

## PAPER

# A Unified Statistical Rating Method for Team Ball Games and Its Application to Predictions in the Olympic Games

Eiji KONAKA<sup>†a)</sup>, *Member*

**SUMMARY** This study tries to construct an accurate ranking method for five team ball games at the Olympic Games. First, the study uses a statistical rating method for team ball games. A single parameter, called a rating, shows the strength and skill of each team. We assume that the difference between the rating values explains the scoring ratio in a match based on a logistic regression model. The rating values are estimated from the scores of major international competitions that are held before the Rio Olympic Games. The predictions at the Rio Olympic Games demonstrate that the proposed method can more accurately predict the match results than the official world rankings or world ranking points. The proposed method enabled 262 correct predictions out of 370 matches, whereas using the official world rankings resulted in only 238 correct predictions. This result shows a significant difference between the two criteria.

**key words:** sports, ball games, prediction model, rating and ranking method

## 1. Introduction

This study tries to construct an accurate ranking method for five team ball games (basketball, handball, hockey, volleyball, and water polo) at the Olympic Games.

Accurate ranking systems are important for players, event organizers, and sports enthusiasts. Players use rankings to evaluate and estimate their skill levels. Event organizers use rankings as a criterion in tournament design tasks such as group draws, player (team) seeding, guest player (team) selection, and so on. Sports enthusiasts use rankings to evaluate the skill of a team and to predict the results of matches. Inaccurate ranking systems confuse and disappoint event organizers, players, and enthusiasts by increasing the gap between predictions and match results. Therefore, accurate rankings aid in creating attractive and consistent sporting events.

The number of wins and the percentage of victories are the most “fair” ranking criteria if all players are matched in a round-robin format. However, a fair round robin is not possible when the number of teams participating is larger than the number of schedulable matches. In particular, the national teams of major sports cannot all compete in a fair round-robin format. As a result, teams have different opponents and play different numbers of matches.

To rank and order teams according to their abilities, the international association of each sport designs its own orig-

inal ranking system. The most popular ranking system is based on an accumulative method [1]. This system calculates *ranking points* for each team. Ranking points are calculated as the sum of the points attributed to international tournaments and the standings in the tournaments. The sum is calculated for a designated period, such as four years. The five ball games examined here determine their world rankings using this method [2]–[5]. The Fédération Internationale de Natation (FINA) does not disclose world rankings and ranking points for water polo on their website. Thus, the rankings and ranking points for water polo used here are collected from personal websites and sports news.

These ranking points have no clear mathematical or statistical basis, therefore, the ranking points do not directly measure of the scoring ability of the teams. For instance, Konaka [6] reports that the Fédération Internationale de Volleyball (FIVB) ranking points have many problems as quantitative measure of teams’ skill owing to their inconsistent design.

A points-exchange is another possible ranking system. Here, each team has a ranking point, which they exchange based on match results. For example, several points may be moved from the losing team to the winning team after a match. The most popular points-exchange system is the Elo rating [7] used in chess ranking. Rugby also uses a modified Elo-based ranking system [8]. In these systems, the calculated ranking points converge to the real values if the abilities of all teams are constant and a sufficiently large number of matches are played within a certain period. In general, ranking points in a points-exchange system require more calculation than those in accumulative points systems.

### 1.1 Ranking and Rating

Here, we define *ranking* and *rating* as follows:

- *ranking*: the order of teams.
- *rating*: a quantitative value associated with the ability of each team.

The objective of this study is to create a ranking based on ratings.

Assume that the following two elements affect the result of a match:

1. the stable and constant skill and ability of each team.
2. condition, form, luck, and other unstable and non-constant elements.

Manuscript received September 18, 2018.

Manuscript revised December 17, 2018.

Manuscript publicized March 11, 2019.

<sup>†</sup>The author is with the Dept. of Information Engineering, Meijo University, Nagoya-shi, 468–8502 Japan.

a) E-mail: konaka@meijo-u.ac.jp

DOI: 10.1587/transinf.2018EDP7315

The ranking points in the accumulative method include both sets of elements. On the other hand, a point-exchange system estimates the first set of elements by denoising the effects of the second set. In this study, the rating is a quantitative value calculated using a statistical method based on the first set of elements.

## 1.2 Sports Analysis as an Information System

Sports-analysis systems are increasingly being viewed as information systems, including sensing and statistical analyses. Two different approaches, specific and unified, are necessary to construct a sports-analysis system.

Many statistical skill-assessment studies have been conducted for various sports. For instance, volleyball studies have examined how elementary techniques (service, reception, spike, dig, block, and set) and strategies contribute to scores and wins [9]–[14].

The Association for Professional Basketball (APBR) [15], established in 1997, analyzes basketball using objective evidence.

Detailed and sport-specific skill assessments and analyses are assumed to improve the skills of players or the tactics of a team. These analyses require complex information-processing systems, including video, wearable medical sensors, and so on. The construction and information-processing costs of such systems cannot be ignored.

The rating system proposed here uses only the scores of matches and has a light computation cost. This approach tries to construct a unified evaluation method for different sports. By comparing the actual and predicted results, players/teams can seek to improve their performance. In addition, as described in the previous section, the proposed method can replace conventional sports-specific ranking systems. Figure 1 summarizes this section.

## 1.3 Objective

As mentioned above, few studies have examined quantitative ability-evaluation methods for national ball game teams. In particular, there is no unified prediction model reported for different ball games.

The main objective and contribution of this study is to use a simple and unified rating framework for different ball games, and to evaluate its prediction performance. The unified method should use only commonly recorded values among the different sports. All five ball games considered here have a common value: a score. A single parameter, called a rating, shows the strength and ability of each team. We assume that the difference between the rating values explains the scoring ratio in a game, based on a logistic regression model. The rating values are estimated from major international competition results, including those of world championships, worldwide league competitions, and the Olympic continental and world qualifying tournaments held before the Rio Olympic Games.

The results of these ball games in the Rio Olympic

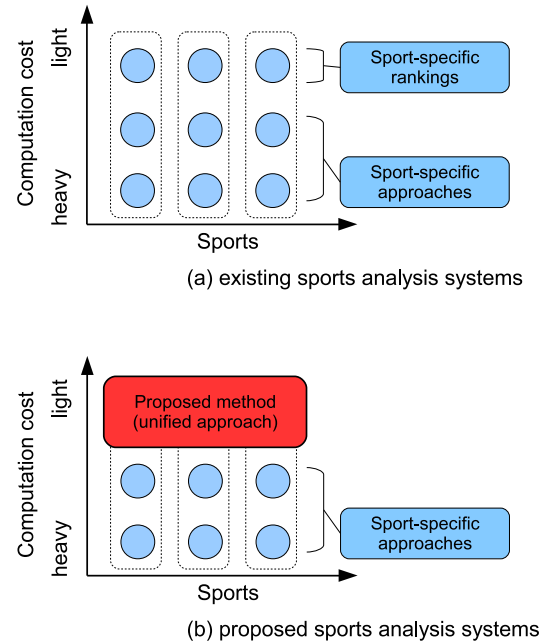


Fig. 1 Sports analysis systems

Games are estimated based on the calculated rating values. The prediction results demonstrate that using the proposed method can more accurately predict a result than when using official world rankings or world ranking points. The prediction method correctly predicted 262 out of 370 matches in 10 events, whereas the official world rankings made only 238 correct predictions. This result shows a significant difference between the two criteria. The method also correctly predicted 10 out of 30 medals, together with their medal colors (33.3%). Moreover, we made 19 correct predictions of podium finishes (63.3%). These prediction results are better than those provided by Sports Illustrated (26.7%, 53.3%), USA Today (23.3%, 46.7%), and Gracenote (33.3%, 46.7%). Note that these results do not show any statistically significant difference because the samples are too small.

This method can be utilized to evaluate the inherent prediction difficulties for each event, and to compare the randomness between sports. This problem is discussed in Lundh [16]. In this study, a “tournament stability index” is calculated from the match results of the evaluation target tournament (e.g., Olympic Games) to quantify the randomness and uncertainty for different tournaments and sports. In contrast to conventional works, this study proposes an index that evaluates the skill distribution of the participating teams before the evaluation target tournament.

## 2. Definition and Calculation of Rating

### 2.1 Current Ranking Systems

The FIVB, the world governing body for volleyball, regularly reports the rankings of its member nations’ teams. The

FIVB Board of Administration has designed a system of point attribution for selected FIVB world and other official competitions [5].

The design shows significant inconsistencies. For instance, there are no clear mathematical and statistical basis on the following point attribution designs.

- The champions of several competitions each awarded equally 100 points.
- The differences in points between standings.
- The continental championships all awarded the same ranking points.

Basketball [2], handball [3], and hockey [4] have similar accumulative ranking systems, essentially based on the standings in international competitions, but do not explain the mathematical fundamentals of the systems. In fact, as of 2016, FINA no longer even discloses the world rankings for water polo.

## 2.2 Proposed Method

As mentioned above, official ranking points do not directly measure scoring ability of each team.

We propose a unified and simple statistical estimation method of scoring ratios based on the score in each match, which is always officially recorded and is common to all ball games.

Assume that the scoring ratio of team  $i$  in a match against team  $j$  ( $i$  and  $j$  are team indices), denoted as  $p_{i,j}$ , is estimated as

$$p_{i,j} = \frac{1}{1 + e^{-(r_i - r_j)}}, \quad (1)$$

where  $r_i$  is defined as the *rating* of team  $i$ . Given  $(s_i, s_j)$ , the actual scores in a match between  $i$  and  $j$ ,

$$s_{i,j} = \frac{s_i}{s_i + s_j} = p_{i,j} + \epsilon_{i,j}, \quad (2)$$

where  $s_{i,j}$  and  $\epsilon_{i,j}$  are the actual scoring ratio and the estimation error, respectively.

This mathematical structure is the well-known *logistic regression model*. It is widely used in areas such as the winning probability assumption of Elo ratings in chess games [7], and the correct answer probability for questions in item response theory [17], [18].

The update method is designed to minimize the sum of the squared error between the result and the prediction  $E^2$ , defined by the following equation:

$$E^2 = \sum_{(i,j) \in \text{all matches}} (s_{i,j} - p_{i,j})^2. \quad (3)$$

It is straightforward to obtain the following update based on the steepest-descent method:

$$r_i \leftarrow r_i - \alpha \cdot \frac{\partial E^2}{\partial r_i}, \quad (4)$$

where  $\alpha$  is a constant.

Of these five sports, hockey matches have the lowest scores. Shut-out results such as 1 – 0 or 3 – 0 occur frequently. Thus, a simple scoring ratio can result in invalid skill evaluation. Therefore, for hockey, the scoring ratio is modified to

$$s_{i,j} = \frac{s_i + 1}{(s_i + 1) + (s_j + 1)} = \frac{s_i + 1}{s_i + s_j + 2}. \quad (5)$$

This modification is known as Colley's method [19], and was originally used to rank college football teams.

By definition, the rating is an interval scale. Therefore, its origin,  $r = 0$ , can be selected arbitrarily and a constant value can be added to all  $r_i$ . For example,

$$r \leftarrow r - (\max r) \cdot 1 \quad (6)$$

implies that  $r = 0$  always shows the highest rating, and  $r < 0$  shows the distance from the top team.

### 2.2.1 Convert Rating on Scoring Ratio to Winning Probability

The rating  $r_i$  in (1) explains the scoring ratio. This differs between sports in terms of showing how the scoring ratio affects the winning probability.

Once we have the scoring ratio  $p_{i,j}$  given in (1), assume that the following independent Bernoulli process is executed  $N$  times, starting from  $(s_i, s_j) = (0, 0)$  and with the parameter  $0 < \beta \leq 1$ .

$$\begin{cases} s_i \leftarrow s_i + 1 & \text{with probability } \beta p_{i,j}, \\ s_j \leftarrow s_j + 1 & \text{with probability } \beta (1 - p_{i,j}), \\ s_i \leftarrow s_i, s_j \leftarrow s_j & \text{with probability } (1 - \beta). \end{cases} \quad (7)$$

This is a unified (and approximated) model of a scoring process for five different ball games, where  $s_i$  and  $s_j$  model the scores of teams  $i$  and  $j$ , respectively. By definition,  $E(s_i + s_j) = N\beta$ ,  $E(s_i) = N\beta p_{i,j}$ , and  $E(s_i / (s_i + s_j)) = p_{i,j}$ .

The parameters  $N$  and  $\beta$  vary among the sports and between definitions of a unit of play. For example, in volleyball, the only one net sport of the five sports, a unit of play is defined from service to scoring. Under this definition,  $\beta = 1$  and  $N \simeq 45$  in one set of a volleyball match. The other four sports are goal sports with different durations. In these sports, a unit of play is a short period of time. For example, in basketball, if a unit of play is defined as 10[s], we have  $N = 40[\text{min}] \times 60[\text{s}] / 10 = 240$ .  $\beta$  is determined as  $\beta = E(s_i + s_j) / N$ .

At the end of the match,  $s_i > s_j$  shows that team  $i$  wins against team  $j$ . Figure 2 shows the simulated winning probability for different  $N\beta$  and rating gap  $(r_i - r_j)$ , with  $N = 240$ . This probability is expressed by the cumulative distribution function for a normal distribution. In many applications, it is common to use a logistic regression model rather than a cumulative distribution [20].

Based on the discussions above, we convert the rating on the scoring ratio to that of a winning probability, as follows:

$$w_{i,j} = 1 \text{ (} i \text{ wins), or } 0 \text{ (} j \text{ wins),} \quad (8)$$

which denotes a win or loss for team  $i$  against team  $j$ . Find  $D_k^*$ , where  $k$  is an index of sports, that satisfies

$$\hat{w}_{i,j} = \frac{1}{1 + \exp(-D_k(r_i - r_j))}, \quad (9)$$

$$D_k^* = \arg \min_{D_k} \sum (w_{i,j} - \hat{w}_{i,j})^2. \quad (10)$$

Then,  $r_i$  is converted as follows:

$$\bar{r}_i = D_k^* r_i, \quad i = 1, 2, \dots, N_T. \quad (11)$$

Therefore,  $\bar{r}_i$  is a rating that explains the winning probability, and it can be utilized in match result predictions.

In Eqs. (3) and (10), the sum of squared errors is used as a loss function instead of the cross-entropy. This is because these problems are regression problems, not classification ones.

### 2.3 Event Competitiveness Measured by Entropy Functions

Once the estimated winning probability  $\hat{w}$  is calculated, the following binary entropy function  $I_{TC}$  can be used to evaluate the distribution of the competitive strength of the teams participating in an event:

$$I_{TC} = -\frac{1}{N} \sum (\hat{w}_{i,j} \log_2 \hat{w}_{i,j} + (1 - \hat{w}_{i,j}) \log_2 (1 - \hat{w}_{i,j})), \quad (12)$$

where  $(i, j)$  is taken from a set of match-ups of the event and  $N$  is the number of matches in the event. By definition,  $I_{TC} \in [0, 1]$ . Here,  $I_{TC} = 1$  implies that all teams have equal strength, that is  $\hat{w}_{i,j} = 0.5$  for all matches. On the other hand,

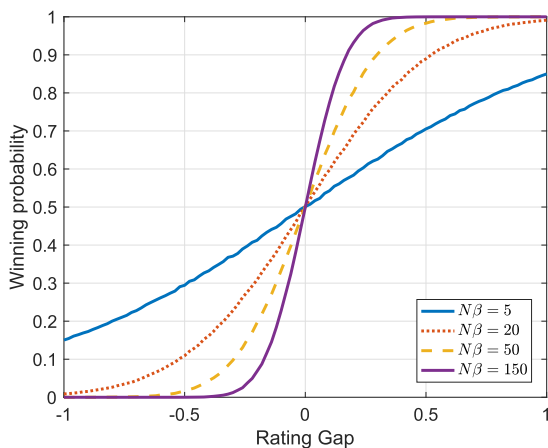


Fig. 2 Rating gap and winning probability

a small  $I_{TC}$  implies that the skill gaps between the teams are large and many one-sided games are included in the event.

### 3. Rating Calculation for Five Ball Games and Its Application to Match Result Predictions in the Rio Olympic Games

#### 3.1 Data Set

We calculate the rating values for the national teams of the following five ball games: basketball, handball, hockey, volleyball, and water polo. The match results used in the rating calculation include the following:

- Rio Olympics qualifying tournaments, including continental championships.
- Worldwide tournaments: for example, world championships, and the World League (men's volleyball), held from 2014 to 2016/8 (just before Rio 2016).

The number of teams participating in at least one tournament and the number of matches in the data set are listed in Table 1.

The following official world rankings are also used in the discussion:

- Basketball: FIBA ranking, 2016/7.
- Handball: IHF ranking, 2016/7.
- Hockey: FIH ranking, 2016/6.
- Volleyball: FIVB ranking, 2016/7.
- Water polo: FINA ranking, 2014/8.

#### 3.2 Results

Figure 3 shows the results of all 38 matches (30 group round-robin matches, four quarterfinals, two semifinals, and two medal matches) of the men's basketball in Rio 2016. The horizontal and vertical axes show the predicted scoring ratio from the calculated rating values and the real scoring ratio, respectively.

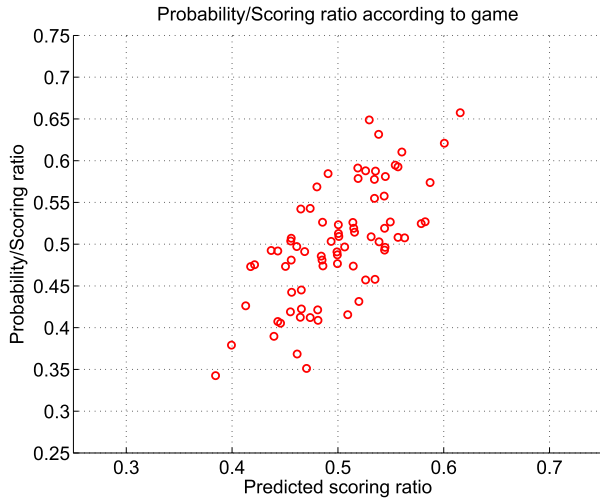
As a comparison, in Fig. 4, the horizontal axis now shows the difference in the official world rankings.

Figure 5 shows the relation between the FIBA ranking points for men (horizontal axis) and the proposed normalized rating (vertical axis) just before the Rio Olympic

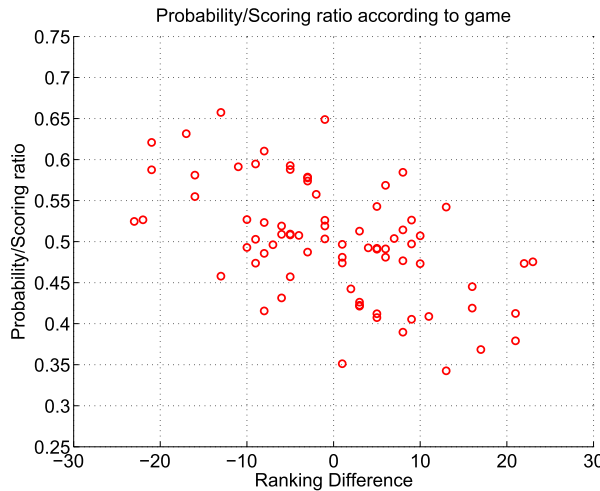
Table 1 Number of teams and matches

Sport	Sex	Teams	Matches
Basketball	M	69	334
Handball	M	69	375
Hockey	M	48	280
Volleyball	M	43	466
Water polo	M	31	346
Sport	Sex	Teams	Matches
Basketball	W	57	238
Handball	W	44	311
Hockey	W	42	265
Volleyball	W	36	337
Water polo	W	26	294

Games. Spearman's rank correlation coefficient  $\rho_k$  between the official ranking points and the proposed normalized rating is calculated for ten events. The values are listed in Table 2.



**Fig. 3** Predicted and real scoring ratio in each game (Rio 2016, basketball, men)

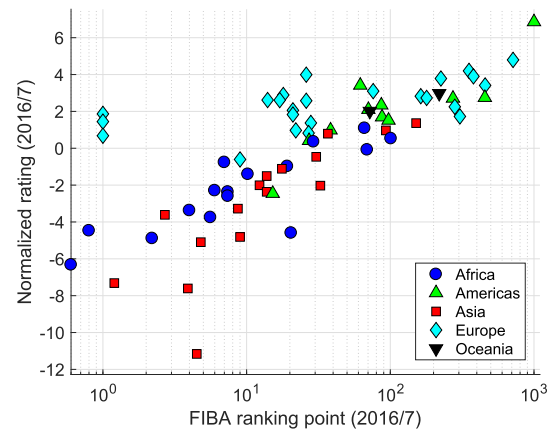


**Fig. 4** Ranking gap and real scoring ratio in each game (Rio 2016, basketball, men)

ble 2.

Table 3 compares the prediction accuracies of the proposed method and the official world rankings. The prediction law is simple: “a team with a higher rating (ranking) scores more.” Draws are judged as incorrect in both methods. The column “Corr. Coeff.” lists the following values:

- Rating: the correlation coefficient between the scoring ratio and the predicted scoring ratio from the rating gap
- Ranking: the correlation coefficient between the scoring ratio and the ranking gap
- Ideal: the correlation coefficient between the scoring ratio and the predicted scoring ratio from the ideal rating gap. “Ideal rating”, denoted as  $r_{ideal}$ , refers to rating values calculated from the actual results of the Rio Olympic Games. The ideal (i.e., maximum) num-



**Fig. 5** FIBA ranking points (men, 2016/7) and proposed rating

**Table 2** Spearman's rank correlation coefficient between the official ranking points and the proposed normalized rating

		Sex	$\rho_k$			Sex	$\rho_k$
Basketball	M		0.7557	Basketball	W		0.7225
Handball	M		0.5412	Handball	W		0.6094
Hockey	M		0.9710	Hockey	W		0.9646
Volleyball	M		0.7165	Volleyball	W		0.8873
Water polo	M		0.8215	Water polo	W		0.7991

**Table 3** Prediction accuracy in Rio 2016

		Correct			Corr. Coeff.			$I_{TC}$	
		Matches	Rating	Ranking	Ideal	Rating	Ranking		Ideal
Basketball	M	38	<b>30</b>	29	32	<b>0.679</b>	-0.542	0.878	0.6082
Handball	M	38	<b>25</b>	20	30	<b>0.592</b>	-0.492	0.654	0.7927
Hockey	M	38	21	21	30	0.725	<b>-0.729</b>	0.863	0.4847
Volleyball	M	38	<b>30</b>	27	32	0.731	<b>-0.790</b>	0.853	0.6124
Water polo	M	42	<b>27</b>	20	32	<b>0.560</b>	-0.438	0.644	0.6797
Basketball	W	38	<b>33</b>	28	36	<b>0.818</b>	-0.698	0.902	0.4950
Handball	W	38	22	<b>30</b>	33	<b>0.579</b>	-0.572	0.785	0.6785
Hockey	W	38	<b>21</b>	18	31	<b>0.764</b>	-0.608	0.847	0.6380
Volleyball	W	38	<b>34</b>	31	36	<b>0.731</b>	-0.663	0.900	0.5571
Water polo	W	24	<b>19</b>	14	22	<b>0.905</b>	-0.697	0.926	0.5716
All	M	194	<b>133</b>	117	156				
All	W	176	<b>129</b>	121	158				
All		370	<b>262</b>	238	314				

**bold:** better performance



ber of correct predictions is also listed in the column “Correct”-“Ideal.”

This table also lists  $I_{TC}$  defined in Sect. 2.3.

Table 4 lists the normalization parameters  $D_k^*$ .

Table 5 lists the detailed predictions for the men’s basketball. The rating values are normalized using  $D_k^*$ , and are shifted so that the lowest rating is zero. All 38 matches are simulated  $10^6$  times. The table lists the average values. The underlined and bold numbers denote the prediction and the result, respectively.

The teams winning medals are predicted for 10 events in five sports. The prediction is evaluated from two viewpoints, “Medal with color” and “Podium finishes.” For example, the prediction in Table 5 tells us that the gold, silver, and bronze medals would have been awarded to USA, ESP, and SRB, respectively. The actual result is USA, SRB, and ESP. In this case, the proposed method predicts one medal with color and three podium finishes.

The proposed prediction result is compared to the predictions seen in

- Official rankings,
- Sports Illustrated (SI) [21],
- USA Today [22], and
- Gracenote [23].

Table 6 shows the results. Bold numbers show the most accurate prediction.

Figures 6 and 7 show the rating distributions in 10 events in five sports. All teams participating in at least one match in the data set are included in these figures. The rating values are normalized by  $D_k^*$ . Figure 8 shows the normalized rating of the qualified teams for Rio 2016. In these figures, the rating values are shifted so that the top-rated team

**Table 4** Normalization parameters  $D_k^*$

	Sex	$D_k^*$		Sex	$D_k^*$
Basketball	M	11.660	Basketball	W	9.193
Handball	M	12.299	Handball	W	9.090
Hockey	M	4.509	Hockey	W	3.463
Volleyball	M	15.019	Volleyball	W	9.868
Water polo	M	5.288	Water polo	W	4.055

**Table 5** Medal prediction (basketball, men)

Team	Rating (normalized)	Group	Gold	Silver	Bronze	4th
FRA	3.3714	A	0.0236	0.1281	0.1692	0.1849
USA	5.9376	A	<b>0.7933</b>	0.1077	0.0695	0.0060
VEN	0.7395	A	0.0000	0.0001	0.0003	0.0019
SRB	3.6863	A	0.0453	<b>0.2225</b>	<u>0.2658</u>	0.1836
CHN	0.0000	A	0.0000	0.0000	0.0000	0.0004
AUS	3.3972	A	0.0248	0.1272	0.2097	<b>0.2392</b>
ARG	2.1873	B	0.0008	0.0112	0.0143	0.0522
ESP	4.3037	B	0.1081	<u>0.3611</u>	<b>0.1796</b>	0.0656
BRA	2.2445	B	0.0007	0.0122	0.0266	0.0851
LTU	2.3811	B	0.0024	0.0173	0.0419	0.1135
CRO	2.1984	B	0.0010	0.0125	0.0229	0.0659
NGR	0.9001	B	0.0000	0.0001	0.0002	0.0017

underline: prediction, **bold**: result

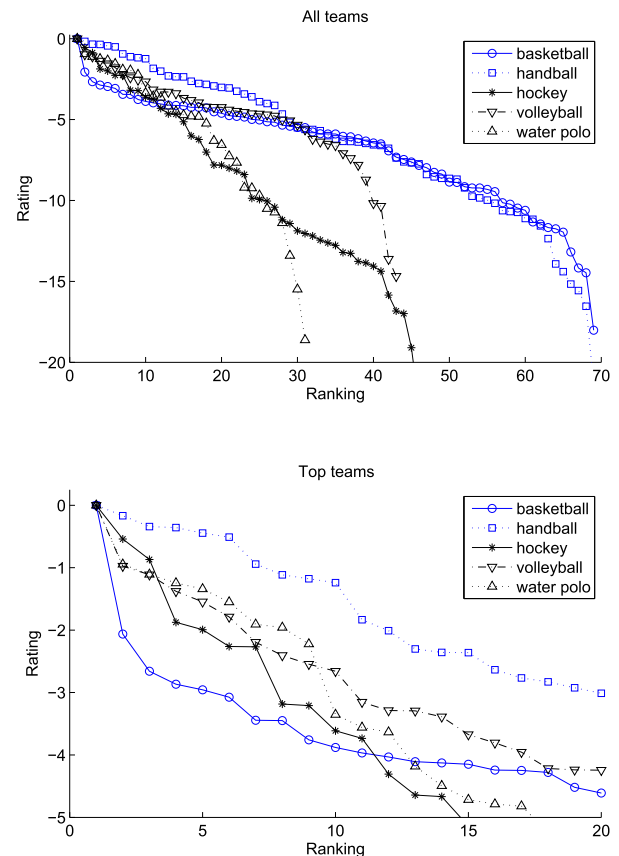
is zero.

Figure 9 shows another view of the ability distribution in Rio 2016. This figure shows the distribution of the pre-

**Table 6** Medal predictions

	All medals	Medal with color	Podium finishes
Proposed	30	<b>10</b>	<b>19</b>
Official Rankings	30	6	14
SI	30	8	16
USA Today	30	7	14
Gracenote	30	<b>10</b>	14

**bold**: best prediction



**Fig. 6** Normalized rating of five sports (men)

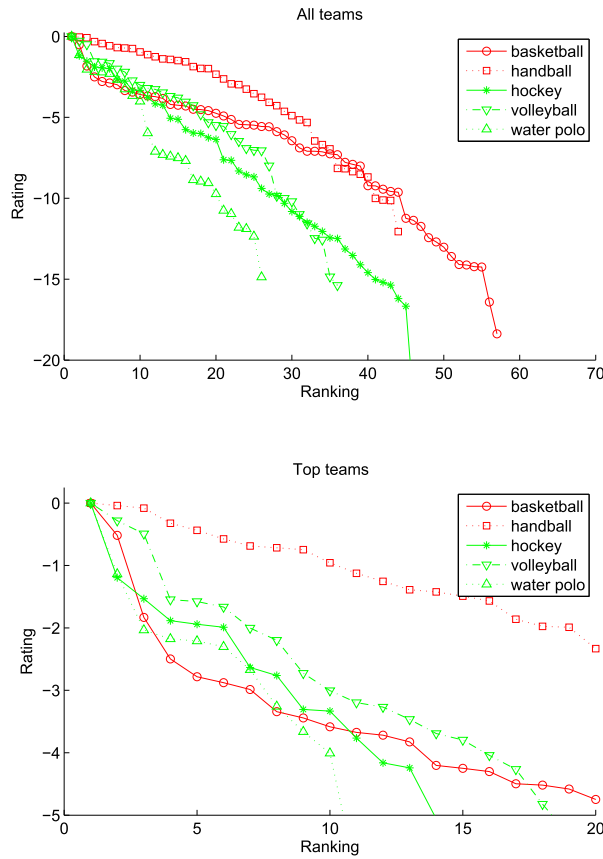


Fig. 7 Normalized rating of five sports (women)

dicted winning probability of the highly rated teams for every match in 10 events.

### 3.3 Discussion

Figure 5 and Table 2 show that the FIBA ranking cannot accurately measure the scoring skill for each team. For example, some European teams (indicated by diamond markers) with similar ratings (approximately 3.0) have very different ranking points (ranging from almost zero to 500). On the other hand, teams with very few ranking points around zero are evaluated as totally different scoring skill (from  $-11.0$  to  $2.0$ ). Spearman's rank correlation implies that what the ranking measures depends on their design. For instance, it seems that the IHF ranking for handball measures something other than scoring skills.

Table 3 shows that the proposed rating method realizes a more accurate prediction (262 correct out of 370 matches, 70.8%) than that using the official (accumulative) world ranking system (238 correct out of 370 matches, 64.3%).

Table 7 classifies the prediction results by the proposed rating and the official ranking. The null hypothesis that “the prediction accuracy of the proposed method is the same as that of the official world ranking system” is rejected by McNemar's  $\chi^2$  test with  $p = 6.0 \times 10^{-3} < 0.01$ . The script written by Cardillo [24] is used to obtain the  $p$ -value.

Moreover, the correlation between the predicted and

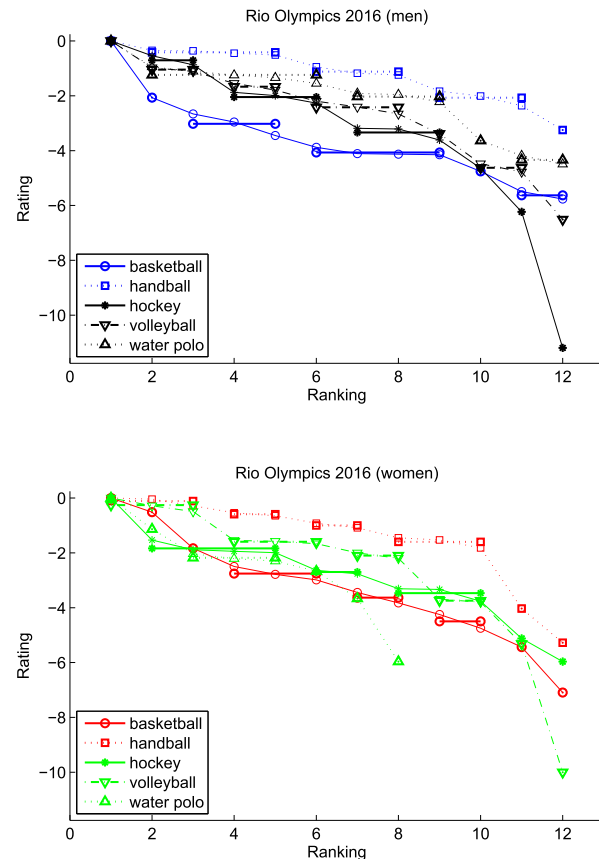


Fig. 8 Clustering result of normalized rating of five sports in Rio 2016 for qualified teams

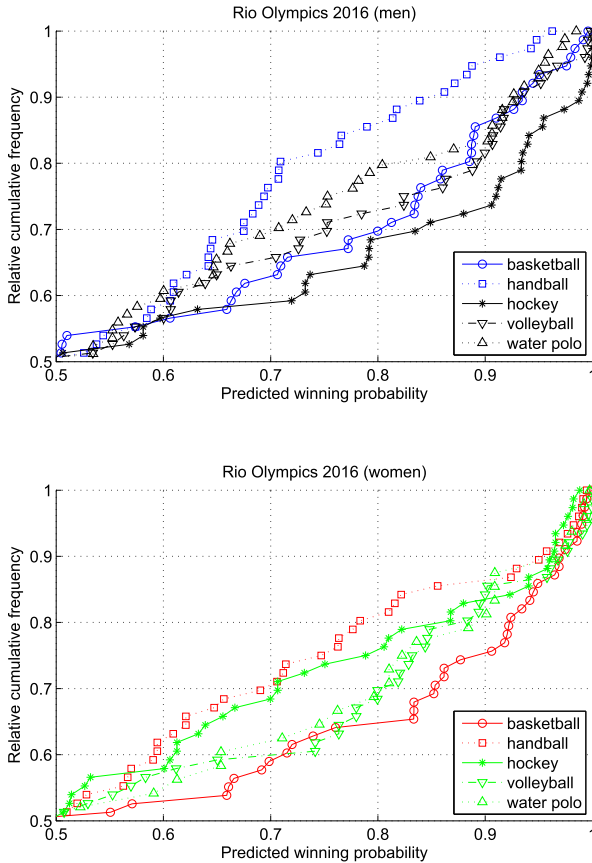
Table 7 Classification table

		Ranking		
		Correct	Incorrect	
Rating	Correct	215	47	262
	Incorrect	23	85	108
		238	132	370

the real scoring ratios is stronger than that between the ranking gap and the ratio. This result implies that the proposed rating value is a better quantitative measure of the ability of national teams of these five ball games than the official world ranking.

Table 4 shows that  $D_k^*$  is larger in men's events than in women's events in the same sport.  $D_k^*$  is a parameter used to convert the rating on the scoring ratio to a rating on the winning probability. A large  $D^*$  implies that many men's teams are equally matched and that many matches are closely contested; that is, the scoring ratio is around 0.5. Table 3 also shows that the official ranking system does not provide accurate ability evaluations, especially for men's competitions.

Table 6 shows that the proposed method provides better predictions than those provided by the official world rankings, a well-known sports magazine (Sports Illustrated), and a nationwide newspaper (USA Today). These are compared with the statistics provided by a company (Gracenote). However, the advantage of the proposed method for medal



**Fig. 9** Distribution of predicted winning probability of highly rated teams

predictions cannot be tested statistically because of the small sample sizes.

Surprisingly, the proposed method achieves better prediction results than those of the official ranking system and professional sports journalists, even though the proposed method uses one unified model and does not include features specific to each sport and event.

Figure 8 shows the normalized rating values of the probability of winning for the qualified teams. The rating values can be compared between different sports because they are normalized. These figures and the prediction results imply the following:

- It is difficult to predict the results of hockey because the matches have low scores (4.973 and 3.395 goals per match in men's and women's competitions, respectively). In other words, the ability gap between two teams are rarely reflected in the actual score and score the difference.
  - The low scores in hockey matches lead to frequent draws. In Rio 2016, six games resulted in a draw in each of the men's and women's events.
- In handball, there is no clearly strongest team. Six teams with  $\bar{r} > -1$  qualified for Rio 2016 in both the men's and the women's events. Therefore, it is

difficult to predict the match results ( $\bar{r} = -1$  implies that the team beats the top-rated team with probability  $1/(1 + e^1) = 0.2689$ ). As a result, the prediction accuracy of the proposed method was not good.

- The other four sports have one to three outstanding teams (i.e.,  $\bar{r} > -1$ ).
- Except for the abovementioned outstanding teams, the slope of the plot of the men's rating is more moderate than that of the women's rating. This implies that there are many equally matched teams in the men's event. In the women's event, match results tend to follow the match previews because there are clear differences in the abilities of the teams. Therefore, the prediction accuracy for the women's event (73.3%, 129 correct out of 176 matches) is higher than that for the men's event (68.6%, 133 correct out of 194 matches).

Figure 9 can also be used to evaluate the competitiveness of each event. In this figure, an event is competitive if the corresponding plot lies in the upper-left section of the graph (e.g., men's handball;  $I_C = 0.7927$  is the largest value among the men's events). On the other hand, if the plot lies in the bottom-right section, then the corresponding event had many one-sided games (e.g., women's basketball;  $I_C = 0.4975$  is the smallest value among the women's events).

#### 4. Conclusion

This paper has presented the prediction results of five ball games, namely, basketball, handball, hockey, volleyball, and water polo, in the Rio Olympic Games based on a unified statistical rating method. Both a unified rating method and its calculation method are proposed. The rating values for all teams participating in Olympic qualification tournaments within one or two years are calculated.

Surprisingly, the proposed method achieves better prediction results than the official ranking system and professional sports journalists, even though the proposed method uses a unified model and does not include features specific to each sport and event.

Future work will extend the proposed framework to other sports in upcoming Olympic Games, especially Tokyo 2020. The proposed method can be applied to sports involving individuals, not only team events. For example, badminton, fencing, judo, table tennis, and wrestling could be covered by the proposed method because worldwide competitions with top players are held regularly in these sports. On the other hand, soccer and baseball are difficult to predict using the proposed method. In the Olympic Games, soccer has a different age restriction (players should be younger than 23) to that of standard international A-matches. In the case of international baseball, there are too few competitions. Thus, the skill of the national teams cannot be evaluated.



## References

- [1] S. Ray, "The methodology of officially recognized international sports rating systems," *Journal of Quantitative Analysis in Sports*, vol.7, no.4, 2011.
- [2] FIBA, "FIBA world ranking," <http://www.fiba.com/rankingmen>, 2016, accessed 2016/12/22.
- [3] IHF, "Ranking table," <http://www.ihf.info/en-us/thegame/rankingtable.aspx>, 2016, accessed 2016/12/22.
- [4] FIH, "FIH men's and women's Hero world ranking," <http://www.fih.ch/rankings/outdoor/>, 2017, accessed 2017/4/7.
- [5] FIVB, "FIVB volleyball world rankings," <http://www.fivb.org/en/volleyball/Rankings.asp>, 2016, accessed 2016/6/14.
- [6] E. Konaka, "Statistical rating method for volleyball national teams and its application to result prediction and competition format design," *Proceedings of the Institute of Statistical Mathematics*, vol.65, no.2, pp.251–269, 2017 (in Japanese).
- [7] A.E. Elo, *Ratings of Chess Players Past and Present*, hardcover ed., Harper Collins Distribution Services, 1979.
- [8] World Rugby, "Rankings explanation," <http://www.worldrugby.org/rankings/explanation>, 2014, accessed 2016/6/14.
- [9] H.J. Eom and R.W. Schutz, "Statistical analyses of volleyball team performance," *Research Quarterly for Exercise and Sport*, vol.63, no.1, pp.11–18, 1992. PMID: 1574656.
- [10] E. Zetou, A. Moustakidis, N. Tsigilis, and A. Komninakidou, "Does effectiveness of skill in complex i predict win in men's olympic volleyball games?," *Journal of Quantitative Analysis in Sports*, vol.3, no.4, 2007.
- [11] L.W. Florence, G.W. Fellingham, P.R. Vehrs, and N.P. Mortensen, "Skill evaluation in women's volleyball," *Journal of Quantitative Analysis in Sports*, vol.4, no.2, 2008.
- [12] R.M. Araújo, J. Castro, R. Marcelino, and I.R. Mesquita, "Relationship between the opponent block and the hitter in elite male volleyball," *Journal of Quantitative Analysis in Sports*, vol.6, no.4, pp.1–12, 2010.
- [13] M. Ferrante and G. Fonseca, "On the winning probabilities and mean durations of volleyball," *Journal of Quantitative Analysis in Sports*, vol.10, no.2, pp.91–98, 2014.
- [14] T. Burton and S. Powers, "A linear model for estimating optimal service error fraction in volleyball," *Journal of Quantitative Analysis in Sports*, vol.11, no.2, pp.117–129, 2015.
- [15] The Association for Professional Basketball, APBR.org, <http://www.apbr.org/>, accessed 2018/8/3.
- [16] T. Lundh, "Which ball is the roundest? - a suggested tournament stability index," *Journal of Quantitative Analysis in Sports*, vol.2, no.3, 2006.
- [17] R. Hambleton, H. Swaminathan, and H. Rogers, *Fundamentals of Item Response Theory (Measurement Methods for the Social Science)*, new edition, Sage Publications, 1991.
- [18] R.J. de Ayala, *The Theory and Practice of Item Response Theory (Methodology in the Social Sciences)*, 1st ed., Guilford Press, 2008.
- [19] W.N. Colley, "Colley's bias free college football ranking method: The Colley matrix explained," <http://www.colleyrankings.com/>, 2002, accessed 2018/8/3.
- [20] J. Lasek, Z. Szilávik, and S. Bhulai, "The predictive power of ranking systems in association football," *Int. J. of Applied Pattern Recognition*, vol.1, no.1, pp.27–46, 2013.
- [21] B. Cazeneuve, "Olympic medal predictions: Picking gold, silver, bronze in all 306 events," <http://www.si.com/olympics/2016/08/01/rio-2016-olympics-medal-picks-predictions-projected-medal-count>, 2016, accessed 2016/8/1.
- [22] USA Today, "2016 Rio Olympics medal projections," <http://www.usatoday.com/story/sports/olympics/2016/07/30/2016-rio-olympics-medal-projections/87779154/>, 2016, accessed 2016/8/1.
- [23] Gracenote, "Gracenote's data analytics predicts winners and losers of 2016 rio olympics," <http://www.gracenote.com/gracenotes-data-analytics-predicts-winners-losers-2016-rio-olympics/>, 2016, accessed 2016/8/1.
- [24] G. Cardillo, "McNemar test: perform the mcnemar test on a 2x2 matrix," <http://www.mathworks.com/matlabcentral/fileexchange/15472>, 2007.



**Eiji Konaka** received his B.E., M.E., and Ph.D. degrees in Electrical Engineering from Nagoya University, Japan, in 2000, 2002, and 2005, respectively. Currently, he is an Associate Professor at the Department of Information Engineering, Meijo University. His research interests are in the areas of intelligent control systems and statistic prediction models of sports. He is a member of IEEJ, IEICE, SICE, and IEEE.