrieved from REF 2014 website: http://results.ref.ac.uk/, 2014.
[2] G. Abramo, T. Cicero, A. D'Angelo, A field-standardized application of DEA to national-scale research assessment of universities, Journal of Informetrics, 5 (2011) 618-628.
[3] G. Abramo, A. D'Angelo, F. Pugini, The measurement of Italian universities' research productivity by a non parametric-bibliometric methodology, Scientometrics, 76 (2008) 225-244.
[4] T. Agasisti, G. Catalano, P. Landoni, R. Verganti, Evaluating the performance of academic departments: an analysis of research-related output efficiency, Research Evaluation, 21 (2012) 2-14.
[5] T. Agasisti, A. Dal Bianco, P. Landoni, A. Sala, M. Salerno, Evaluating the Efficiency of Research in Academic Departments: an Empirical Analysis in an Italian Region, Higher Education Quarterly, 65 (2011) 267-289.
[6] S. Aoki, K. Inoue, R. Gejima, Data envelopment analysis for evaluating Japanese universities, Artificial Life and Robotics, 15 (2010) 165-170.
[7] B. Arnold, N. Balakrishnan, N. Nagaraja, A First course in Order Statistics, John Wiley & Sons, Inc., New York, 1992.
[8] M. Balinski, R. Laraki, Majority Judgment: Measuring, Ranking and Electing, MIT Press, Boston, 2010.
[9] C. Bana e Costa, M. Oliveira, A multicriteria decision analysis model for faculty evaluation, Omega, 40 (2012) 424-436.

[10] F. Barron, Selecting a best multiattribute alternative with partial information about attribute weights, Acta Psychologica, 80 (1992) 91-103.
[11] E. Bayraktar, E. Tatoglu, S. Zaim, Measuring the relative efficiency of quality management practices in Turkish public and private universities, Journal of the Operational Research Society, 64 (2013) 1810–1830.
[12] M.-L. Bougnol, J. Dulá, Validating DEA as a ranking tool: An application of DEA to assess performance in higher education, Annals of Operations Research, 145 (2006) 339-365.
[13] R. Carrington, T. Coelli, P. Rao, The performance of Australian universities: conceptual issues and preliminary results, Economic Papers: A journal of applied economics and policy, 24 (2005) 145-163.
[14] J.-P. Chavas, B. Barham, J. Foltz, K. Kim, Analysis and decomposition of scope economies: R&D at US research universities, Applied Economics, 44 (2011) 1387-1404.
[15] S. Corrente, S. Greco, A. Ishizaka, Combining analytical hierarchy process and Choquet integral within non-additive robust ordinal regression, Omega, (2016) 2-18.
[16] M. Druzdel, Verbal Uncertainty Expressions: Literature Review, Carnegie Mellon University, Pittsburgh, 1989.
[17] W. Edwards, H. Barron, SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement, Organizational Behavior and Human Decision Processes, 60 (1994) 306-325.
[18] G. Fandel, On the performance of universities in North Rhine-Westphalia, Germany: Government's redistribution of funds judged using DEA efficiency measures, European Journal of Operational Research, 176 (2007) 521-533.
[19] C. Giannoulis, A. Ishizaka, A Web-based decision support system with ELECTRE III for a personalised ranking of British universities, Decision Support Systems, 48 (2010) 488-497.
[20] S. Hora, Eliciting Probabilities from Experts, in: R. Miles Jr, D. von Winterfeldt (Eds.) Advances in Decision Analysis From Foundations to Applications, Cambridge University Press, Cambridge, 2007, pp. 129-153.
[21] A. Ishizaka, P. Nemery, Multi-Criteria Decision Analysis, John Wiley & Sons Inc, Chichester (United Kingdom), 2013.
[22] A. Ishizaka, G. Resce, B. Mareschal, Visual management of performance with PROMETHEE productivity analysis, Soft Computing, (2017) advance online publication, doi: 10.1007/s00500-00017-02884-00500.
[23] G. Johnes, Efficiency in English higher education institutions revisited: a network approach, Economics Bulletin, 33 (2013) 2698-2706.
[24] J. Johnes, Data envelopment analysis and its application to the measurement of efficiency in higher education, Economics of Education Review 25 (2006) 273–288.
[25] J. Johnes, Measuring teaching efficiency in higher education: An application of data envelopment analysis to economics graduates from UK Universities 1993, European Journal of Operational Research, 174 (2006) 443-456.
[26] J. Johnes, G. Johnes, Research funding and performance in U.K. University Departments of Economics: A frontier analysis, Economics of Education Review, 14 (1995) 301-314.
[27] L. Kao, C. Lu, C. Chiu, The training institution efficiency of the semiconductor institute programme in Taiwan-application of spatiotemporal ICA with DEA approach, Journal of the Operational Research Society, 62 (2011) 2162-2172.
[28] M. Karanik, S. Gramajo, L. Wanderer, M. Giménez, D. Carpintero, Multi-Criteria Decision Model based on AHP and Linguistic Information, Journal of Computer Sciences and Technology, 14 (2014) 16-24.
[29] M. Katharaki, G. Katharakis, A comparative assessment of Greek universities' efficiency using quantitative analysis, International Journal of Educational Research, 49 (2010) 115-128.
[30] G. Kempkes, C. Pohl, The efficiency of German universities–some evidence from nonparametric and parametric methods, Applied Economics, 42 (2010) 2063-2079.
[31] K. Kounetas, A. Anastasiou, P. Mitropoulos, I. Mitropoulos, Departmental efficiency differences within a Greek university: An application of a DEA and Tobit analysis, International Transactions in Operational Research, 18 (2011) 545-559.

[32] R. Lahdelma, J. Hokkanen, P. Salminen, SMAA - Stochastic multiobjective acceptability analysis, European Journal of Operational Research, 106 (1998) 137-143.
[33] R. Lahdelma, K. Miettinen, P. Salminen, Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA), European Journal of Operational Research, 147 (2003) 117-127.
[34] R. Lukman, D. Krajnc, P. Glavič, University ranking using research, educational and environmental indicators, Journal of Cleaner Production, 18 (2010) 619-628.
[35] P. Murias, J. de Miguel, D. Rodríguez, A Composite Indicator for University Quality Assesment: The Case of Spanish Higher Education System, Social Indicators Research, 89 (2008) 129-146.
[36] A. Mustafa, M. Goh, Multi-criterion models for higher education administration, Omega, 24 (1996) 167-178.
[37] J. Nazarko, J. Šaparauskas, Application of DEA method in efficiency evaluation of public higher education institutions, Technological and Economic Development of Economy, 20 (2014) 25-44.
[38] S. Nisel, An extended VIKOR method for ranking online graduate business programs, International Journal of Information and Education Technology, 4 (2014) 103-107.
[39] S. Nisel, R. Nisel, Using VIKOR Methodology for Ranking Universities by Academic Performance, Journal of Mathematics, Statistics and Operations Research, 2 (2013) 86-92.
[40] M. Pidd, J. Broadbent, Business and management studies in the 2014 Research Excellence Framework, British Journal of Management, 26 (2015) 569-581.
[41] Y. Politis, Y. Siskos, Multicriteria methodology for the evaluation of a Greek engineering department, European Journal of Operational Research, 156 (2004) 223-240.
[42] C. Rao, M. Goh, Y. Zhao, J. Zheng, Location selection of city logistics centers under sustainability, Transportation Research Part D: Transport and Environment, 36 (2015) 29-44.
[43] J. Rezaei, Best-worst multi-criteria decision-making method, Omega, 53 (2015) 49-57.
[44] T. Rosenmayer, Using Data Envelopment Analysis: A Case of Universities, Review of Economic Perspectives, 14 (2014) 34–54.
[45] T. Saaty, M. Takizawa, Dependence and Independence: from Linear Hierarchies to Nonlinear Networks, European Journal of Operational Research 26 (1986) 229-237.
[46] C. Sarrico, R. Dyson, Using DEA for planning in UK universities-an institutional perspective, Journal of the Operational Research Society, 51 (2000) 789-800.
[47] T. Sexton, C. Comunale, S. Gara, Efficiency-based funding for public four-year colleges and universities, Education Finance and Policy, 7 (2012) 331-359.
[48] M. Spiegel, Theory and Problems of Statistics, McGraw-Hill Book Company, New York, 1961.
[49] E. Thanassoulis, M. Kortelainen, G. Johnes, J. Johnes, Costs and efficiency of higher education institutions in England: a DEA analysis, Journal of the Operational Research Society, 62 (2011) 1282-1297.
[50] K. Tochkov, N. Nenovsky, K. Tochkov, University efficiency and public funding for higher education in Bulgaria, Post-Communist Economies, 24 (2012) 517-534.
[51] D. Turner, Benchmarking in universities: league tables revisited, Oxford Review of Education, 31 (2005) 353-371.
[52] H.-Y. Wu, J.-K. Chen, S. Chen, H.-H. Zhuo, Ranking universities based on performance evaluation by a hybrid MCDM model, Measurement, 45 (2012) 856-880.
[53] J.-W. Youn, K. Park, University development models and efficiency analysis, Journal of Service Science, 1 (2009) 9-30.