

カーリング各国代表チームの実力の定量的分析

小中 英嗣
名城大学

和文概要 カーリングは1998年の長野大会より冬季オリンピックの正式競技に採用されるなど、世界的に普及している冬のスポーツである。世界カーリング連盟 (World Curling Federation, WCF) によれば、カーリングの世界大会の起源は1959年に開催され、カナダとスコットランドのみが参加した Scotch Cup である (5戦してカナダが5勝)。この大会はこれ以降参加国を増やして世界選手権へと発展し、また大陸毎の選手権や下位グループの大会が新設されるなど、現在ではおよそ50カ国・地域の代表チームが国際大会で競う世界的なスポーツに成長している。本稿ではWCF管理下の全ての国際大会の試合結果から、各国代表チームの実力を定量的に評価する手法を提案する。それに基づき、世界的な普及、および実力分布の歴史的な経緯について議論することを目的とする。

キーワード: 予測, 数理モデル, 確率モデル, スポーツ, カーリング

1. 研究背景

カーリングは1998年の長野大会よりオリンピックの正式競技に採用されるなど、世界的に普及している冬のスポーツである。世界カーリング連盟 (World Curling Federation, WCF) によれば、カーリングの世界大会のルーツは1959年に開催され、カナダとスコットランドの2ヶ国のみが参加した Scotch Cup [8] である (5戦してカナダが5勝)。この大会はこれ以降参加国を増やし世界選手権へと発展し、また大陸毎の選手権や下位グループの大会が新設されるなど、現在ではおよそ50カ国・地域の代表チームが国際大会で競う世界的なスポーツに成長している。

本稿ではまずカーリングの主要国際大会の歴史を簡潔に示し、その参加国数の増加および地理的な広がりによりカーリングの普及を概観する。次にWCFが管理している全ての国際大会の試合結果に基づき、各国の実力の評価値としてのレーティングおよびその計算手法を提案する。具体的には、チーム間のレーティング差が勝率 (または得点率) をロジスティック回帰モデルにより説明するモデルを仮定し、そのモデルに基づき試合結果を最もよく説明する値をレーティングとして算出する。カーリングの個々の技術・戦略に関する分析は散見されるが学術論文の形式で公表・集積は小規模であり、かつ試合結果から実力を定量的に評価した研究成果の報告は (著者の調査の範囲内では) 皆無である。

提案手法により算出したレーティングは間隔尺度となり、その差の絶対値を距離とした階層型クラスタリングを用いることで実力の分布を定量的に評価することが可能となる。この事実に基づき、代表チームの実力分布の歴史的変化、および男女間での違いを分析する。この手法はカーリングに限らず、ある程度の得点が入る全てのスポーツでの実力分布の分析にも適用可能であり、オリンピックの正式種目となっている他のスポーツについての同様の分析を紹介する。

2. カーリングの国際大会の歴史

表1にカーリングの主要国際大会とその創設年を，また図1に大会ごとの参加チーム数を示す。

表 1: Major international competitions of curling

Competition Name	Gender	From	Teams in latest event
World Curling Championships (WCC)	Men	1959(*1)	12
	Women	1979	12
European Curling Championships (ECC)	Men	1975	26
	Women	1975	20
European Curling Championships "C" Division (ECC-C)(*2)	Men	2010	11
	Women	2010	8
Pacific-Asia Curling Championships (PCC)	Men	1991	8
	Women	1993	5
Winter Olympic Games (WOG)(*3)	Men	1998	10
	Women	1998	10
Winter Olympic Qualifying Event (WOQ)(*3)	Men	2013	8
	Women	2013	7

(*1): The competition is named as "Scotch cup" (1959–1967), and "Silver Bloom" (1968–1985).

(*2): be separated from European Curling Championships.

(*3): be held once every four years.

1959年にスコットランドとカナダの対抗戦として創設された Scotch Cup は参加チーム数の増加と名称変更を経て，近年は世界12カ国による世界選手権として定着している。

ヨーロッパ選手権が男女ともに1975年に創立され，女子の世界選手権はそれより後の1979年に第1回が開催されている。その後ヨーロッパ選手権はその規模を拡大し，2000年代には男子で30カ国が参加する大規模なものとなっている。その後2010年より下位チームの大会がヨーロッパ選手権Cとして分離されている。

北米のカナダ，アメリカ合衆国はカーリングの強豪国である(特にカナダは2016年までの世界選手権で男女それぞれ35回，13回優勝している)が，WCFが主催する大陸選手権は開催されていない。

欧米外へ目を向けると，太平洋アジア選手権(Pacific Asia Championship)が男女ともに1990年代に創設されている。当初は英連邦のオーストラリア，ニュージーランドに日本を加えた3カ国で始まり，その後規模を拡大している。ヨーロッパ，太平洋アジアそれぞれの大会規模の拡大に決定的な要因となったのがオリンピック正式種目としての採用(1998年長野大会)であった。オリンピック出場権を争う形式も変更され，2014年ソチ大会の予選として2013年にOlympic Qualifying Eventが新設されている。これら主要国際大会へ参加経験のあるチーム数は男女それぞれでおよそ50，40である。

3. レーティングの定義および算出方法

前章でカーリングの国際大会の拡大の歴史を概観した。本節ではその全ての試合結果に基づき、各国の実力を定量的に評価するレーティングを定義し、その算出アルゴリズムもあわせて提案する。この定量的評価により、増加したチーム間で実力が拮抗しているのかどうかを定量的に評価可能となる。

2つのチームが対戦した場合、大きく二つに分けて、

1. それぞれのチームが安定的に持っている技能・実力
2. 当日の調子・運, 相性

の要因が得点や勝敗に影響すると仮定する。本稿では、狭義のレーティングとして「上記の(1)を統計的に算出したもの」と定義する。カーリングに関して、試合結果から実力を定量的に評価した研究成果の報告は(著者の調査の範囲内では)皆無である。

WCFが公開している試合結果 [7] において、全ての大会で共通して得られたのは各試合の各チームの総得点であった(古い試合では各エンドの得点が記録されていないものがある)。この利用できるデータから推定可能なパラメータとして、対戦チーム間での値の差が各チームの勝率、または得点率を説明するものを提案する。具体的には、各チーム i の実力を示すパラメータ(今後、これをチーム i のレーティングと呼ぶ)を r_i で示し、チーム i と j が対戦した場合、 i の勝率(または得点率) $p_{i,j}$ が次式に従うと仮定する。

$$p_{i,j} = \frac{1}{1 + e^{-(r_i - r_j)}}. \quad (3.1)$$

試合結果からレーティングを算出するアルゴリズムを以下に示す。表 2 に記号の意味を示す。

表 2: Notations

N_T	Number of teams
$\mathbf{r} = (r_1, \dots, r_{N_T})^T$	Rating vector
N_G	Number of games
$\langle i, j, s_i, s_j \rangle$	Result of one game. Team i and j scored s_i and s_j points in a game. N_G tuples are stored in database.
ϵ_{th}	Threshold value
K	Parameter used in rating update
k	Iteration index
$\mathbf{0}, \mathbf{1}$	Column vector composed of zeros and ones with suitable dimensions
$\ \mathbf{x}\ $	Euclidean norm of vector x

レーティング算出アルゴリズム

1. 勝率、または得点率のどちらに対するレーティングを算出するか選ぶ。

$\mathbf{r}^{(0)} = \mathbf{0}$ とし、収束判定のための $\epsilon_{th} > 0$ を小さな値に設定する。 $K > 0$ を定める。経験上、 10^{-3} かそれ以下の値が望ましい。繰り返し回数を $k = 0$ とする。 N_G 試合の結果を保持したデータベースを用意する。各試合の結果を $\langle i, j, s_i, s_j \rangle$ で示す。

2. データベースの各試合結果をランダムにソートする。

3. ソートされたデータベースから試合結果を順に取り出し、これを $\langle i, j, s_i, s_j \rangle$ とする。
4. レーティング r_i および r_j を以下の規則で更新する。

- 勝率に対するレーティングを算出する場合：

$$p = \frac{1}{1 + e^{-(r_i^{(k)} - r_j^{(k)})}}, \quad s = 1 \text{ (if } s_i > s_j), 0 \text{ (if } s_i < s_j) \quad (3.2)$$

- 得点率に対するレーティングを算出する場合：

$$p = \frac{1}{1 + e^{-(r_i^{(k)} - r_j^{(k)})}}, \quad s = \frac{s_i}{s_i + s_j}, \quad (3.3)$$

$$r_i^{(k+1)} = r_i^{(k)} + K(s - p), \quad r_j^{(k+1)} = r_j^{(k)} + K((1 - s) - (1 - p)). \quad (3.4)$$

データベース内の全試合に対して上記の更新を行う。

5. $\|r^{(k+1)} - r^{(k)}\| < \epsilon_{th}$ が成り立つ場合、 $r^{(k+1)}$ を出力としてアルゴリズムを終了する。そうでない場合 $k \leftarrow k + 1$ としてステップ 2. へ戻る。 □

3.1. 提案手法の位置づけ

対の比較結果から全体の順位付けを行う他の手法として、Bradley-Terry モデル (BT モデル)[3, 1] がよく知られており、スポーツのレーティングに関する研究も多い。しかし、さまざまな競技団体 (スポーツやテーブルゲームを含む) のランキング制度に関する文献 [4] の調査では、BT モデルに基づくレーティング手法を採用している競技団体は無い。

それに対し、提案手法はいくつかの競技団体で (その変種も含めて) いくつか採用されているイロレーティング (Elo rating)[2]¹ と類似の手法である。

イロレーティングは試合ごとにその結果に基づき両チームのレーティングを修正する手法であり、試合数が少ないチームのレーティングの変化量が小さいことや、引退・新規参入に伴いその値が増加傾向を示すなどの問題点が指摘されている。それに対し、提案手法はある一定期間内での平均のレーティングを算出していると解釈でき、イロレーティングの問題点を解決できている。

さらに、イロレーティングおよび提案手法はレーティングが間隔尺度、つまりその差が対戦時の勝率を説明するモデルを採用しており、このことから提案手法は間隔尺度に基づくクラスタリング手法と一貫性があることが分かる。レーティングのクラスタリングに基づく実力分布の定量的分析については第 4.2 節で詳細に述べる。

4. レーティングに基づく実力分布の分析

WCF では当年を含めた 6 年間の国際大会の成績をそれぞれ定量化し、その総和により各国のランキングポイントを算出している [6]。本研究ではこれに倣い、当年を含めた 6 年間の全ての試合結果に基づき、各国のレーティングを前節で提案した手法に基づき算出する。たとえば、2014 年末のレーティングは 2009 年から 2014 年に行われた全試合結果から算出する。レーティングの算出は各年に対して行う。

試合結果は WCF が公式サイト [7] で Web ページとして公開しているものからスクレイピングにより取得した。取得時期は 2016 年 6 月下旬から 7 月上旬であり、この期間より後の

¹もともとチェスプレイヤーの順序付け手法として開発された。

国際大会の結果，および公式サイトの入力ミス・データベース損傷などにより取得できないデータについては取り扱わない．取得できた試合数を図2に示す．総試合数は12526試合（男子7034試合，女子5492試合）であった．

4.1. レーティング算出結果および評価

本節では勝率および得点率を説明するレーティングの算出結果を示す．レーティングは間隔尺度であり，定数を加算して原点(0)を移動しても問題ない．したがって，得られたレーティングベクトル \mathbf{r} に対し，

$$\mathbf{r} \leftarrow \mathbf{r} - (\max \mathbf{r}) \cdot \mathbf{1} \quad (4.1)$$

とすることで，レーティングの最上位を必ず0とし，そこからの差を表すことができる．

男女それぞれの勝率を説明するレーティングを図3および図4に示す．図の煩雑さを避けるため，個別の国名は示さず大陸（ヨーロッパ，北米，太平洋アジア）のみを示した．またレーティングは全ての年に対して算出しているが，図中では2年ごとに示している．

レーティング分布の経年変化を確認するため，横軸にレーティングに基づくランキング，縦軸にレーティングをとり，6年おきに描画したグラフを図5から図8に示す．それぞれ勝敗・得点率のレーティングを男女別に示し，上図は上位チーム，下図は全チームを示す．

2015年末のレーティングをWard法[5]の階層型クラスタリングによりクラスタリングした結果を図9に示す．クラスタ間距離の閾値は，勝率のレーティング，得点率のレーティングそれぞれに対して1.0，0.1としている．

前年末のレーティングで当年の試合結果を予測した結果を図10に示す．予測結果の評価は対戦チーム間のレーティング差から予測される勝率または得点率と試合結果との相関係数に基づく．

4.2. レーティングに基づく実力分布の定量的評価

図3および図4より，男女ともに北米（カナダ）が長く高い実力を示していることや，ここ10年間での太平洋アジア各国の実力の上昇が示されている．図5から図8も，参加国の増加が全体的に競技力を押し上げていることを示している．近年，上位で実力が拮抗しているチームが増加しており，世界選手権（12チーム），オリンピック（10チーム）の出場権を得るための競争が激しさを増していることが示されている．たとえば，男子で1985年の勝敗レーティング10位はおよそ -1.0 であったが，2015年ではおよそ -0.6 である．このことは，（レーティングで）世界1位と対戦したとき，1985年時点では勝率が0.269程度の実力のチームが世界選手権参加へのボーダーラインであったのに対し，2015年時点では期待勝率が0.354程度無いと世界選手権へ出場できなくなっている事を示している．（正確には大陸毎の出場枠があるので，この限りではない）．

図9から，10番手までの実力分布は男子よりも女子のほうが密であるが，それ以降では男子の実力分布が密であることが分かる．

図10は，レーティング差から予測される勝率（得点率）と実際の勝率（得点率）との相関係数が0.5付近であり，中程度の相関があることを示している．このことは，カーリングは選手やチームが安定的に保持している技能以外の要素（当日の調子，運，相性など）が得点および勝敗に与える影響が大きいスポーツであることを示している．

比較対象としてバレーボールでのレーティングの分布および相関係数を示す．リオデジャネイロオリンピック直前1年間の主要国際大会の結果から得点率を説明するレーティングを求め，それを勝率を説明するレーティングに変換した結果を図11に示す．

図 11 を図 9 と比較すると、同順位でバレーボールのレーティングはカーリングよりも低く、かつ下がり方も急であり、同じ閾値でまとめたクラスタ内のチーム数も少ない。このことは、バレーボールでは上位チームといえども実力差が大きく、実力が拮抗したチーム数も少ないことを示している。

また、バレーボールではレーティング差から予測される得点率と実際の得点率との相関は約 0.74 と、より強い。このことは、バレーボールはカーリングと比較して、選手やチームが安定的に保持している技能により得点および勝敗が定まる傾向が強いことを示している。同様の分析は他のスポーツにも適用可能であり、今後の研究課題である。

5. 結論

本論文ではカーリングの国際大会の結果(各試合での得点)に基づき、各チームの実力を示すパラメータ(レーティング)を算出する方法を提案した。レーティングの差が得点率を説明するロジスティック回帰モデルを採用した。1959 年以降に開催された主要な国際大会における約 1 万 2000 試合の試合結果を世界カーリング連盟の公式サイトより取得し、各年について当年を含む過去 6 年間の試合結果に基づきレーティングを算出した。レーティング分布は、近年実力が拮抗しているチームが増加し、オリンピック出場権獲得のための競争が激しさを増していることを示している。また、他競技(バレーボール)とのレーティング分布および実際の得点率との相関係数の比較により、カーリングは選手やチームが安定的に保持している技能以外の要素(当日の調子、運、相性など)が得点および勝敗に与える影響が大きいスポーツであることが分かった。

参考文献

- [1] Roger R. Davidson. On extending the bradley-terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, Vol. 65, No. 329, pp. 317–328, 1970.
- [2] Arpad E. Elo. *Ratings of Chess Players Past and Present*. Harper Collins Distribution Services, hardcover edition, 1979.
- [3] L. L. Kupper and P. V. Rao. Ties in paired-comparison experiments: A generalization of the bradley-terry model. *Journal of the American Statistical Association*, Vol. 62, No. 317, pp. 194–204, 1967.
- [4] Stefani Ray. The methodology of officially recognized international sports rating systems. *Journal of Quantitative Analysis in Sports*, Vol. 7, No. 4, 2011.
- [5] Joe H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, Vol. 58, No. 301, pp. 236–244, 1963.
- [6] World Curling Federation. Men's & women's world rankings. <http://www.worldcurling.org/rankings-men-and-women>, 2016. referred in 2016/11.
- [7] World Curling Federation. Results & statistics. <http://results.worldcurling.org/>, 2016. referred in 2016/11.
- [8] World Curling Federation. Scotch cup 1959. <http://results.worldcurling.org/Championship/Details/5>, 2016. referred in 2016/11.

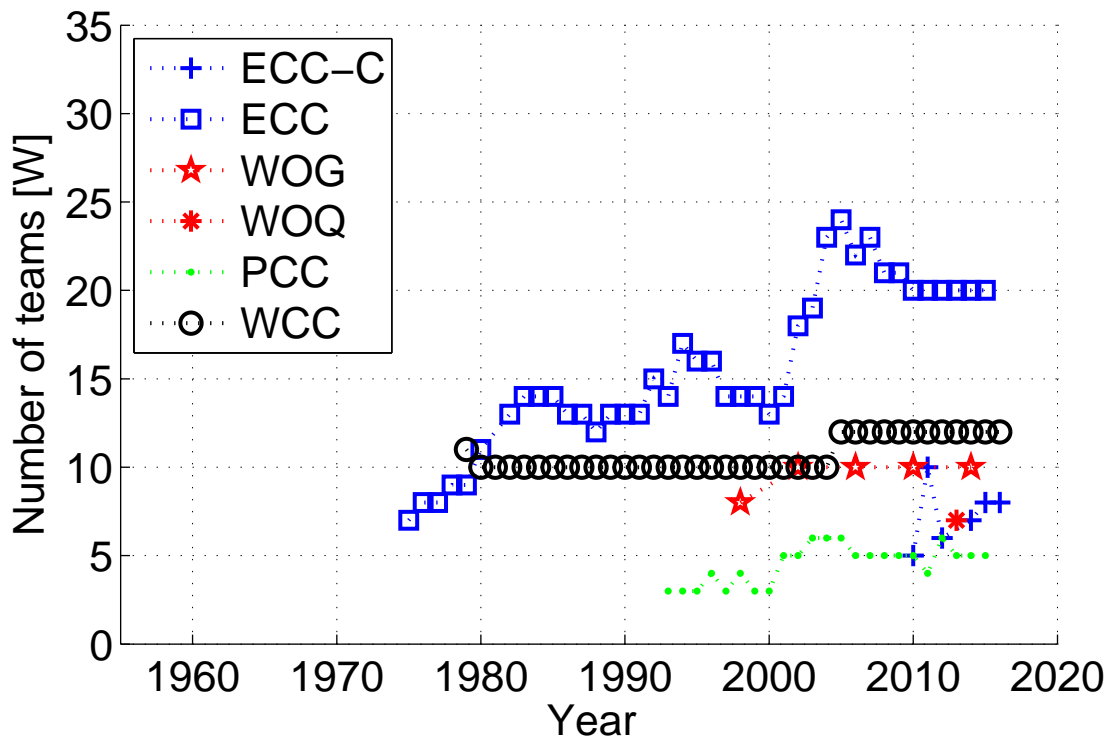
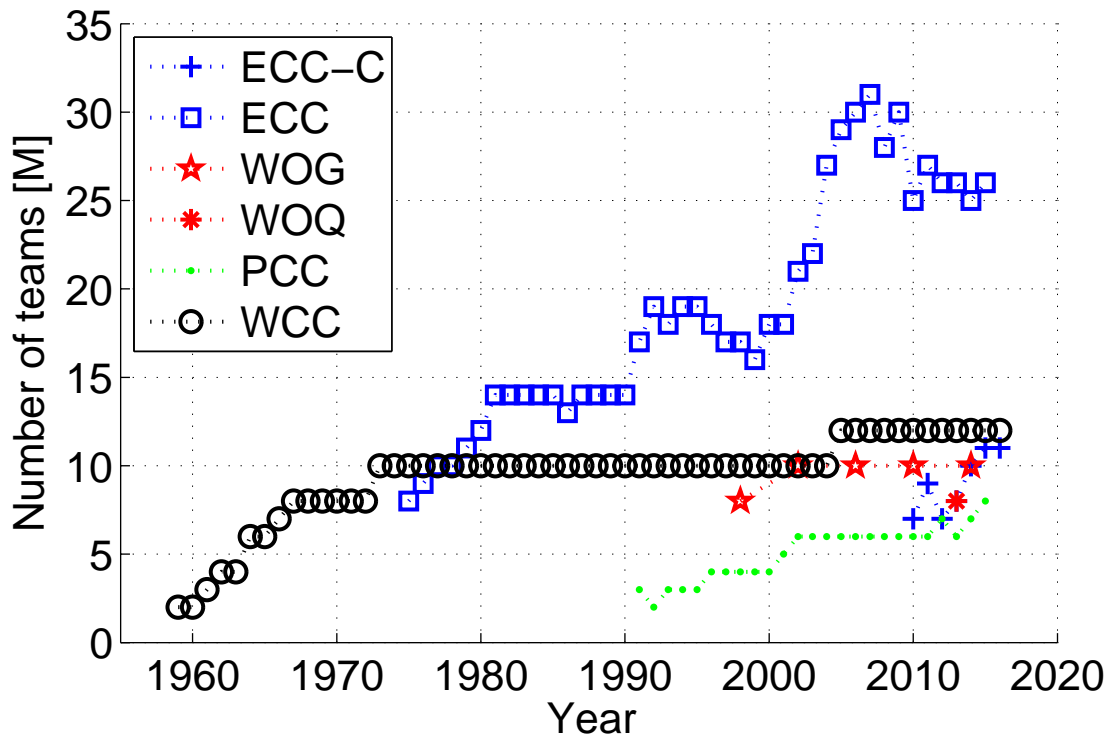


Figure 1: Number of teams in major international competitions of curling

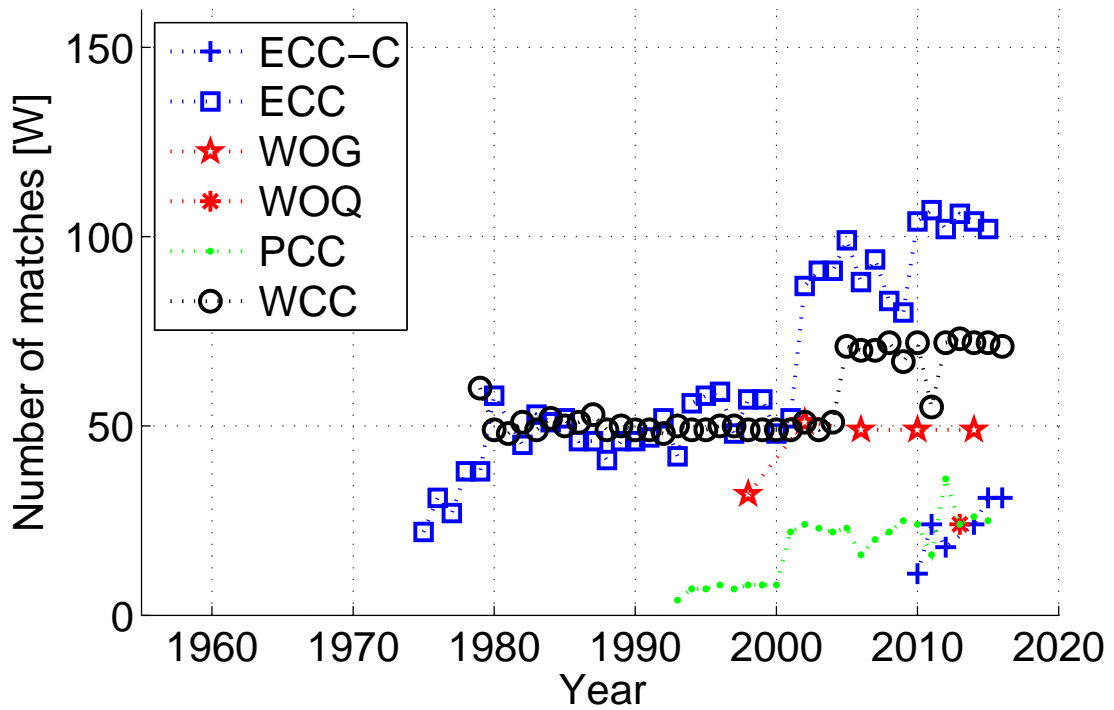
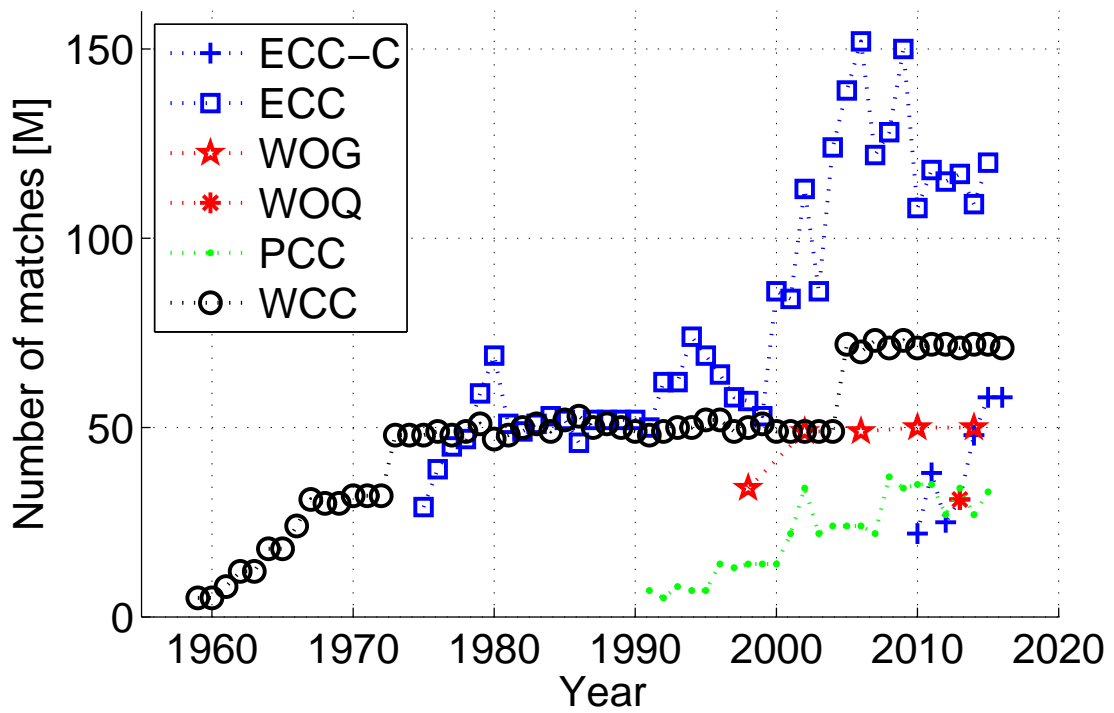


图 2: Number of matches in major international competitions of curling

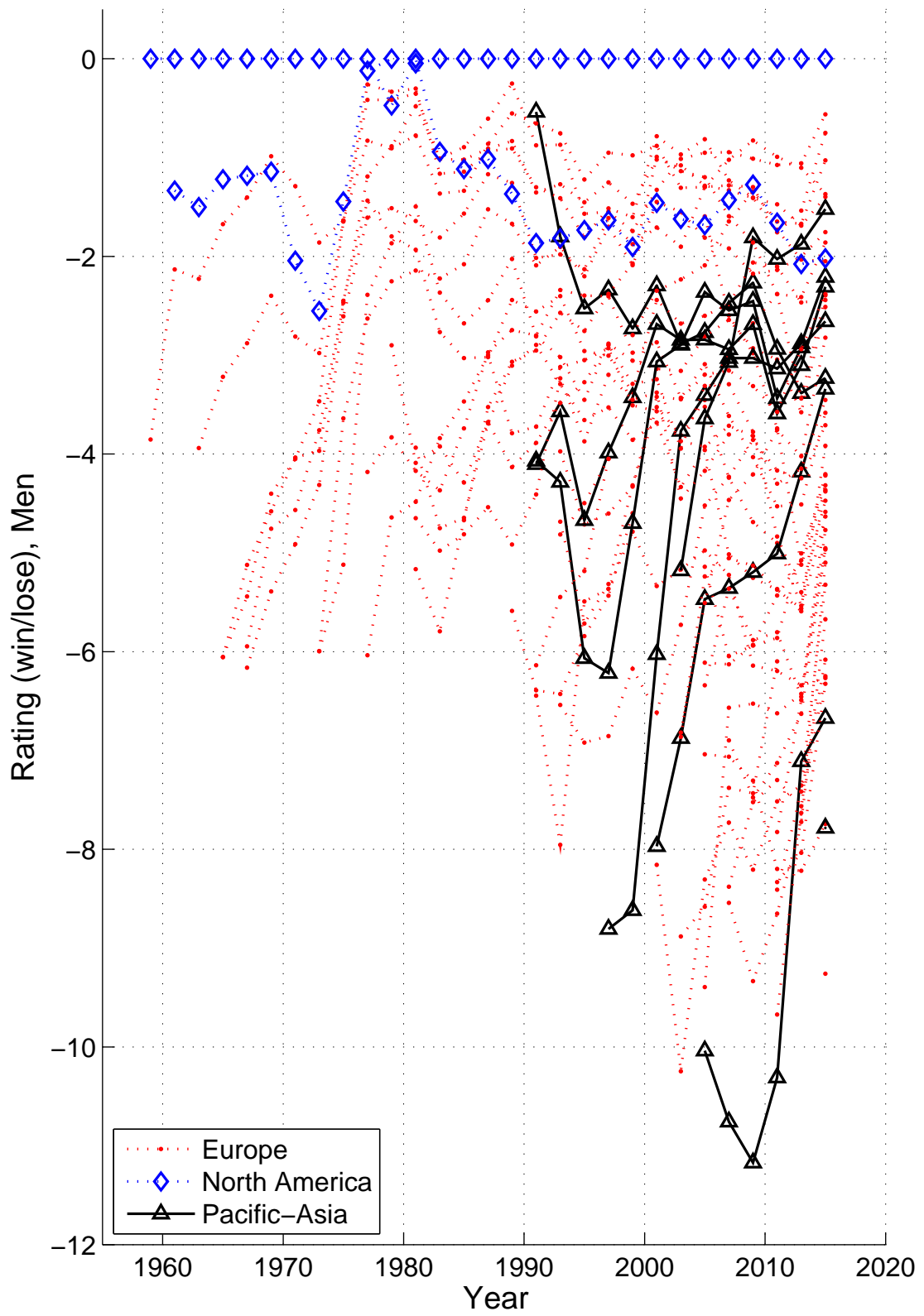


图 3: Rating history (men)

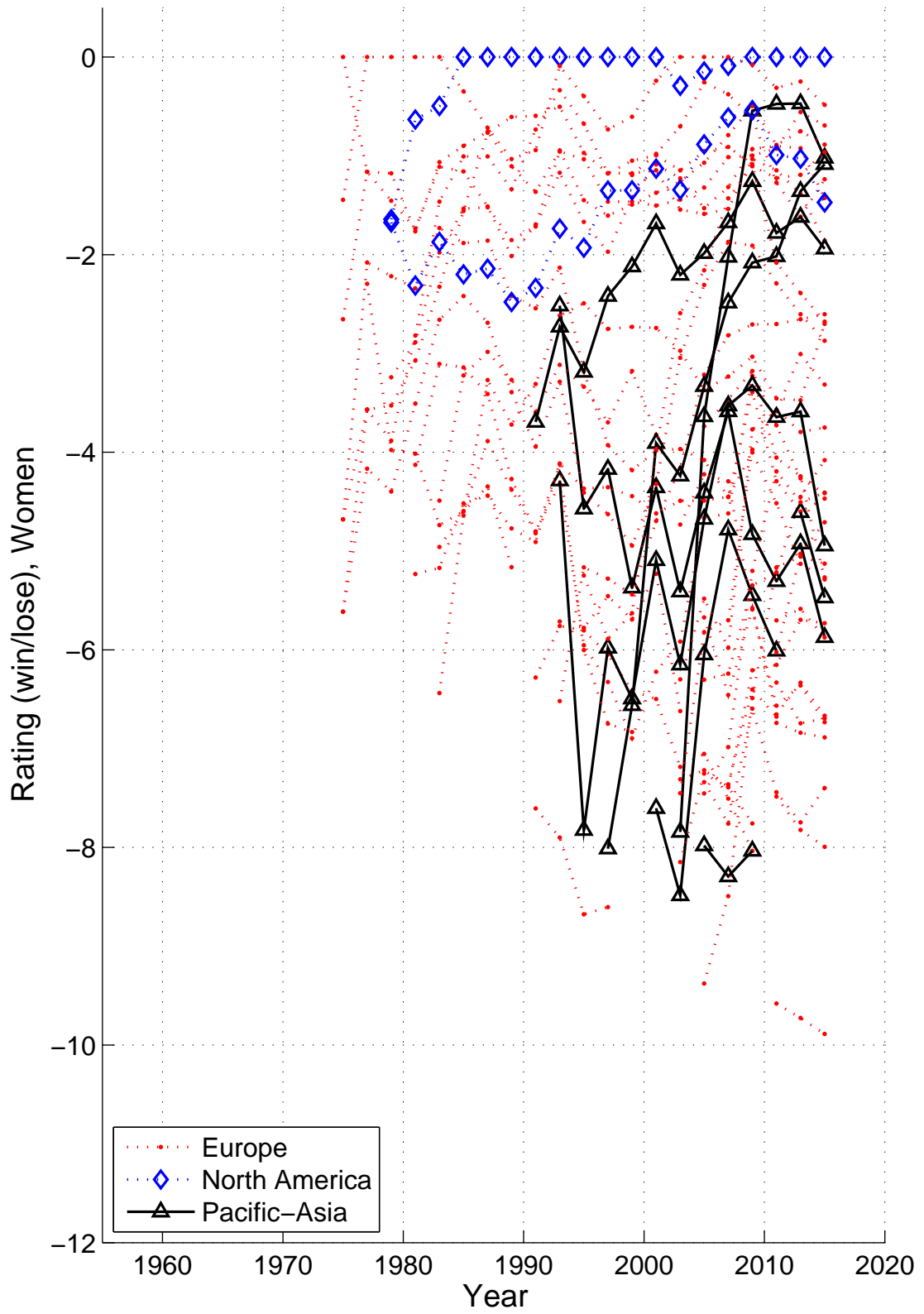
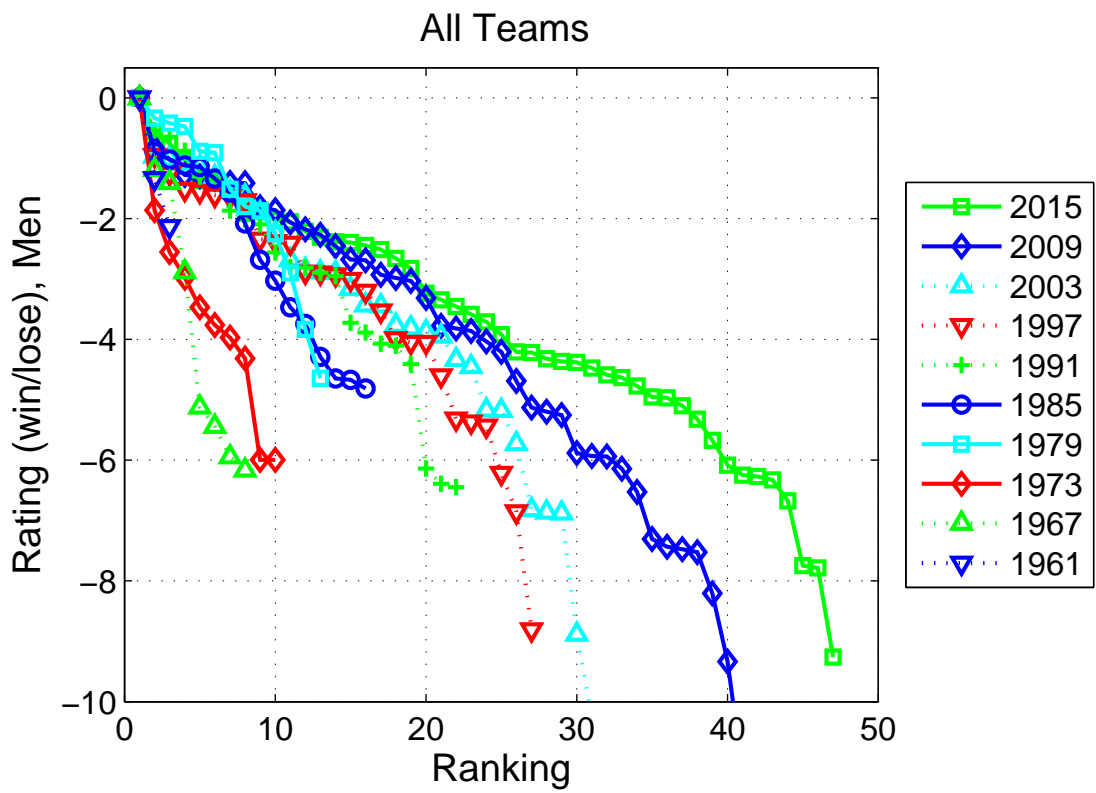
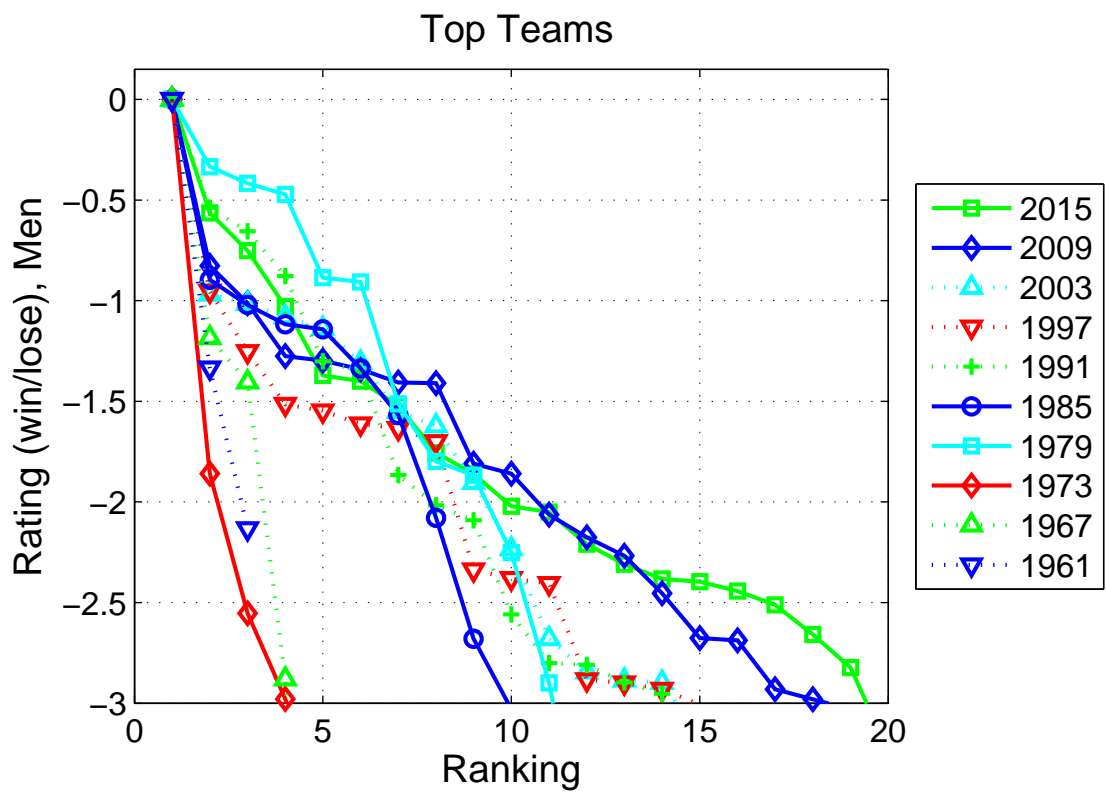


图 4: Rating history (women)



⊠ 5: Rating distribution (win/lose, men)

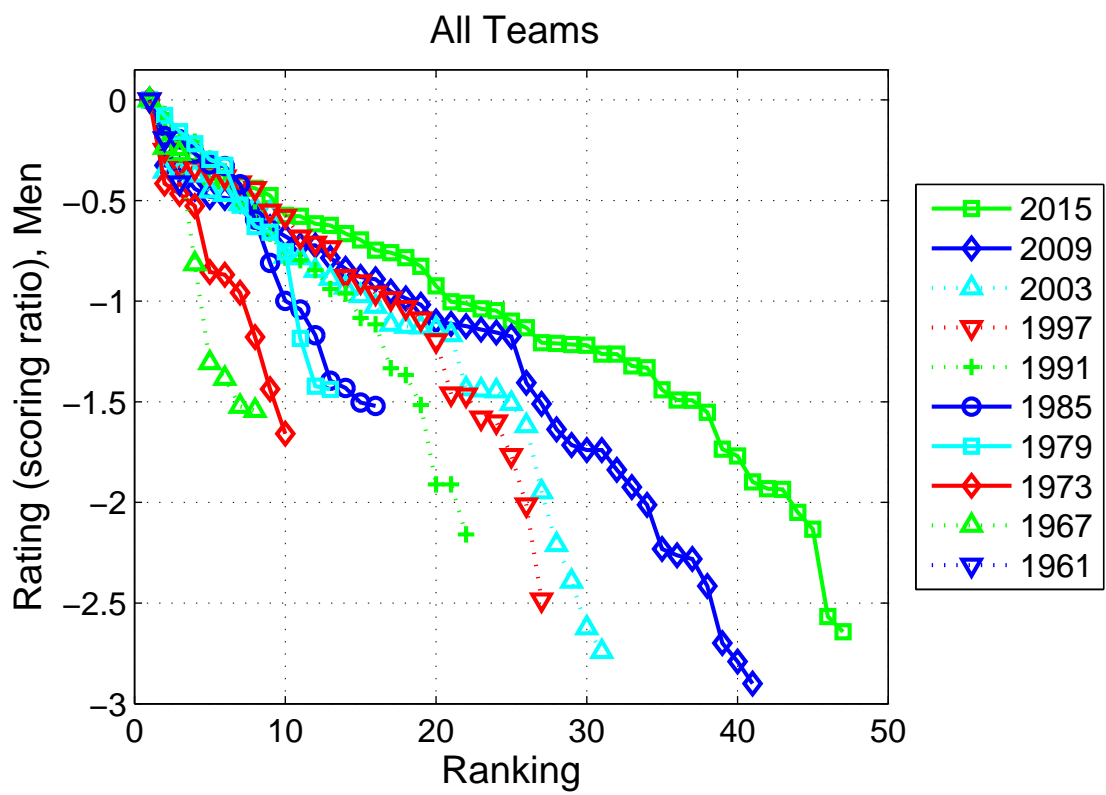
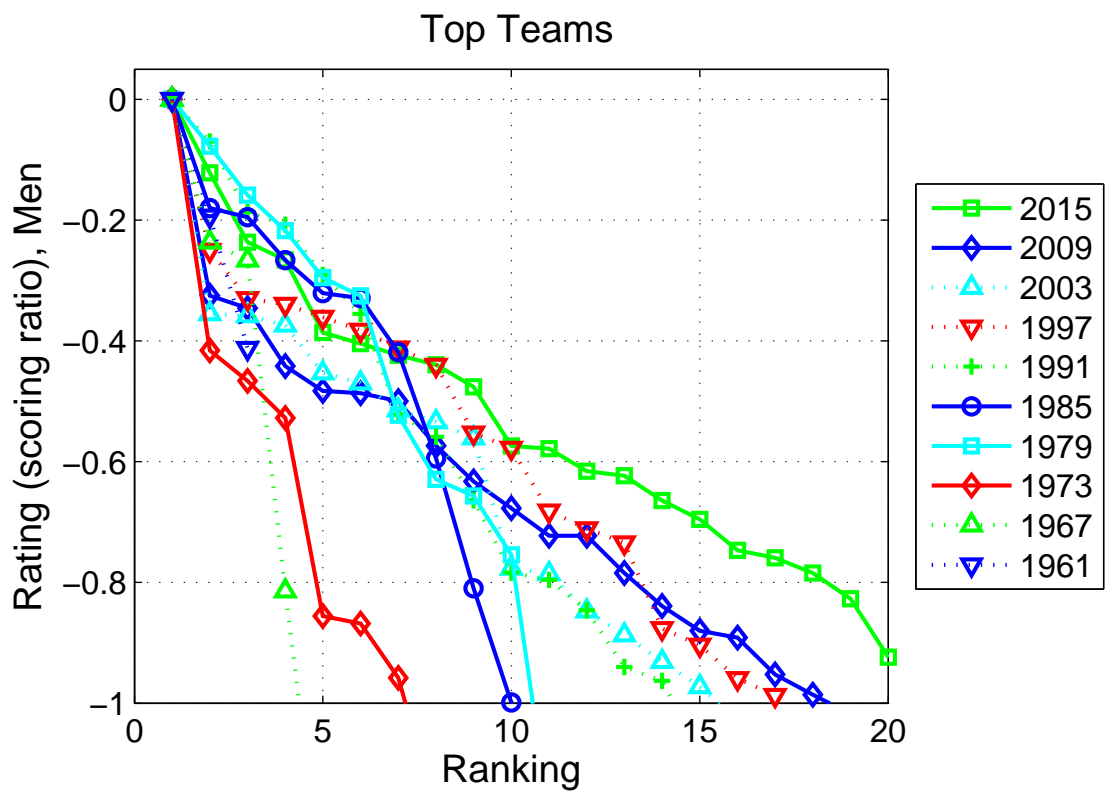


图 6: Rating distribution (scoring rate, men)

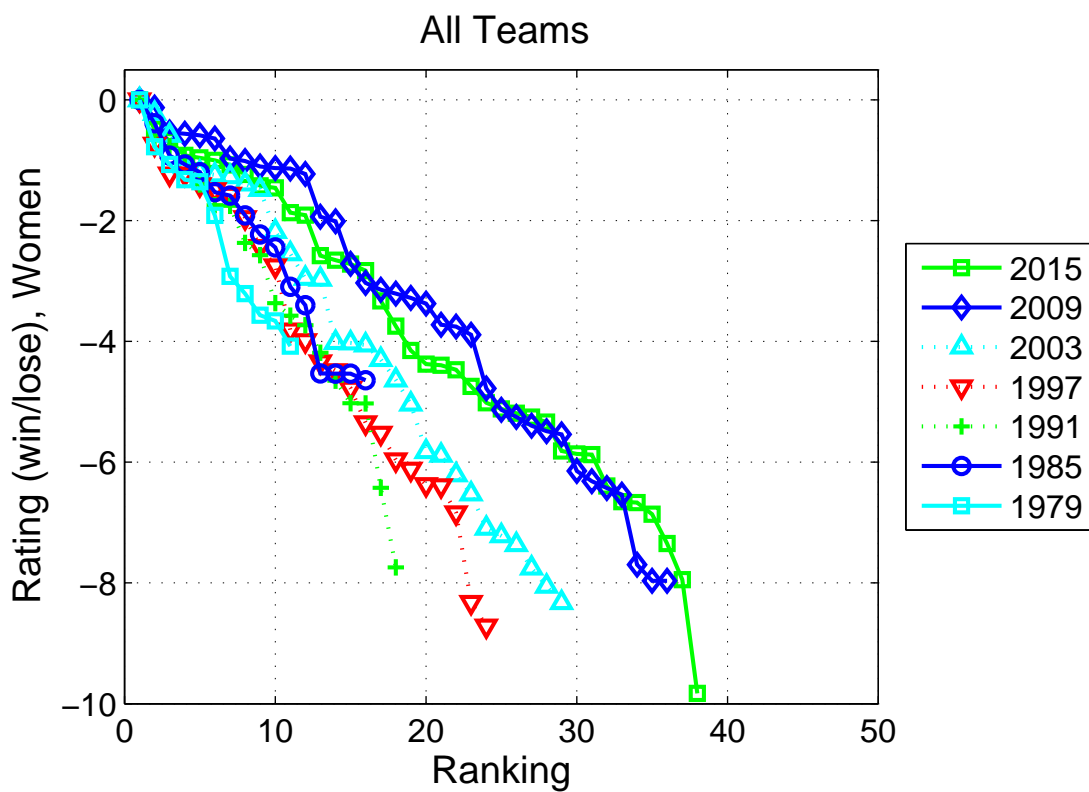
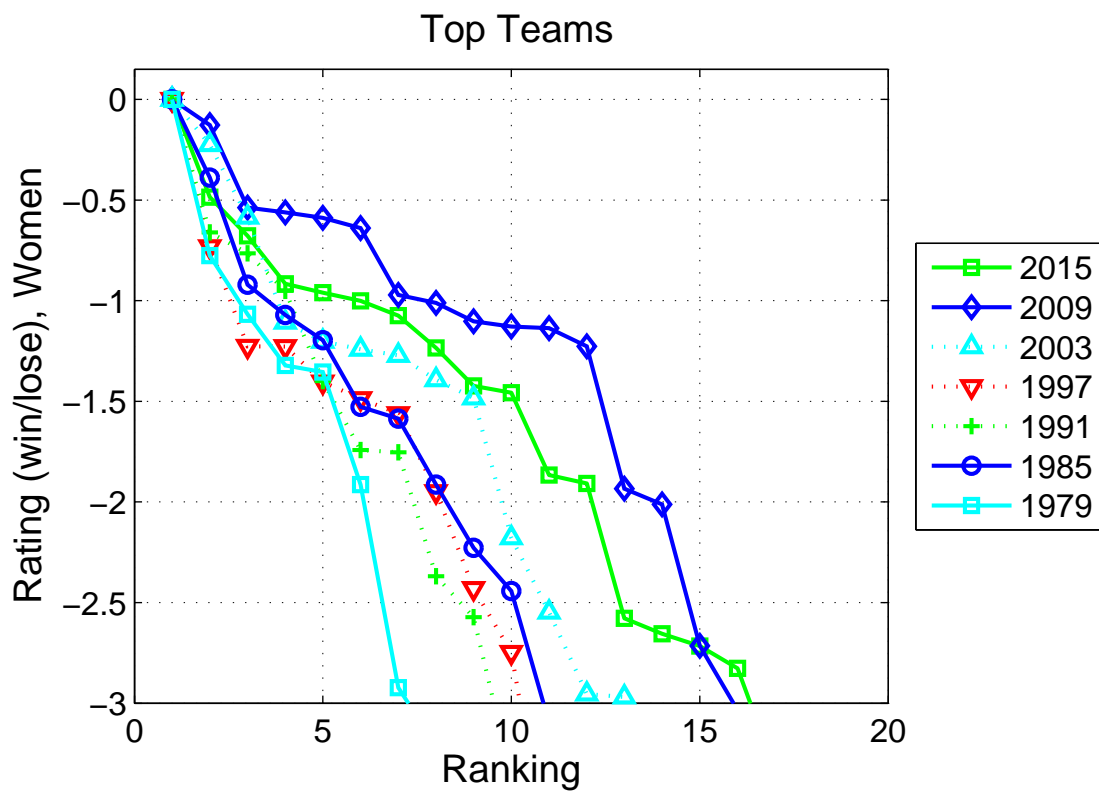


Figure 7: Rating distribution (win/lose, women)

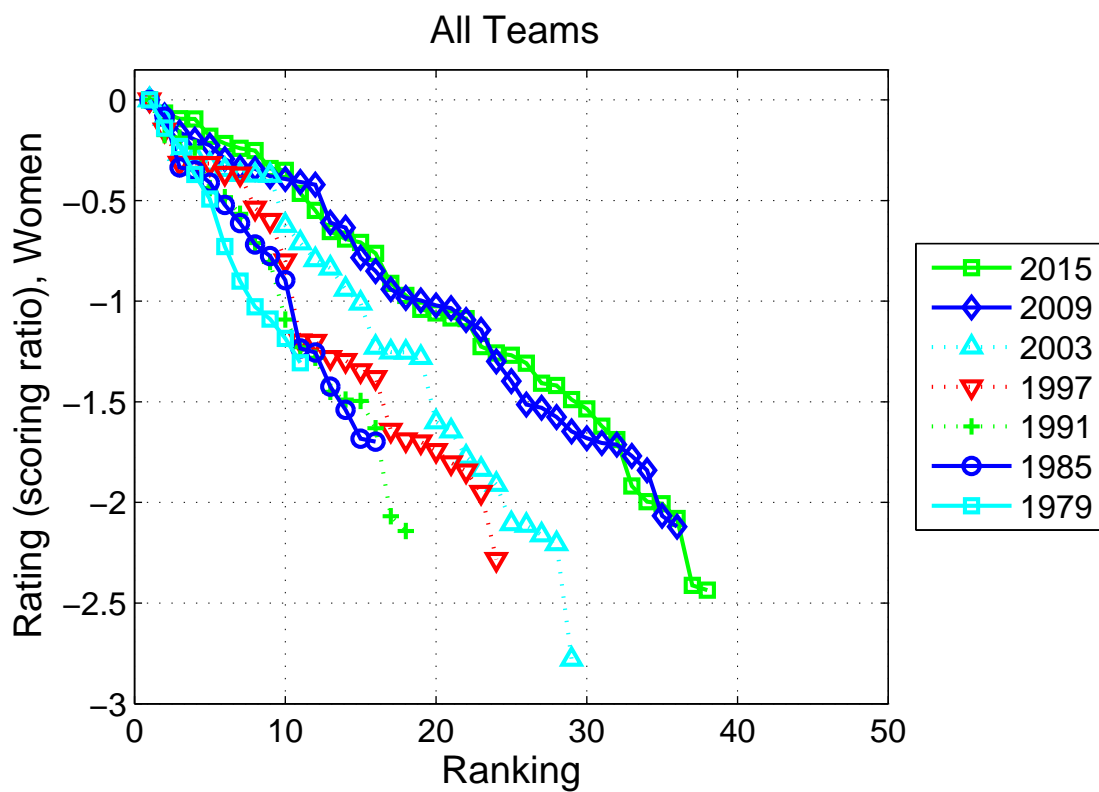
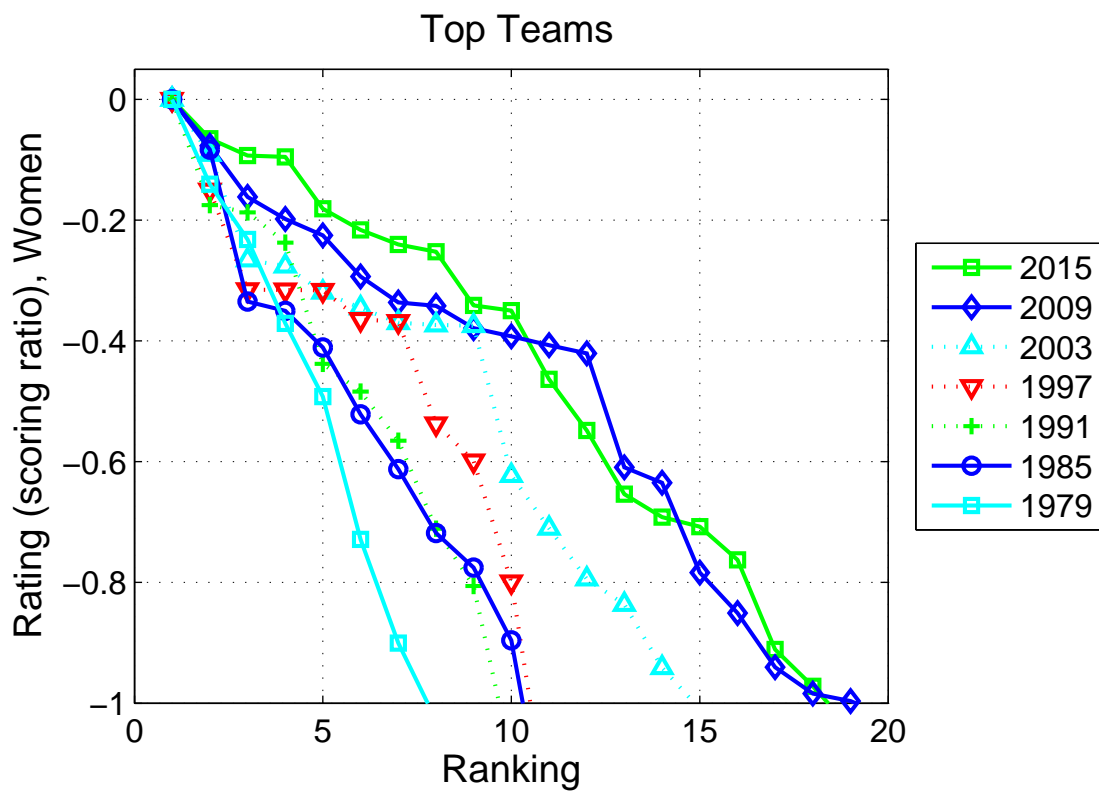


图 8: Rating distribution (scoring rate, women)

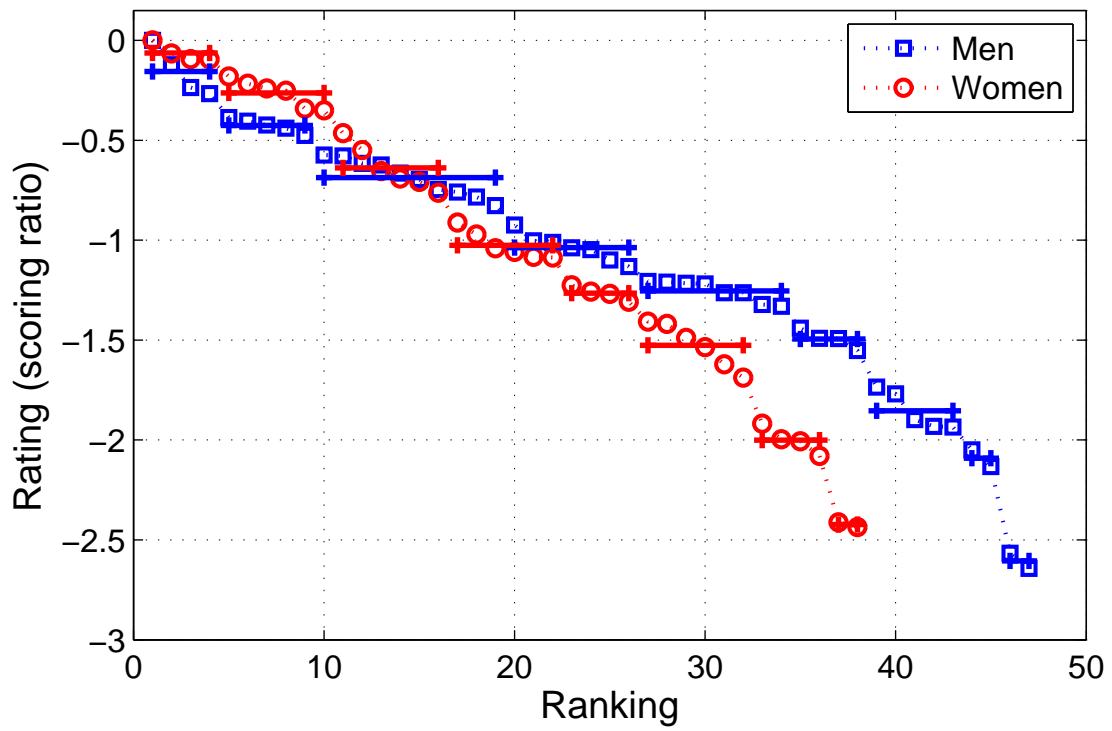
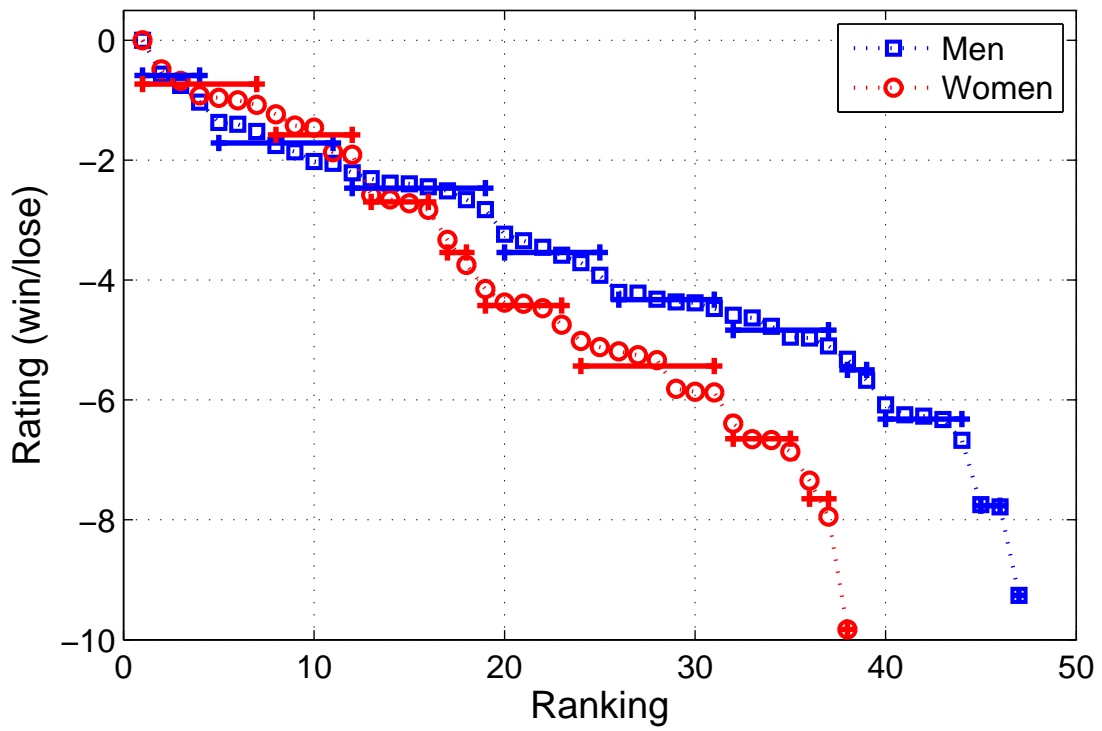
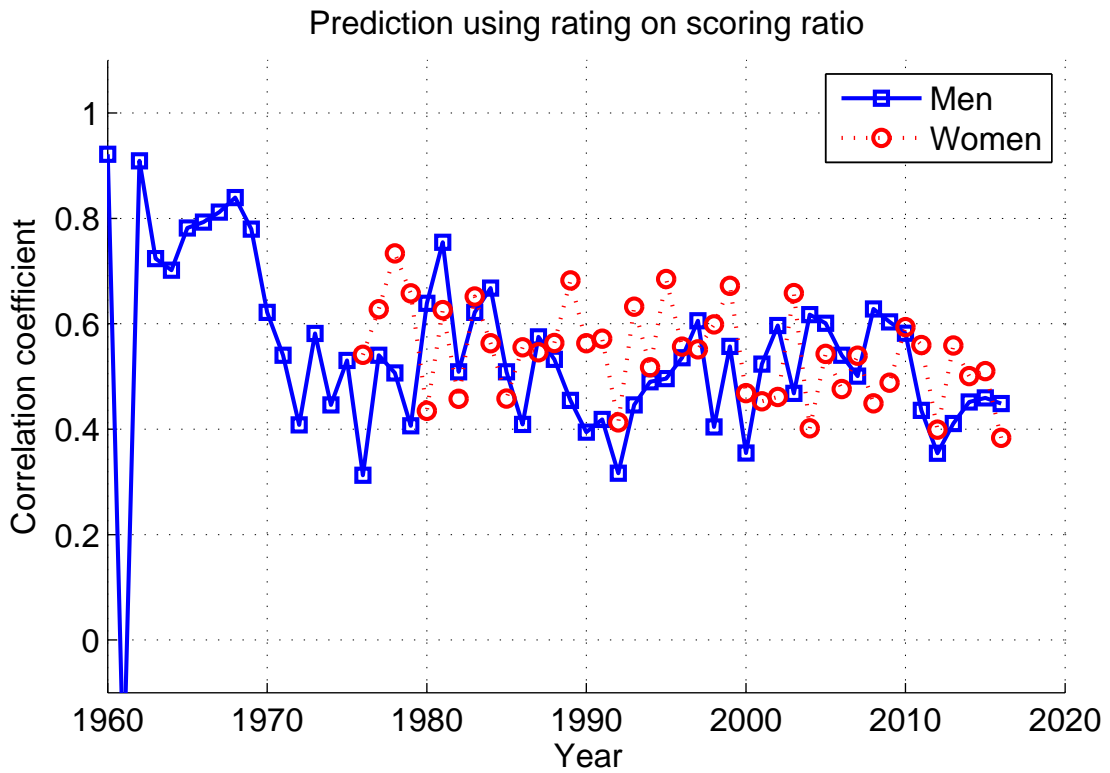
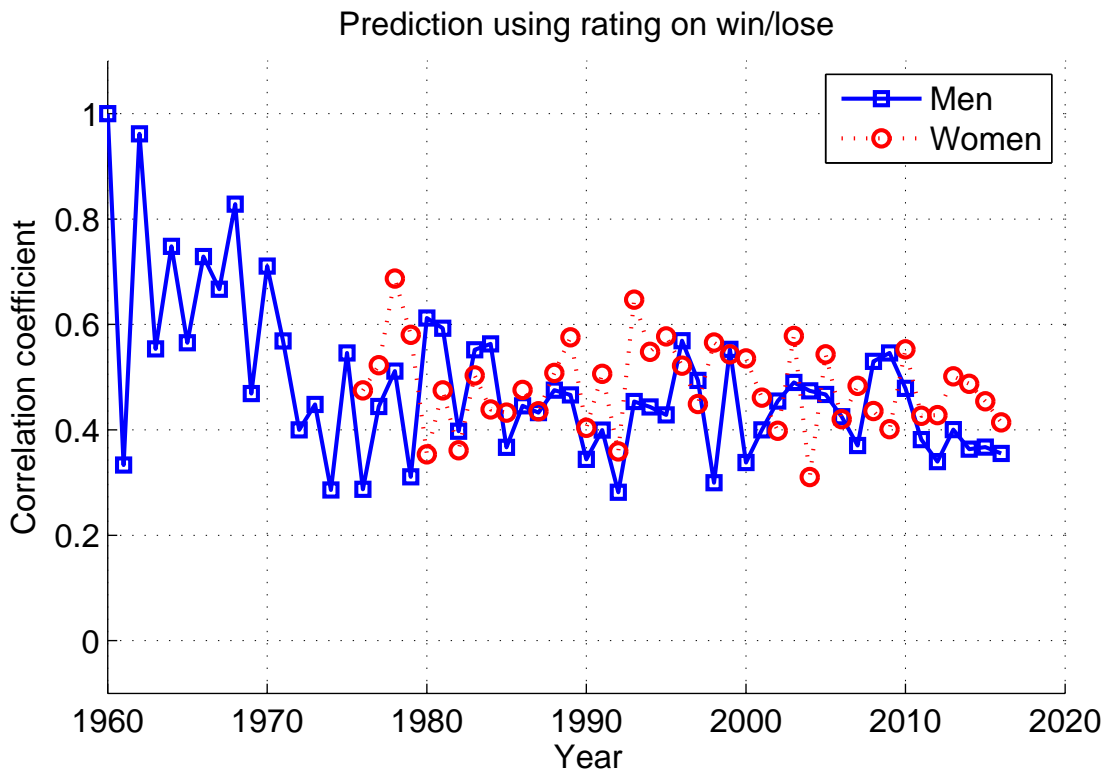


图 9: Rating clustering



☒ 10: Correlation coefficients between predicted and real results

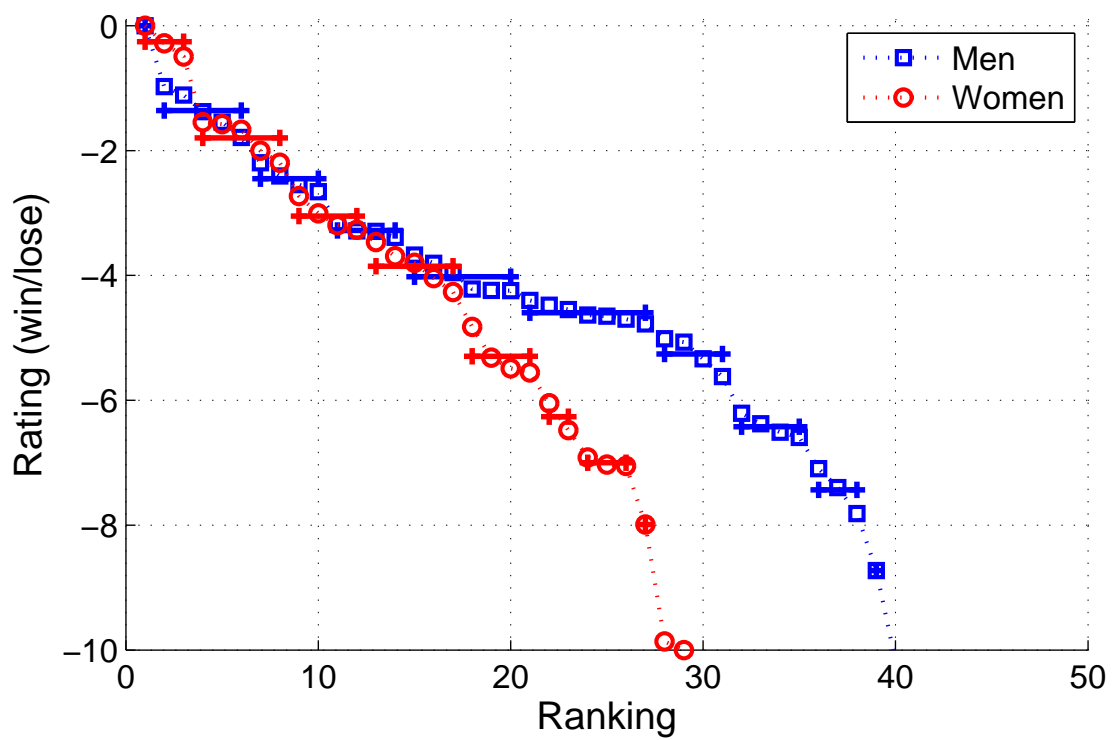


图 11: Rating clustering of volleyball national teams