Hierarchical Bayesian Modeling of English Premier League

Milad Kharratzadeh, Columbia University

Data

- English Premier League, season 2015/2016
- Data from *www.football-data.co.uk*
- 20 teams, 38 weeks, 380 matches
  - Home Team, Away Team, Home Goals, Away Goals

Model – team abilities

- **Team abilities vary over time. For team j:**

$$A_{w,j} \sim \mathbf{N}(A_{w-1,j}, \sigma_{aj}), \qquad w = 2, \ldots, 38$$

Game to game variation for team j

$$\sigma_{aj} \sim \mathbf{N}(0, \tau_a); \quad \tau_a \sim \mathbf{Cauchy}(0, 1)$$

- **Initial team abilities:**

A score in [-1,+1] based on previous season performance

$$A_{1,j} \sim \mathbf{N}(b_{prev} A_{0,j}, \sigma_{a0})$$

$$b_{prev}, \sigma_{a0} \sim \mathbf{N}(0, 1)$$

Model – score differences

$$y_i \sim t_\nu(A_{home\_week(i),\ home\_team(i)} - A_{away\_week(i),\ away\_team(i)} + b_{home}, \sigma_y)$$

Score difference in match i

Possible advantage for the home team

$$b_{home} \sim \mathbf{N}(0, 1); \quad \sigma_y \sim \mathbf{N}(0, 5); \quad \nu \sim \text{Gamma}(2, 0.1)$$

Home effect cannot be too large

Scores can be surprising!

Proposed and analyzed by Juárez and Steel (2010)

Stan code – data

```
data {
  int<lower=1> nteams; // number of teams (20)
  int<lower=1> ngames; // number of games
  int<lower=1> nweeks; // number of weeks
  int<lower=1> home_week[ngames]; // week number for the home team
  int<lower=1> away_week[ngames]; // week number for the away team
  int<lower=1, upper=nteams> home_team[ngames]; // home team ID (1, ..., 20)
  int<lower=1, upper=nteams> away_team[ngames]; // away team ID (1, ..., 20)
  vector[ngames] score_diff;    // home_goals - away_goals
  row_vector[nteams] prev_perf; // a score between -1 and +1
}
```

Stan code – parameters

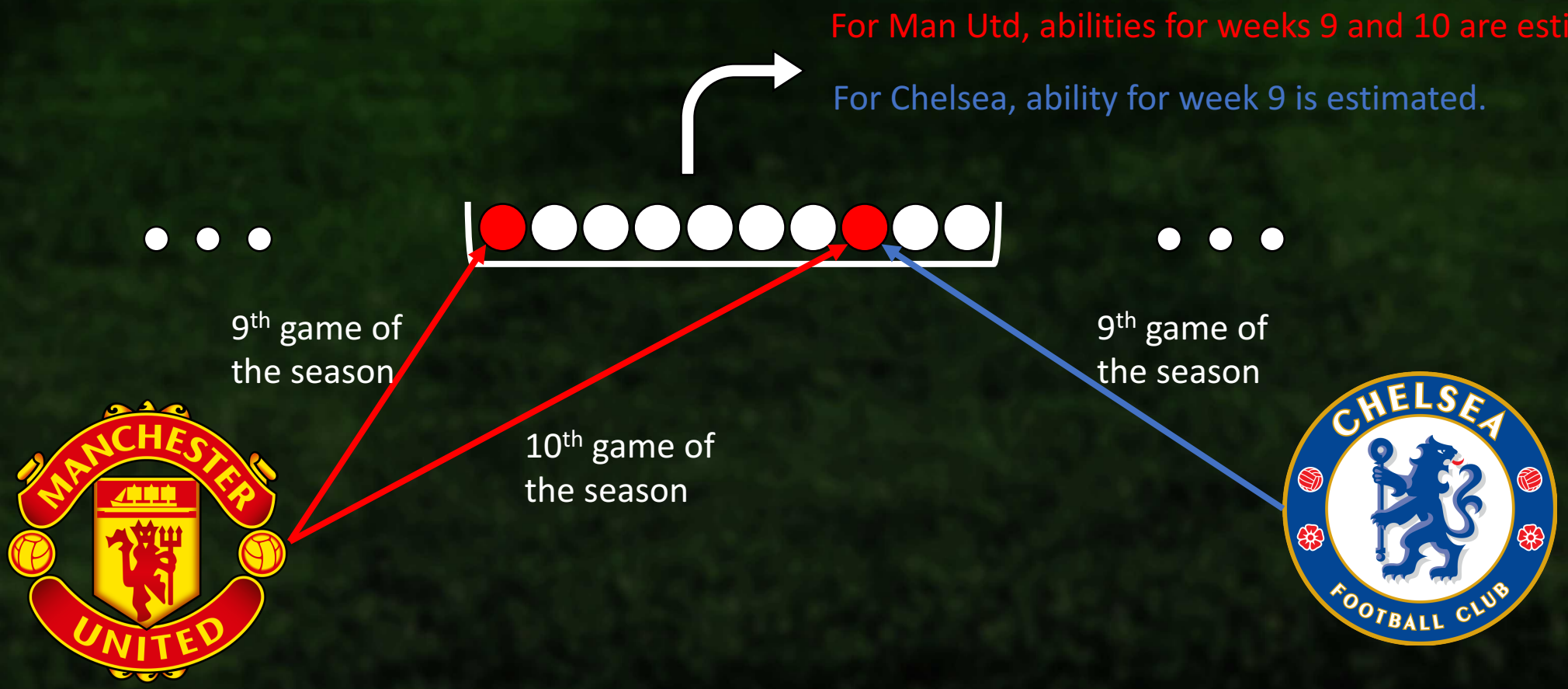
```
parameters {
  real b_home; // the effect of hosting the game in mean of score_diff dist.
  real b_prev;                          // regression coefficient of prev_perf
  real<lower=0> sigma_a0;               // teams ability variation
  real<lower=0> tau_a;                  // hyper-param for game-to-game variation
  real<lower=1> nu;                     // t-dist degree of freedom
  real<lower=0> sigma_y;                // score_diff variation
  row_vector<lower=0>[nteams] sigma_a_raw; // game-to-game variation
  matrix[nweeks,nteams] eta_a;          // random component
}
transformed parameters {
  matrix[nweeks, nteams] a;                       // team abilities
  row_vector<lower=0>[nteams] sigma_a;            // game-to-game variation
  a[1] = b_prev * prev_perf + sigma_a0 * eta_a[1]; // initial abilities (at week 1)
  sigma_a = tau_a * sigma_a_raw;
  for (w in 2:nweeks) {
    a[w] = a[w-1] + sigma_a .* eta_a[w];          // evolution of abilities
  }
}
```

Stan code – model

```
model {
  vector[ngames] a_diff;
  // Priors
  nu ~ gamma(2,0.1);
  b_prev ~ normal(0,1);
  sigma_a0 ~ normal(0,1);
  sigma_y ~ normal(0,5);
  b_home ~ normal(0,1);
  sigma_a_raw ~ normal(0,1);
  tau_a ~ cauchy(0,1);
  to_vector(eta_a) ~ normal(0,1);
  // Likelihood
  for (g in 1:ngames) {
     a_diff[g] = a[home_week[g],home_team[g]] - a[away_week[g],away_team[g]];
  }
  score_diff ~ student_t(nu, a_diff + b_home, sigma_y);
}
```

Fitting the model

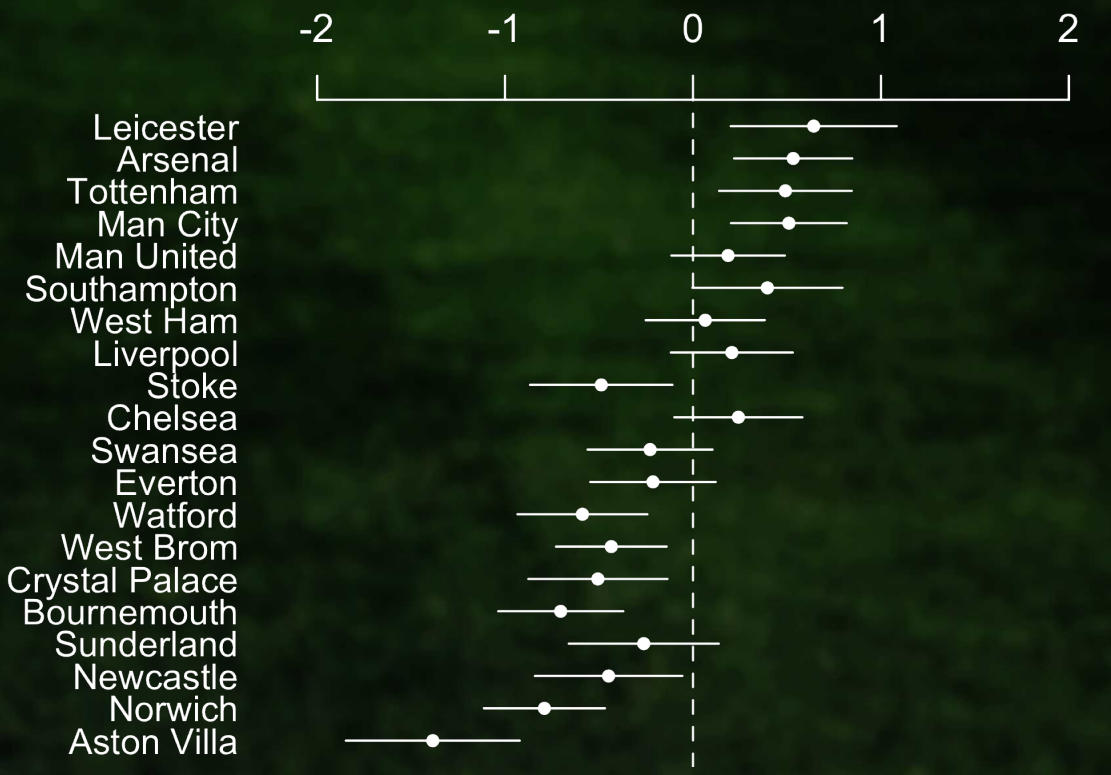
- The model is fitted multiple times – once after every 10 games



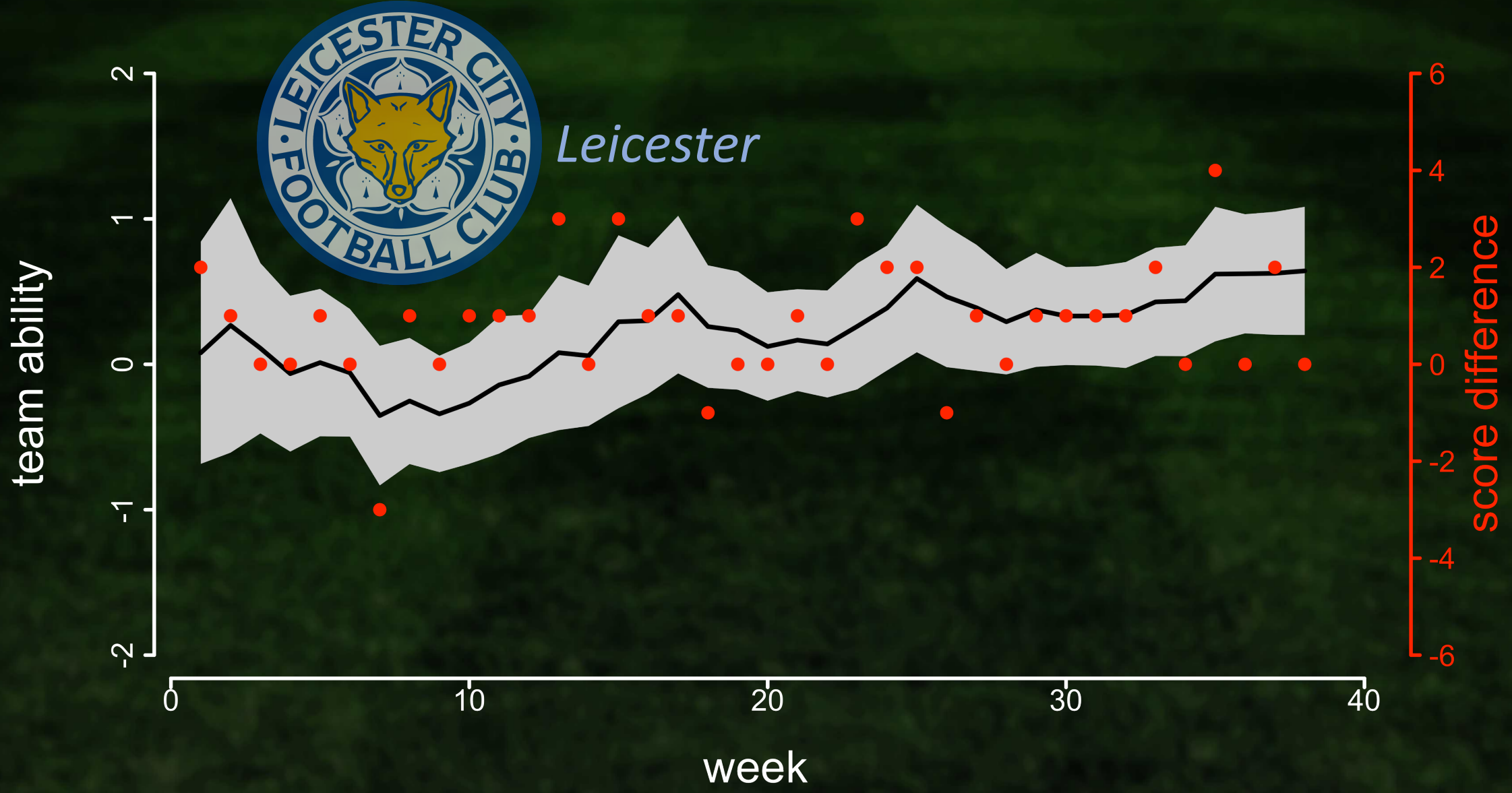
Estimated abilities after week 1 (+/- 1 s.e.)

Teams are sorted according to their performance in previous season (total points)

Estimated abilities after week 38 (+/- 1 s.e.)



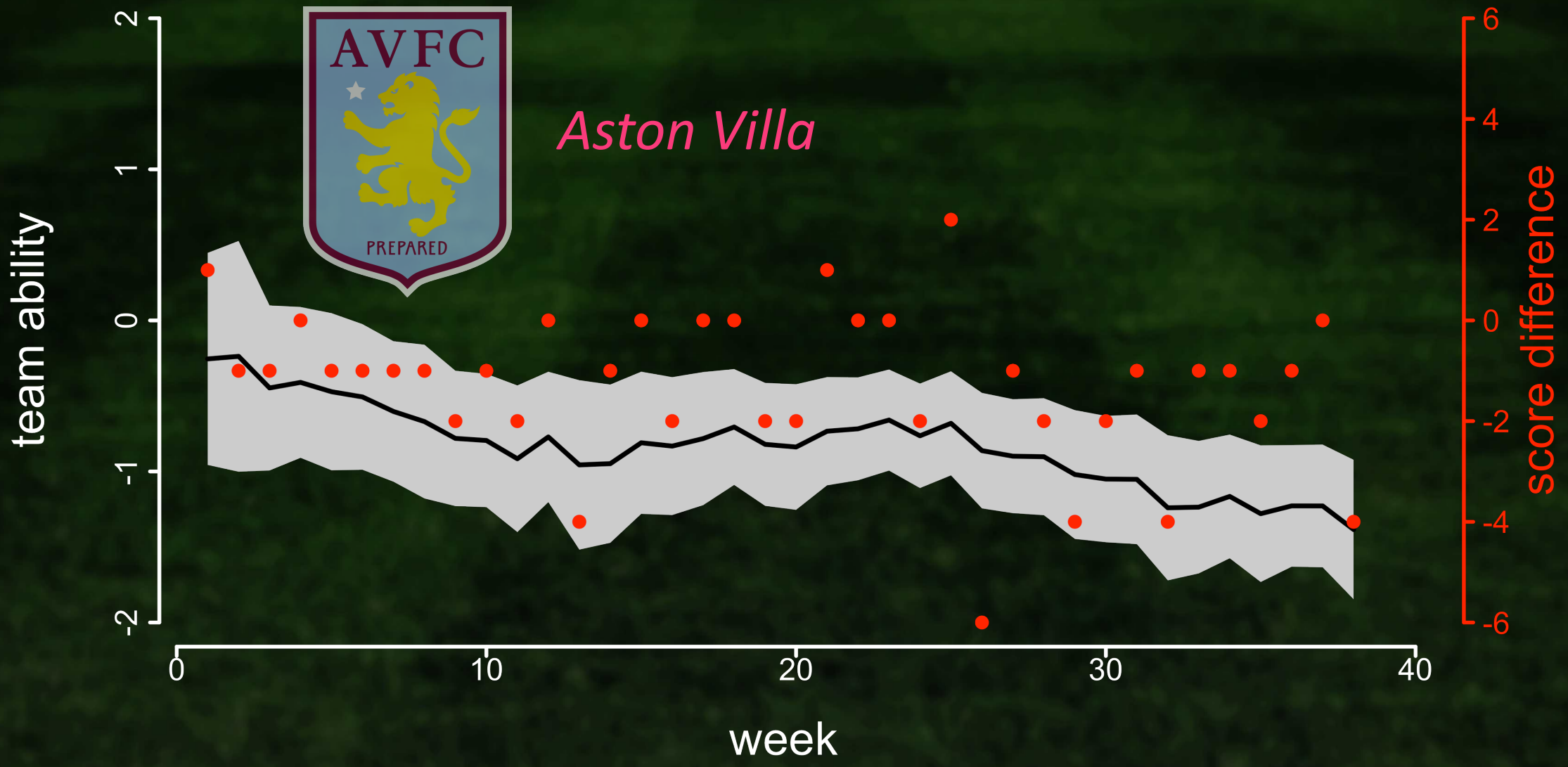
Teams are sorted according to their performance in current season (total points)



Matchday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Ground	H	A	H	A	H	A	H	A	A	H	A	H	A	H	A	H	A	A	H	H	A	A	H	H	A	A	H	H	A	H	A	H	A	H	H	A	H	A
Result	W	W	D	D	W	D	L	W	D	W	W	W	W	D	W	W	W	L	D	D	W	D	W	W	W	L	W	D	W	W	W	W	W	D	W	D	W	D
Position	1	1	1	3	2	3	6	4	5	5	3	3	1	2	1	1	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table from Wikipedia

Aston Villa



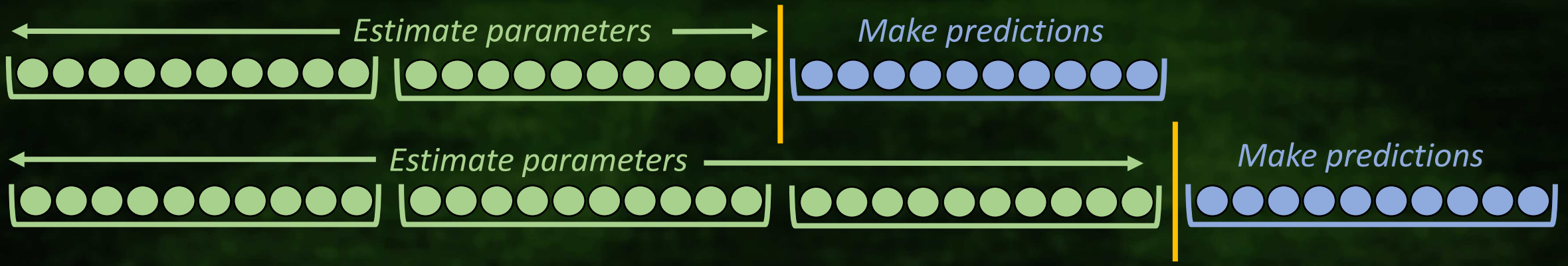
Matchday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Ground	A	H	A	H	A	H	A	H	A	H	A	H	A	H	A	H	A	H	A	A	H	H	A	A	H	H	A	H	A	H	A	H	H	A	H	A	H	A
Result	W	L	L	D	L	L	L	L	L	L	L	D	L	L	D	L	D	D	L	L	W	D	D	L	W	L	L	L	L	L	L	L	L	L	L	L	D	L
Position	5	7	12	12	15	17	18	18	18	19	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20

Table from Wikipedia

Parameter estimates after week 38

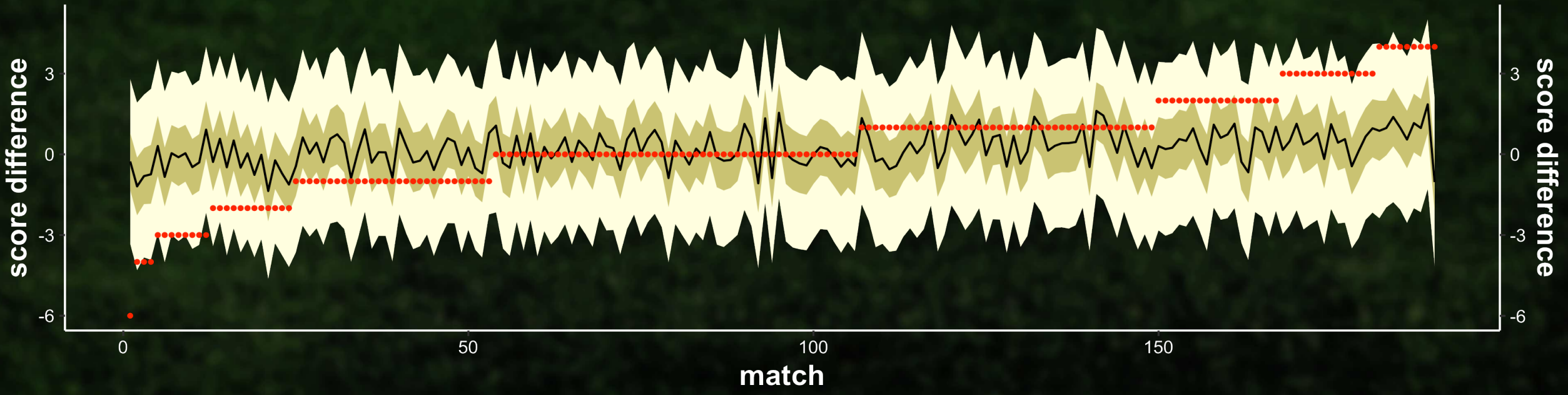
parameter	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
b_{home}	0.28	0.00	0.08	0.13	0.22	0.28	0.33	0.43	1500	1.00
b_{prev}	0.59	0.01	0.18	0.25	0.47	0.59	0.71	0.94	963	1.00
σ_{a0}	0.28	0.01	0.14	0.03	0.19	0.29	0.37	0.56	284	1.02
τ_a	0.08	0.00	0.04	0.01	0.05	0.08	0.10	0.17	372	1.01
ν	26.4	0.35	13.39	9.53	16.34	23.62	33.00	59.73	1500	1.00
σ_y	1.47	0.00	0.07	1.34	1.43	1.48	1.52	1.60	1500	1.00

Predicting score differences



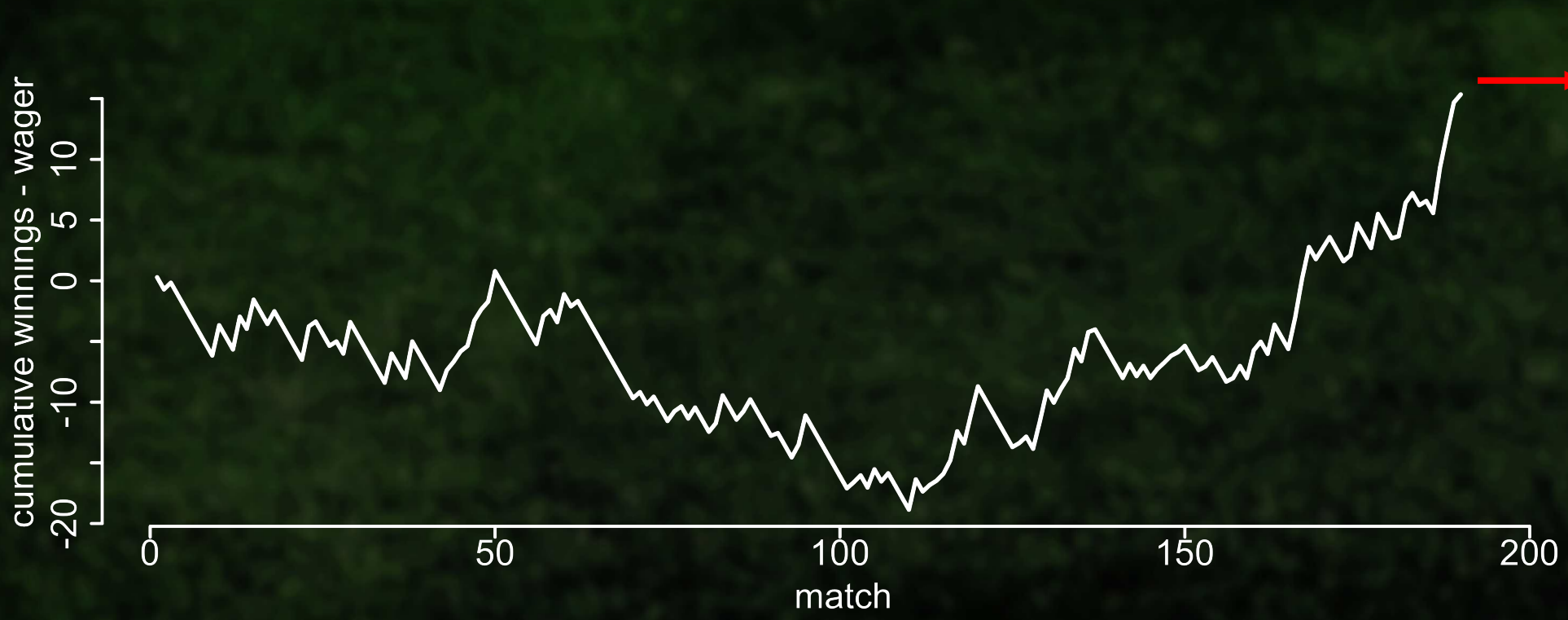
53% of actual values are in the 50% interval. 95% of the actual values are in the 95% interval.

Predictions (black) with 95% intervals (light yellow), 50% intervals (dark yellow), and the actual score differences (red)



Predicting score differences

- Round the median of predicted score differences to closest integer
- Prediction accuracy (win/lose/draw): 87 out of 190 matches (46%)
- What if we bet?
  - using odds given by *Bet365;* wager $1 for each match

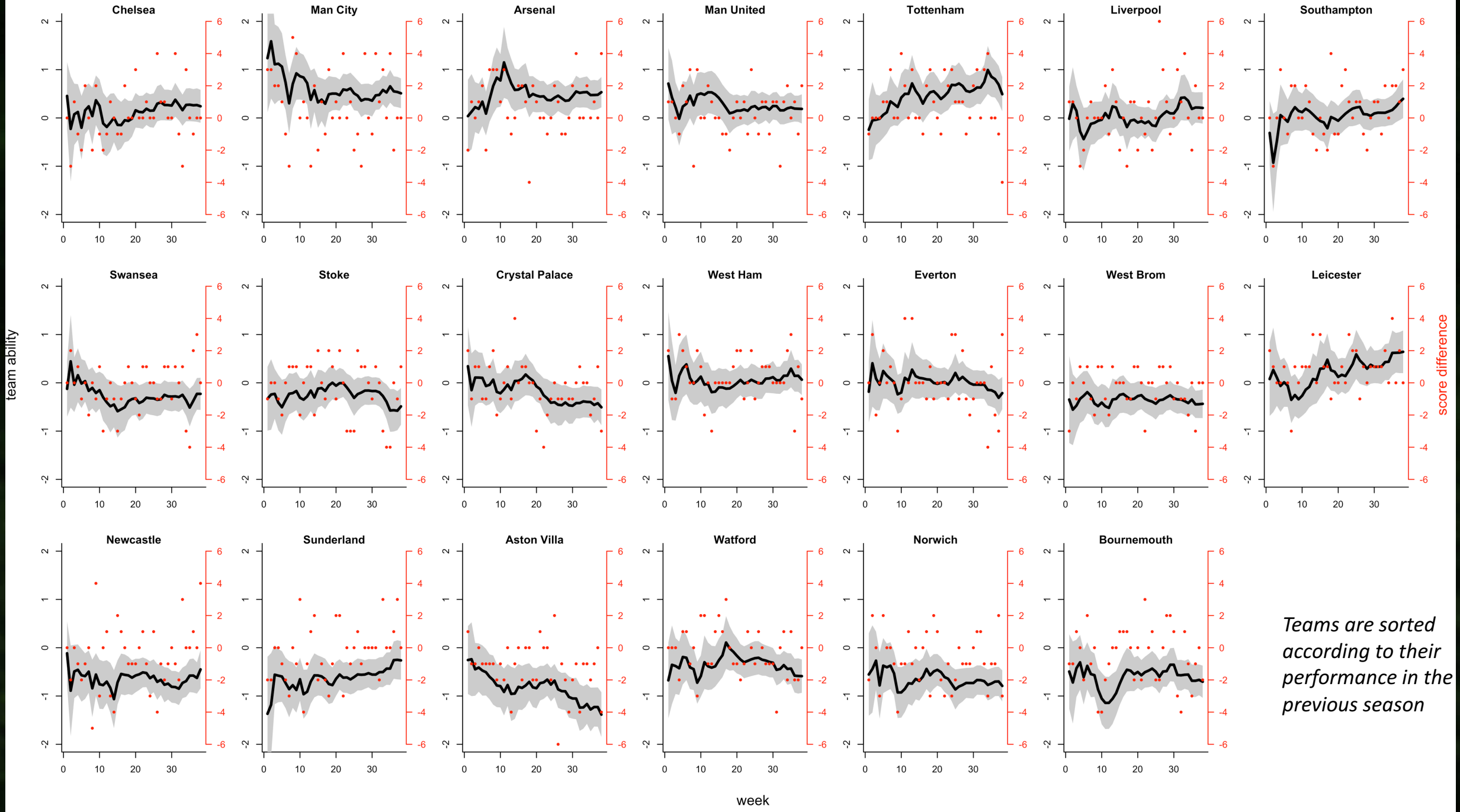


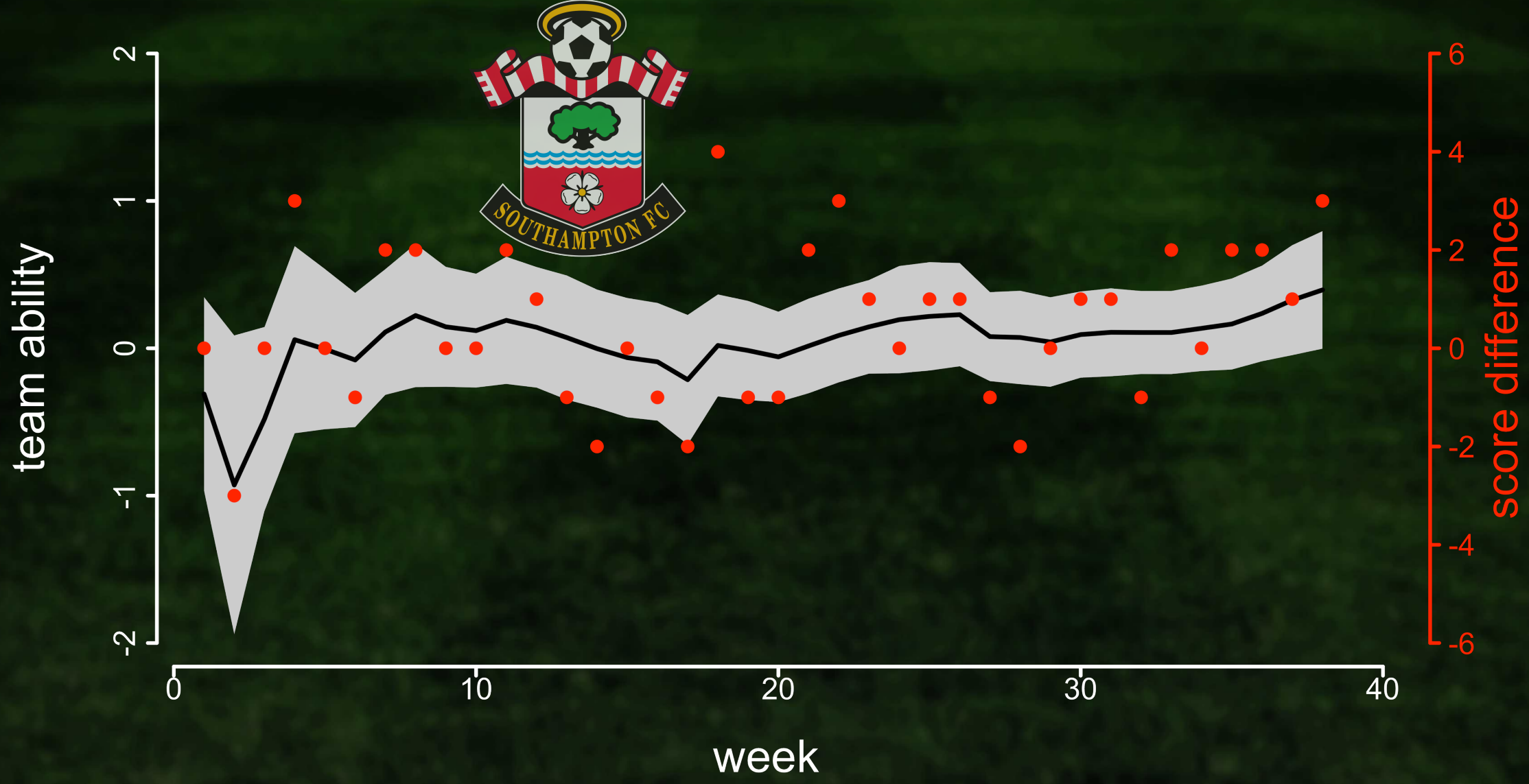
$190 wagered in total
$190+15.3 won
8% return!

Hierarchical Bayesian Modeling of English Premier League

Milad Kharratzadeh, Columbia University

Extra Slides

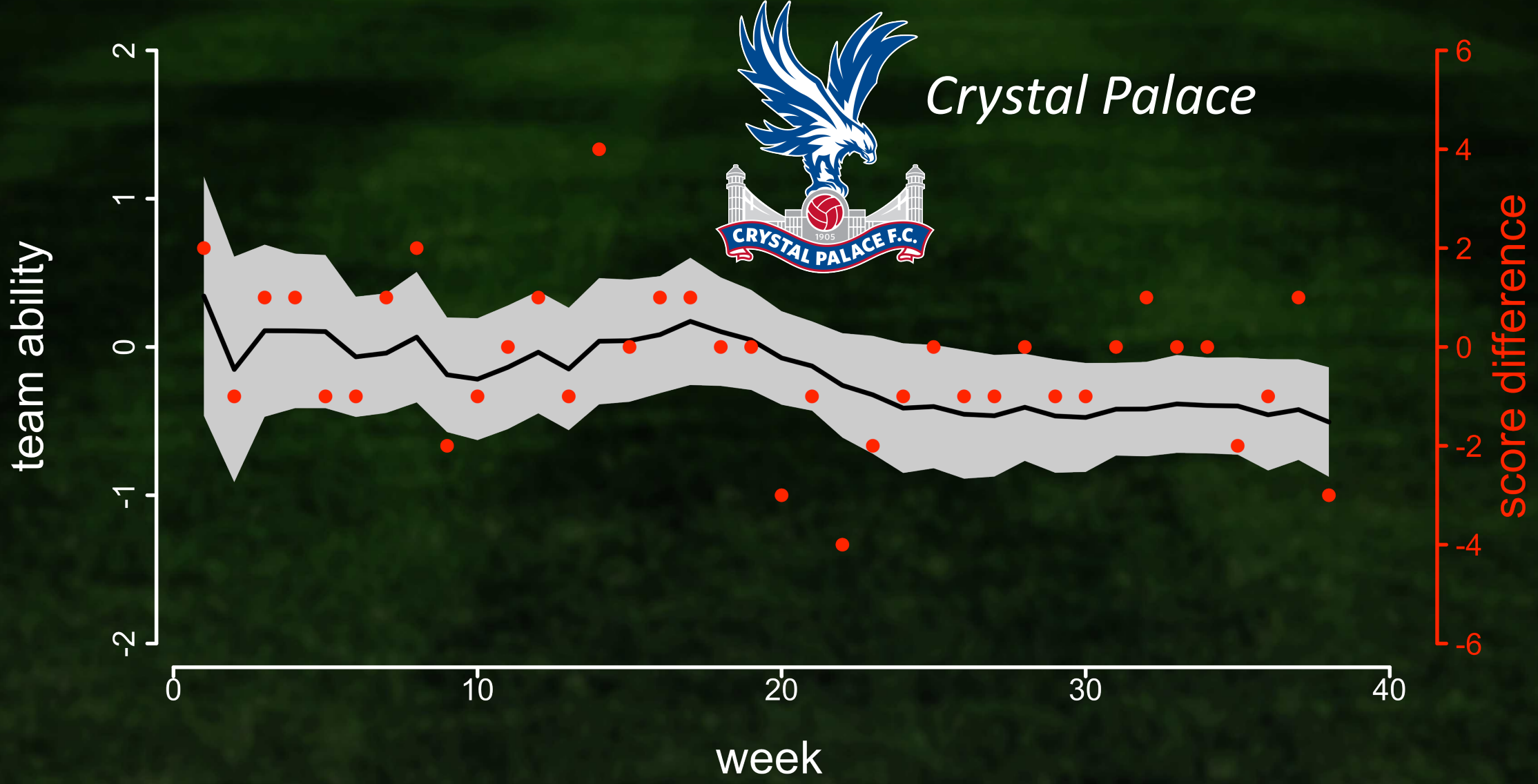

Chelsea
Man City
Arsenal
Man United
Tottenham
Liverpool
Southampton
Swansea
Stoke
Crystal Palace
West Ham
Everton
West Brom
Leicester
Newcastle
Sunderland
Aston Villa
Watford
Norwich
Bournemouth
team ability
score difference
week
Teams are sorted according to their performance in the previous season



Matchday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Ground	A	H	A	H	A	H	H	A	H	A	H	A	H	A	H	A	H	H	A	A	H	H	A	A	H	A	H	A	H	A	H	A	H	A	A	H	A	H
Result	D	L	D	W	W	L	W	W	D	D	W	W	L	L	D	L	L	W	L	L	W	W	W	D	W	W	L	L	D	W	W	L	D	D	W	W	W	W
Position	10	16	18	10	11	16	9	9	8	8	7	7	8	10	12	12	12	12	12	13	12	10	8	7	7	6	7	7	8	7	7	7	7	7	8	7	6	6

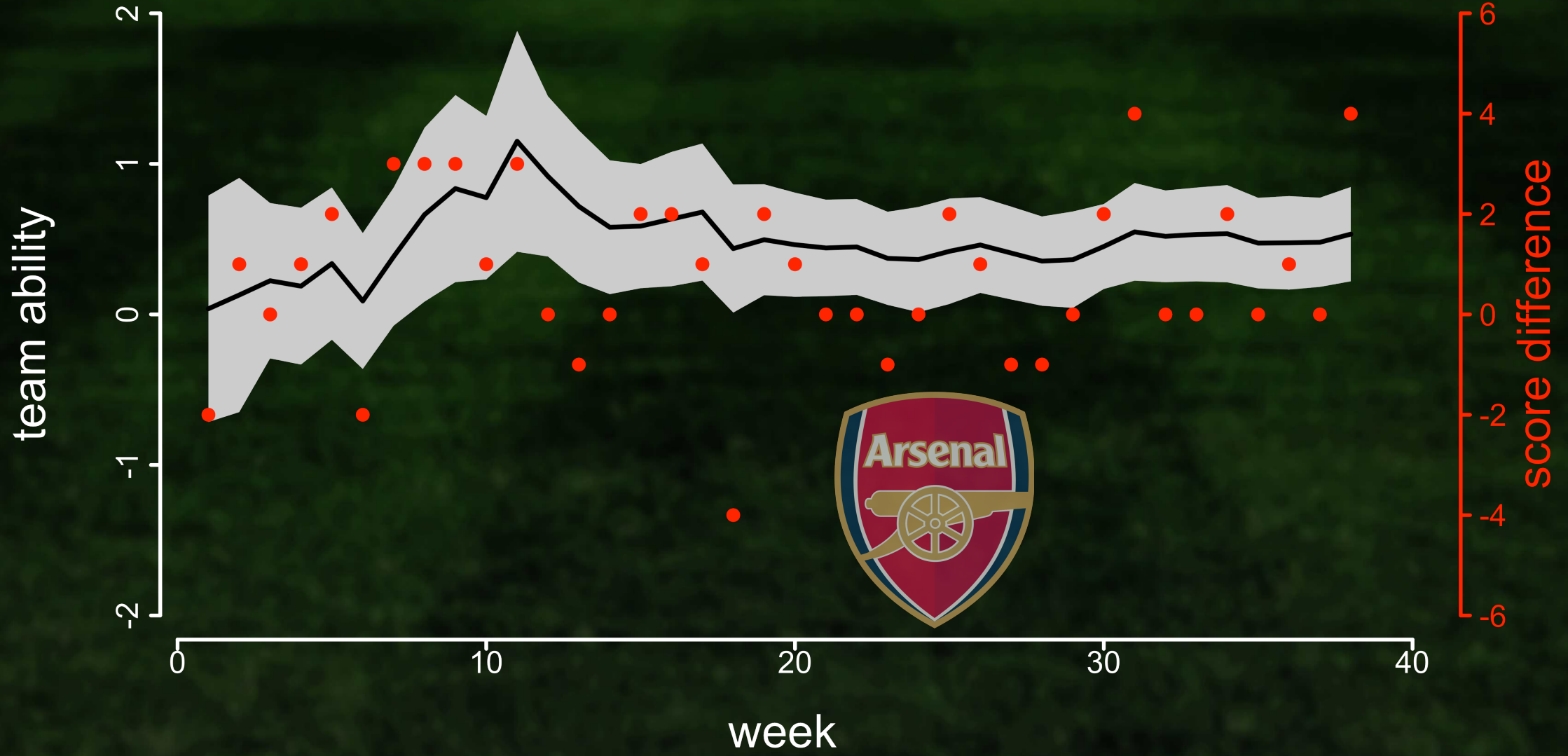
Table from Wikipedia

Crystal Palace



Matchday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Ground	A	H	H	A	H	A	A	H	H	A	H	A	H	H	A	H	A	A	H	H	A	A	H	H	A	H	A	A	H	H	A	H	H	A	A	A	H	A
Result	W	L	W	W	L	L	W	W	L	L	D	W	L	W	D	W	W	D	D	L	L	L	L	L	D	L	L	D	L	L	D	W	D	D	L	L	W	L
Position	3	6	4	2	4	8	6	4	5	7	10	8	10	6	6	6	6	5	5	7	7	8	11	11	12	13	14	14	15	16	16	16	16	16	16	16	14	15

Table from Wikipedia



Matchday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Ground	H	A	H	A	H	A	A	H	A	H	A	H	A	A	H	A	H	A	H	H	A	A	H	H	A	H	A	H	A	A	H	A	H	H	A	H	A	H
Result	L	W	D	W	W	L	W	W	W	W	W	D	L	D	W	W	W	L	W	W	D	D	L	D	W	W	L	L	D	W	W	D	D	W	D	W	D	W
Position	20	11	9	5	3	5	4	2	2	1	2	2	4	4	2	1	2	2	1	1	1	1	3	4	3	3	3	3	3	3	3	3	4	3	4	3	3	2