

A quantitative method for evaluating the skills of national volleyball teams

EIJI KONAKA (MEIJO UNIVERSITY, JAPAN)



Main objective

- Propose quantitative skill-evaluation for international volleyball teams
- Identify design flaws in the official FIVB ranking system
 - Find over/under-estimated teams in the FIVB rankings
 - Prediction of major worldwide tournaments in 2010s.
 - World Championships (WChs) and Olympic Games)
- Case study: Japan men's teams in WChs 2018
- **Main results:**
 - Proposed method has better prediction performance than FIVB ranking
 - European teams have been underestimated in the FIVB rankings.

Agenda

- **Background**
- Ranking systems, including FIVB rankings
- Proposed method
- Main results
- Discussions
- Conclusions

Background: ranking system

- Ranking systems in sports
 - Evaluation of skill levels
 - Criterion in tournament design
 - Group draws, player seeding, ...
- What is a “good” ranking system?
 - Quantify winning/scoring skills
 - High prediction accuracy
- Ranking point calculation method
 - Accumulative or point exchange (e.g., Elo-based method)

Background: prediction in Rio2016

- Prediction in Rio2016 [Konaka (2019)]
 - **Propose Elo-family (points exchange) rating method**
 - The official rankings in five sports using the accumulative method.
 - Accumulative method: Ranking points are calculated as the sum of the points attributed to international tournaments and the standings in the tournaments.

- Prediction results

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

- **The proposed rating is a better prediction method with $p < 0.01$ by McNemar's test.**

Background

- Ranking system in international volleyball
 - FIVB rankings are an accumulative ranking system
- **Problem presentation**
 - Lack of mathematical or statistical basis in FIVB ranking design.
 - Possible over/under-estimation caused by worldwide tournament system.

Agenda

- Background
- **Ranking systems, including FIVB rankings**
- Proposed method
- Main results
- Discussions
- Conclusions

FIVB ranking system

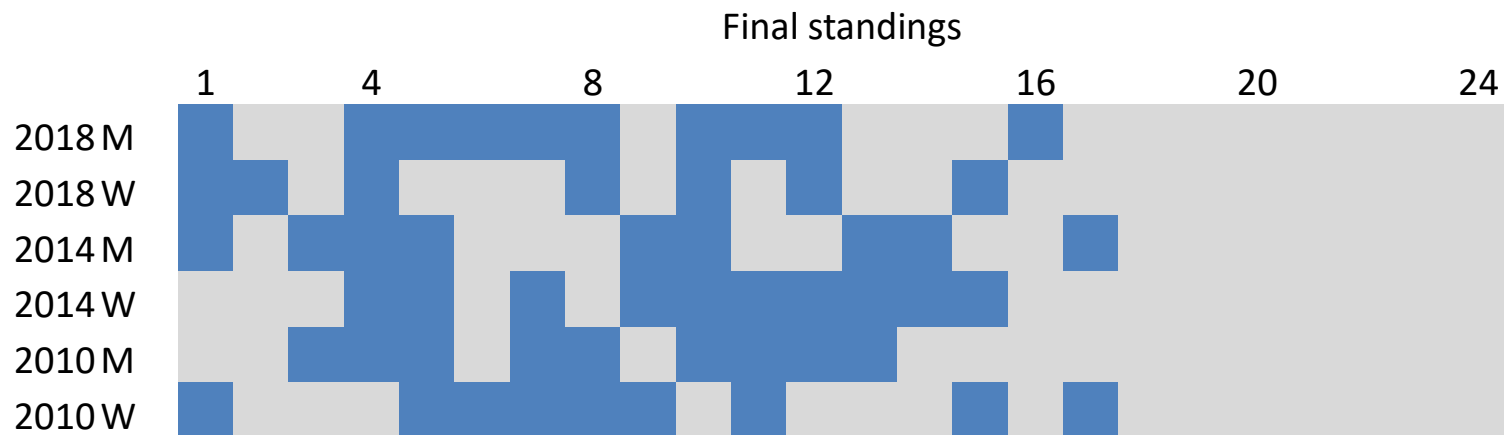
- FIVB ranking point system (2018)
[Excerpt]
- Why are all champions equally awarded 100 points?
- How are the points for each standing designed?
- Next: design flaws in World Cup (third largest tournament)

Table 1: FIVB Ranking Point System (2018)

Standing	Tournament name			
	Olympic Games	World Cup	World Championship	
			Men	Women
1	100	100	100	100
2	90	90	90	90
3	80	80	80	80
4	70	70	70	70
5	50	50	62	58
6	—	40	56	—
7	—	30	50	50
8	—	25	—	—
9	30	5	45	45
10	—	5	—	—
11	20	5	40	40
12	—	5	—	—
13 Tie			36	36
15 Tie			33	33
17 Tie			30	30
21 Tie			25	25

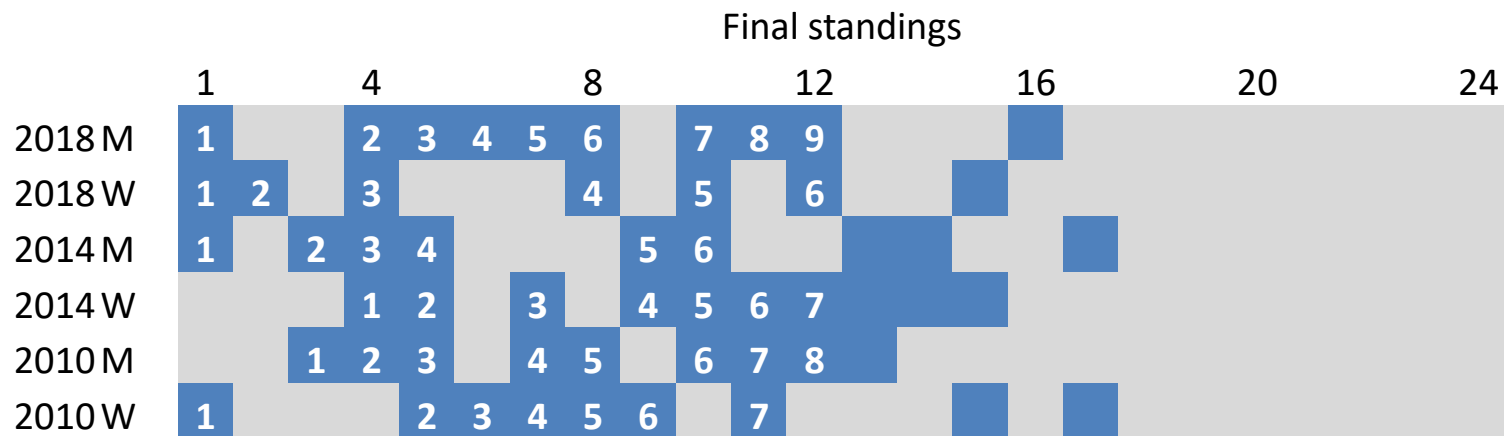
Inconsistent tournament design and underestimation of European teams

- Spot allocation in World Cup volleyball
 - Japan always appears as the host.
 - Ten slots are allocated equally to five confederations.
 - **Only two European teams can appear in this tournament.**
- European teams in WChs



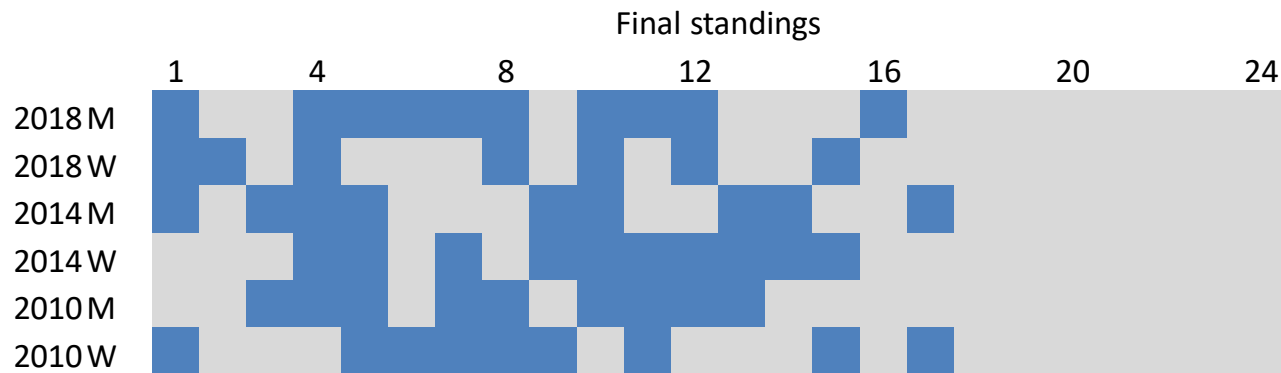
Inconsistent tournament design and underestimation of European teams

- Spot allocation in World Cup volleyball
 - Japan always appears as the host.
 - Ten slots are allocated equally to five confederations.
 - **Only two European teams can appear in this tournament.**
- European teams in WChs



Inconsistent tournament design and underestimation of European teams

- Spot allocation in World Cup volleyball
 - Only two European teams can appear this tournament.
- European teams in WChs



- **European teams could be underestimated in FIVB rankings because of fewer ranking points awarded to Europe from World Cup volleyball**

Agenda

- Background
- Ranking systems, including FIVB rankings
- **Proposed method**
- Main results
- Discussions
- Conclusions

Proposed rating method

- Proposed skill-evaluation method

$$p_{i,j} = 1 / \left(1 + \exp \left(- (r_i + r_{hadv} - r_j) \right) \right)$$

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

Notation	Definition
$i, j \in \{1, \dots, N_T\}$	Indices of teams
r_i	Rating of team i
r_{hadv}	Home advantage (if team i hosts the match)
s_i	Total score of team i in a match
$s_{i,j}$	Actual scoring ratio in a match i against j
$p_{i,j}$	Predicted scoring ratio in a match i against j

Proposed rating method

- Proposed skill-evaluation method

$$p_{i,j} = 1 / \left(1 + \exp \left(- (r_i + r_{hadv} - r_j) \right) \right)$$

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

- Rating estimation

- Simple “steepest descent” method

$$E^2 = \sum_{\substack{(i,j) \in \text{all} \\ \text{matches}}} (s_{i,j} - p_{i,j})^2, \quad r_i \leftarrow r_i - \alpha \cdot \frac{\partial E^2}{\partial r_i}, \quad r_{hadv} \leftarrow r_{hadv} - \alpha \cdot \frac{\partial E^2}{\partial r_{hadv}}$$

Conversion to rating on winning probability

- Proposed skill-evaluation method

$$\hat{w}_{i,j} = 1 / \left(1 + \exp \left(-D_k (r_i + r_{adv} - r_j) \right) \right)$$

$$D_k^* = \arg \min_{D_k} \sum (w_{i,j} - \hat{w}_{i,j})^2, \quad w_{i,j} = 1 \text{ (} i \text{ won) or } 0 \text{ (} j \text{ won)}$$

$$\bar{r}_i = D_k^* r_i$$

Notation	Definition
$w_{i,j}$	Actual won/lost in match i against j
$\hat{w}_{i,j}$	Predicted won/lost probability in match i against j
D_k	Conversion parameter

Conversion to rating on winning probability

- Proposed skill-evaluation method

$$\hat{w}_{i,j} = 1 / \left(1 + \exp \left(-D_k (r_i + r_{adv} - r_j) \right) \right)$$

$$D_k^* = \arg \min_{D_k} \sum (w_{i,j} - \hat{w}_{i,j})^2, \quad w_{i,j} = 1 \text{ (} i \text{ won) or } 0 \text{ (} j \text{ won)}$$

$$\bar{r}_i = D_k^* r_i$$

- Before the prediction target tournament
 - The rating values for every team are calculated by using the major international match results for a couple of years
 - Example: World Cup, Continental Championships, ...

Short-term rating updates during the tournament

- The rating values are updated after every match
 - Based on classical Elo-rating

$$r_i \leftarrow r_i + K(s_{i,j} - p_{i,j}), \quad K = \frac{32 \log_e 10}{400D_k^*}$$

- Summary
 - The difference in rating values explains the scoring ratio via a logistic regression model
 - Rating values are selected to minimize the prediction errors
 - The ratings on winning probability are similarly defined
 - The rating values are updated during tournament, (e.g., WChs.)

Agenda

- Background
- Ranking systems, including FIVB rankings
- Proposed method
- **Main results**
- **Discussions**
- Conclusions

Prediction: target tournaments and datasets

- Prediction target tournaments
 - WChs: 2010, 2014, and 2018.
 - Olympic Games (OL): 2012 and 2016.
- Datasets for prediction model
 - Matches within two years before the target tournament.
 - World Cup: 2011 and 2015
 - Continental Championships
 - Qualifying tournaments
 - Nations league (2018-), World league (Men, -2017), World Grand Prix (Women, -2017)
 - World Grand Champions' Cup: 2013 and 2017
- A total of 733 match results were predicted by using 8,244 match results.

Prediction items

- Prediction methods
 - Proposed method
 - Official FIVB ranking
- Prediction items
 - Win/lose for each match
 - Qualify from the first round
- First round
 - Single round-robin
 - Basically, best four out of six teams qualify to the subsequent round

Prediction results

- Prediction results (match)

		Official			
		Corrects	Incorrects		
Proposed	Corrects	486	79	565	0.771
	Incorrects	58	110	168	
		544	189	733	
		0.742			

McNemar's p -value = 0.0875

Prediction results

- Prediction results (match)

		Official			
		Correct	Incorrect		
Proposed	Correct	486	79	565	0.771
	Incorrect	58	110	168	
		544	189	733	
		0.742			
McNemar's p -value = 0.0875					

- The proposed method can realize better predictions than the FIVB rankings

- Could not prove statistical significance between two methods, i.e., $p = 0.0875 > 0.05$

Prediction results

- Prediction results (qualifying from the first round)

		Official			
		Correct	incorrect		
Proposed	Correct	143	25	168	0.875
	Incorrect	6	18	24	
		149	43	192	
		0.776			

McNemar's p -value = 1.23×10^{-3}

Prediction results

- Prediction results (qualifying from the first round)

		Official		Total	Accuracy
		Correct	incorrect		
Proposed	Correct	143	25	168	0.875
	Incorrect	6	18	24	
		149	43	192	
		0.776			
McNemar's p -value = 1.23×10^{-3}					

- The proposed method can realize better prediction than the FIVB rankings
 - Could prove statistical significance between two methods, i.e., $p = 0.0123 < 0.05$
- Small differences in prediction accuracy would be accumulated through the round-robin format.

Discussion: over/under-estimation in FIVB rankings

- The proposed method is better than the FIVB rankings
- The two methods made different predictions for the following 31 teams.

1 st round result	Proposed method	FIVB rankings	Teams (continents)
Qualify	Qualify	Not qualify	▲ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ▼ [12]
	Not qualify	Qualify	● ▲ [2]
Not qualify	Qualify	Not qualify	● ▲ ■ ■ [4]
	Not qualify	Qualify	● ● ● ● ▲ ▲ ■ ■ ■ ◆ ◆ ◆ ▼ [13]

● Africa, ▲ Asia, ■ Europe, ◆ North and central America, ▼ South America

Discussion: over/under-estimation in FIVB rankings

- The proposed method is better than the FIVB rankings
- The two methods made different predictions for the following 31 teams.

1 st round result	Proposed method	FIVB rankings	Teams (continents)
Qualify	Qualify	Not qualify	▲ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ▼ [12] ←
	Not qualify	Qualify	● ▲ [2]
Not qualify	Qualify	Not qualify	● ▲ ■ ■ [4]
	Not qualify	Qualify	● ● ● ● ▲ ▲ ■ ■ ■ ◆ ◆ ◆ ▼ [13]

● Africa, ▲ Asia, ■ Europe, ◆ North and central America, ▼ South America

- Underestimated teams: 10 out of 12 teams were from Europe

Discussion: over/under-estimation in FIVB rankings

- The proposed method is better than the FIVB rankings
- The two methods made different predictions for the following 31 teams.

1 st round result	Proposed method	FIVB rankings	Teams (continents)
Qualify	Qualify	Not qualify	▲ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ▼ [12]
	Not qualify	Qualify	● ▲ [2]
Not qualify	Qualify	Not qualify	● ▲ ■ ■ [4]
	Not qualify	Qualify	● ● ● ● ▲ ▲ ■ ■ ■ ◆ ◆ ◆ ▼ [13] ←

● Africa, ▲ Asia, ■ Europe, ◆ North and central America, ▼ South America

- Underestimated teams: 10 out of 12 teams were from Europe
- **Overestimated teams: 10 out of 13 teams were from outside Europe**

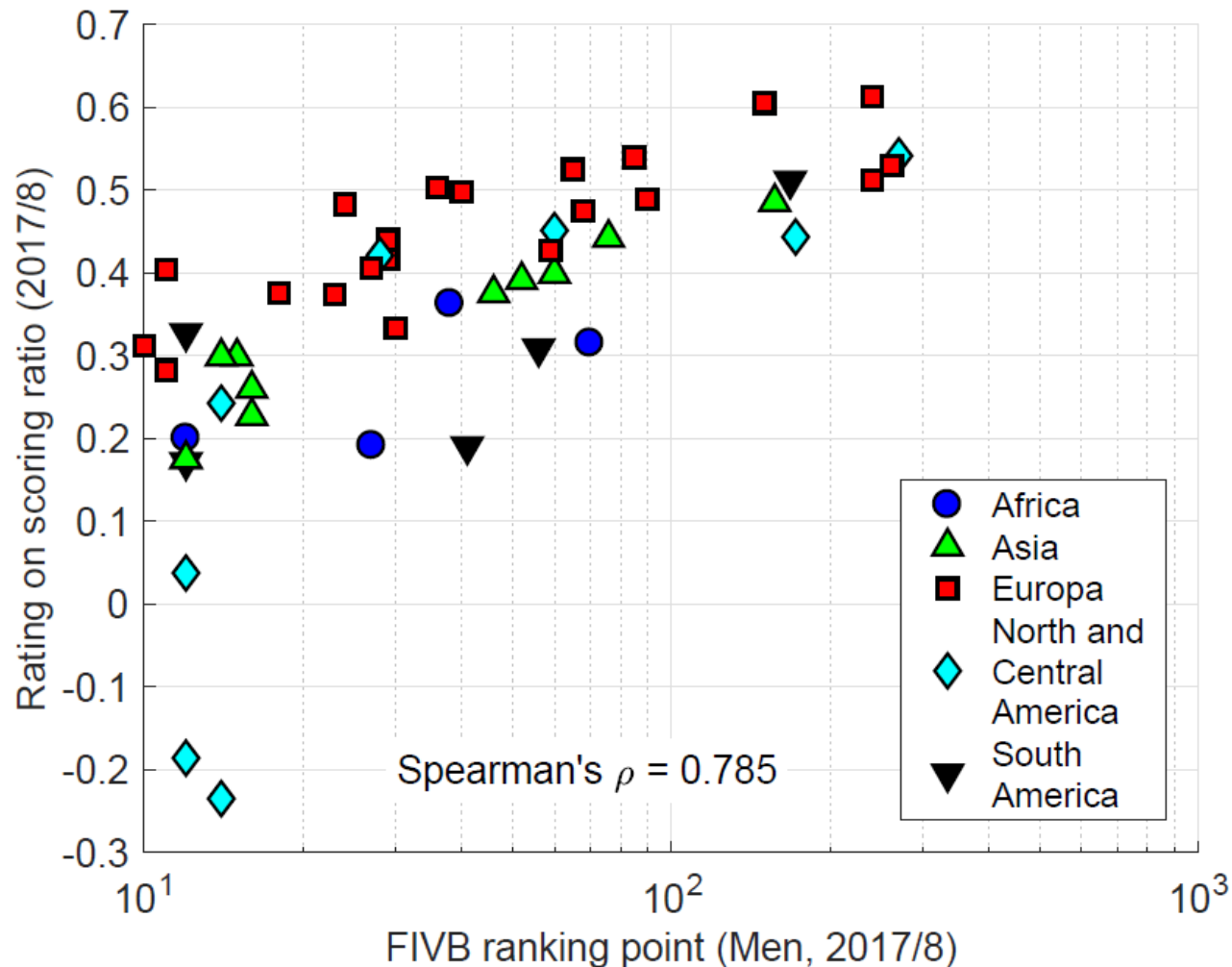
Discussion: over/under-estimation in FIVB rankings

- The proposed method is better than the FIVB rankings
- The two methods made different predictions for the following 31 teams.

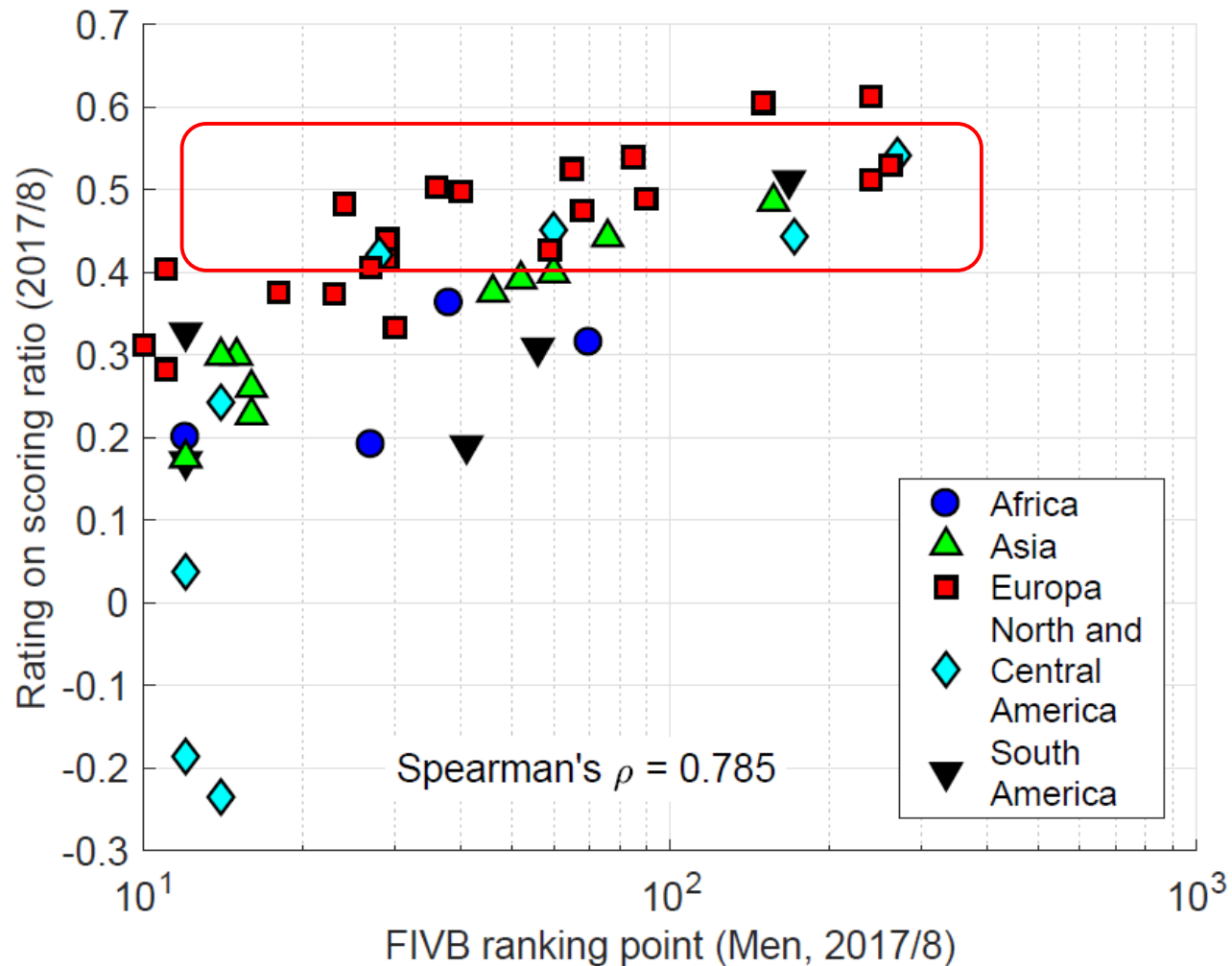
1 st round result	Proposed method	FIVB rankings	Teams (continents)
Qualify	Qualify	Not qualify	▲ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ▼
	Not qualify	Qualify	● ▲
Not qualify	Qualify	Not qualify	● ▲ ■ ■
	Not qualify	Qualify	● ● ● ● ▲ ▲ ■ ■ ■ ◆ ◆ ◆ ▼ ←

European teams are underestimated in the FIVB ranking system

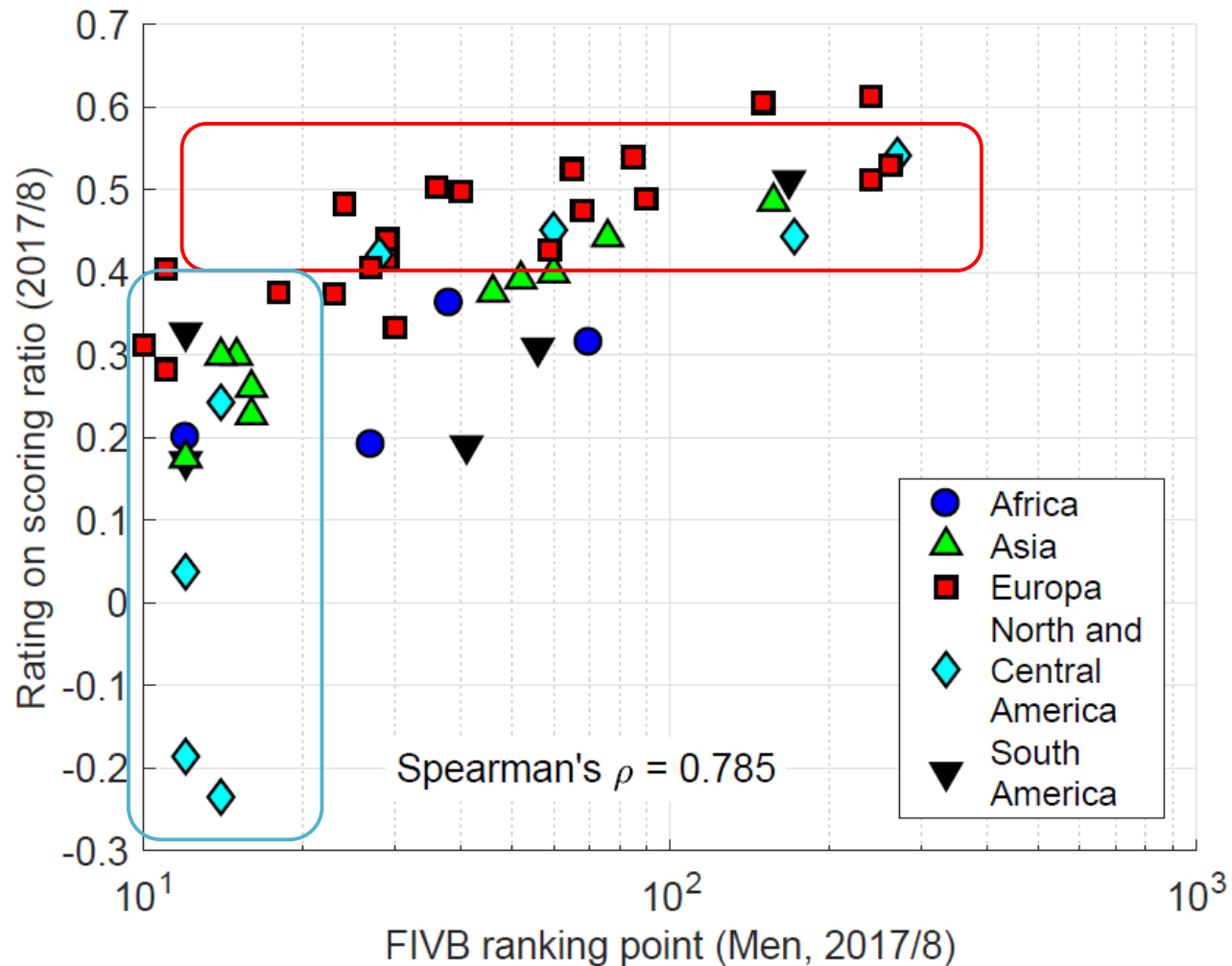
Proposed rating and FIVB ranking points



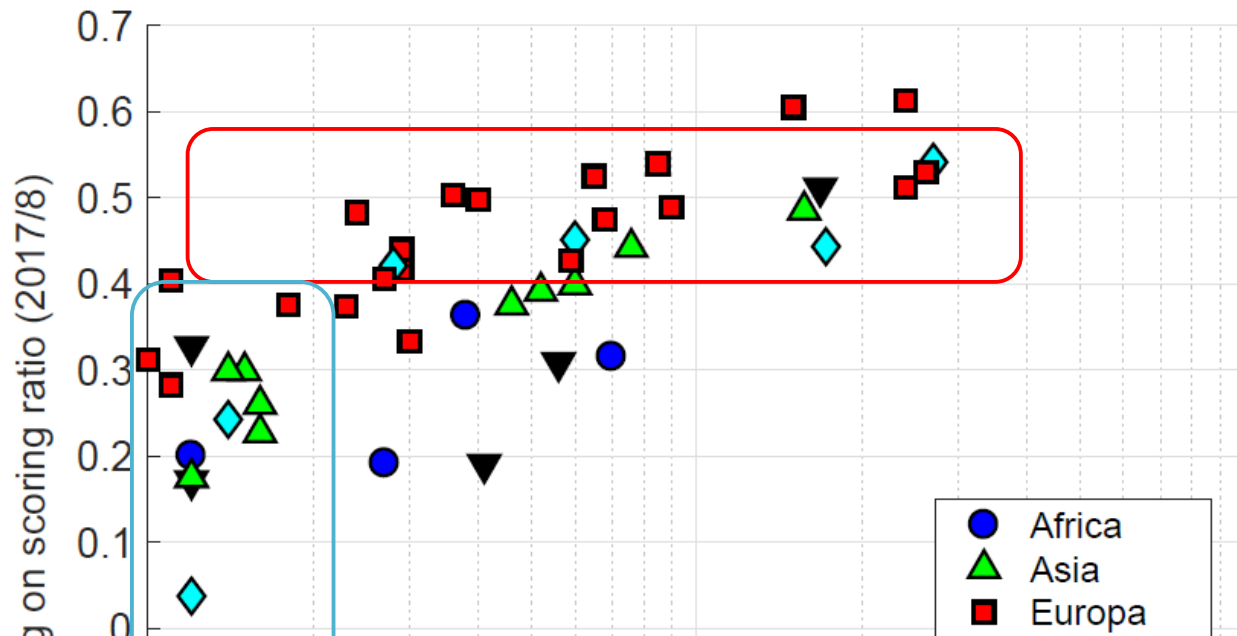
Proposed rating and FIVB ranking points



Proposed rating and FIVB ranking points



Proposed rating and FIVB ranking points



The FIVB ranking system can not measure scoring skill correctly

10^1

10^2

10^3

FIVB ranking point (Men, 2017/8)

Case study: Japan men's team in WCh2018

- Japan men's team in WCh2018

- FIVB ranking: 12

- Third in six teams in Pool A

- Final result: fifth in Pool A

- The main factor: **overestimation in the FIVB ranking**

RANK	TEAMS	MATCHES		
		TOTAL	WON	LOST
1	 ITALY	5	5	0
2	 BELGIUM	5	3	2
3	 SLOVENIA	5	3	2
4	 ARGENTINA	5	2	3
5	 JAPAN	5	2	3
6	 DOMINICAN REPUBLIC	5	0	5

[<https://italy-bulgaria2018.fivb.com/en/results-and-ranking/round1>]

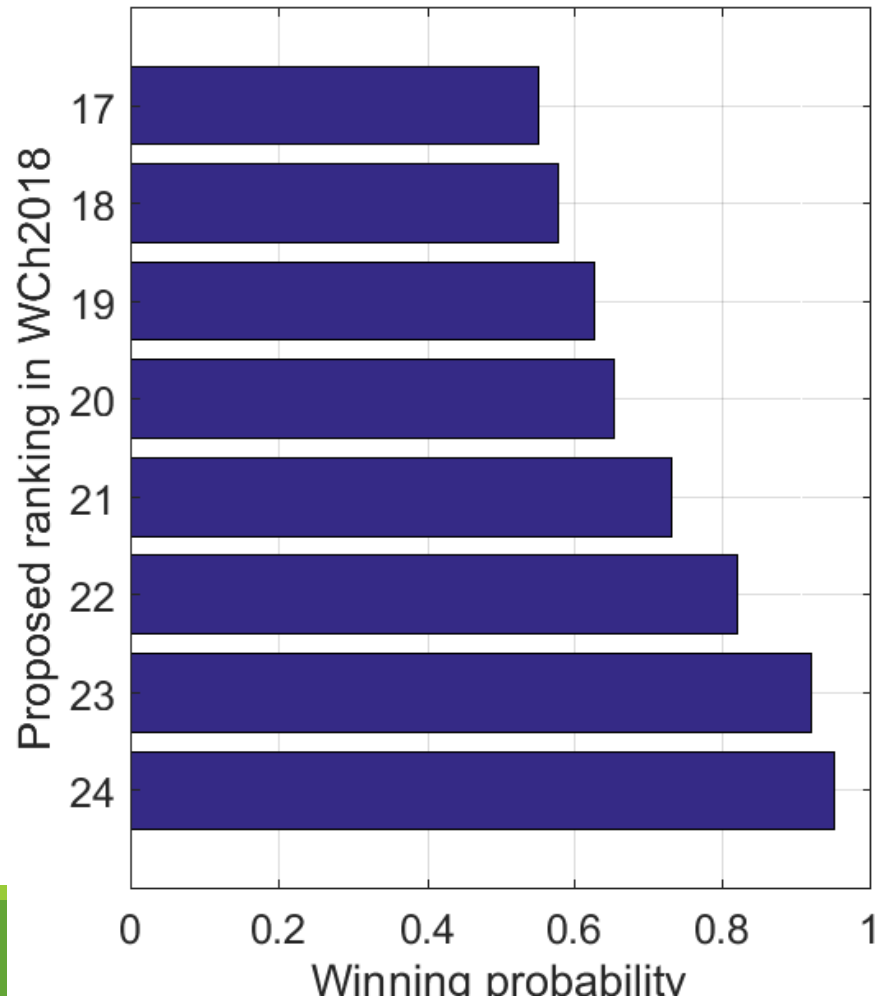
Pool A in WCh2018

- Pool draw (FIVB rankings)
 - ITA(4), ARG(7), **JPN(12)**,
BEL(15), **SLO(23)**, DOM(38)
- Rankings by proposed rating in WCh2018
 - ITA[4], **BEL[8]**, ARG[9],
SLO[11], **JPN[16]**, DOM[23]

RANK	TEAMS	MATCHES		
		TOTAL	WON	LOST
1	 ITALY	5	5	0
2	 BELGIUM	5	3	2
3	 SLOVENIA	5	3	2
4	 ARGENTINA	5	2	3
5	 JAPAN	5	2	3
6	 DOMINICAN REPUBLIC	5	0	5

What happened if the ranking were correct?

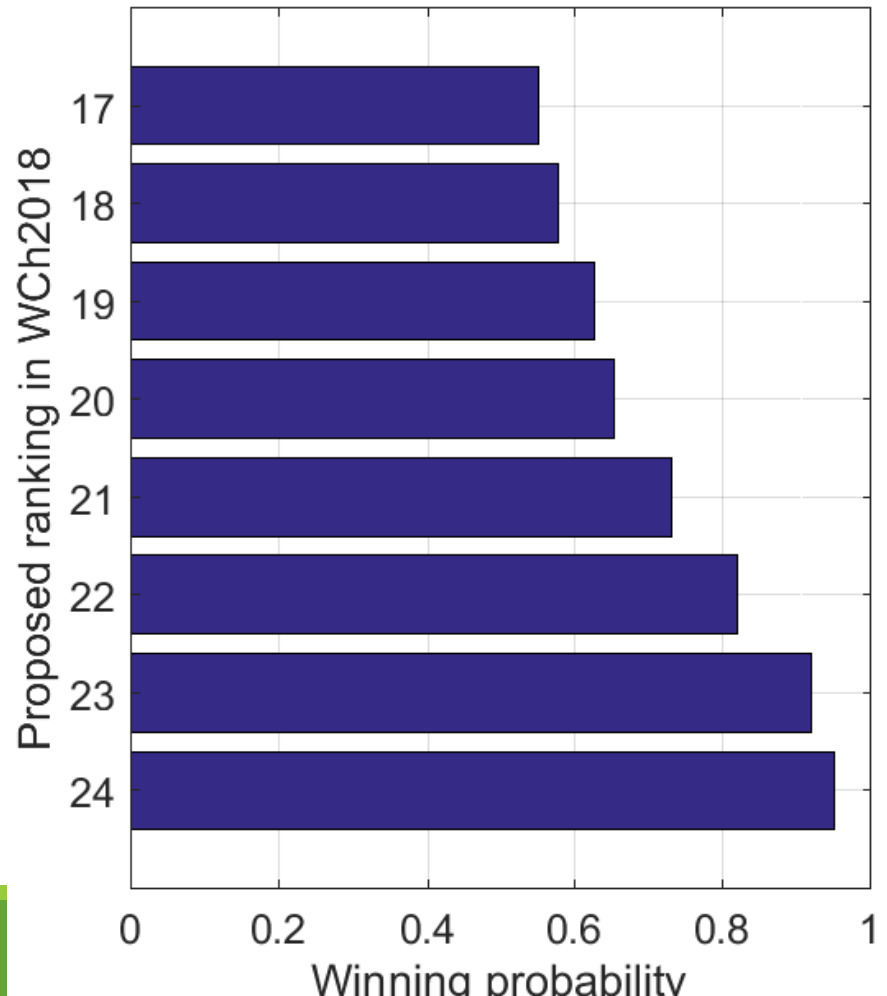
- What happened if the ranking were correct?
 - Japan was 16th by the “correct” ranking
- Predicted winning probability against 17th to 24th teams
 - Japan could have secured fourth place in the first round



What happened if the ranking were correct?

- What happened if the ranking were correct?
 - Japan was 16th by the “correct” ranking
- Predicted winning probability against 17th to 24th teams
 - Japan could have secured fourth place

**Overestimation
prevented “fair”
result**



Conclusion

- A quantitative skill-evaluation for international volleyball teams is proposed
- Identify design flaws in the official FIVB ranking system
- **Main results:**
 - Proposed method has better prediction performance than FIVB ranking
 - European teams have been underestimated in the FIVB rankings



Tournament review: Japan teams in WChs 2018
