

Judging Messi's and Ronaldo's scoring ability in different situations according to the model

Xinrui Chen^{1,3,4}, Yufei Tang²

¹University of California Davis, Davis, 95616, USA, cxrchen@ucdavis.edu

²Stony Brook University, New York, 11794, USA, yufei.tang@stonybrook.edu

³Correspondence author

⁴cxrchen@ucdavis.edu

Abstract. In the past decade, two players have overshadowed others in soccer. Who is better, Lionel Messi or Cristiano Ronaldo, has been debated for over a decade. Unlike basketball, the low-scoring nature of soccer dictates that one usually cannot visually conclude the game. Most people discuss who shines in terms of statistics, but there is no way to know the goal-scoring preferences of either man. This paper explores the goal-scoring ability of the two men in different situations to prove who is more complete based on the goal-scoring records of the 2020-2021 season and the data required for expected goals (xG). The study results prove that Messi is more dominant with long-range shots, and Ronaldo scores goals in all visible ranges. This paper introduces a new method of comparing Messi and Ronaldo and uses it as an example to develop a comparison that applies to all players.

Keywords: Lionel Messi, Cristiano Ronaldo, Soccer, Numerical model.

1. Introduction

Technology in soccer has been booming in recent years. Players' skills and teams' tactics are more efficient than they were in the last century, and a lot of new technology has been generated to aid the game. For example, the Video Assistant Referee (VAR) system, introduced to the game in 2018, has profoundly impacted the game's fairness. "The introduction of VAR has undoubtedly had a significant impact on the game, with decisions now being made with greater accuracy and consistency than ever before" [1]. A decade ago, due to the small-scoring nature of soccer, the only way to get a player's performance was through statistics, such as the number of goals or assists. But in this case, there was no shortage of players who were nowhere near as good as their stats. Maybe they scored only two goals out of twenty shots, which is a worse performance than a player who scored one goal out of five, but statistically, the former is better. So, "With the emergence of new technologies, soccer clubs can now analyze a vast amount of data to gain insights into player performance, enabling them to make more informed decisions when it comes to player recruitment and team selection" [2].

Expected goals (xG) is a metric used to assess the chance of a shot resulting in a goal which is helpful for coaches and practitioners has been used frequently in recent years to indicate a possible scoring chance on one image and is used in certain other sports (like hockey) [3]. The xG of a shot on goal is usually a number between [0,1], and if a shot on goal is scored, then its xG is 1. A more significant number indicates a higher probability of an attainable goal. Although xG looks like a

probability, the media often measures a player's xG performance over multiple games or a team's xG combined, so xG is more like a mathematical expectation. There is no particular uniform definition of xG. While xG may seem like a metric only data analysts care about, xG has emerged and is accepted by the media. Some websites that publish game results and simple statistics often publish xG data for players and teams."xG has revolutionized the way we analyze soccer matches, providing a more accurate representation of a team's attacking prowess" [4].

This paper explores the ability of Messi and Ronaldo to score each goal within the sample based on the data required for the xG model to derive the dominance of the two in different situations.

2. Material selection

This paper chose the 2020-2021 season with a similar number of goals scored by both. Messi scored 38 goals, and Ronaldo scored 36. The data was taken from the website "Transfermarkt" and Youtube videos of both players' plans.

2.1. Data collection

1. Coordinates on the field: "The length of a standard soccer field is 100-110 meters long and 64-75 meters wide. The four corners of the field form a rectangular shape with two goal posts and a crossbar at each end of the field" [5]. We choose the median value of length and width, 105m and 68m. Draw two diagonal lines of the field, with the point of intersection as the origin, as shown in Figure 1. The length is the x-axis, and the width is the y-axis, taking absolute values (no negative numbers). For example, the coordinates of the center point of the goal are (52.5, 0)

2. Shot distance: Meters between the shot location and the central point of the goal line [6]. Using the found x and y coordinates, the straight-line distance between the shot location and the leading end of the goal line is found using the Pythagorean theorem.

3. The shooting area is divided into the customary foot, the non-habitual foot, and other body parts. "The player's preferred foot is identified as the foot with which they have scored the majority of their goals in the past, while the non-preferred foot is the opposite." [7]. For example, Messi's dominant foot is the left foot, while the opposite is true for the non-dominant foot. Other parts of the body are the head or any part of the body other than the foot (hands also count, although they are against the rules, but can be included as long as the referee did not notice and scores, see On June 22, 1986, Maradona handballed the ball into England's goal during the World Cup quarterfinal match between Argentina and England in Mexico)

4. Visible range of the goal: "The visible range of the goal is defined as the angle between the two goalposts as seen from the ball's position at the time of the shot." [8], as shown in Figure 2. It can be found by the law of cosine, which requires knowing the triangle's three sides. It is known that the goal is 7.3m long, and the line connecting the goalposts can be found using the Pythagorean theorem, and the angle is finally obtained.

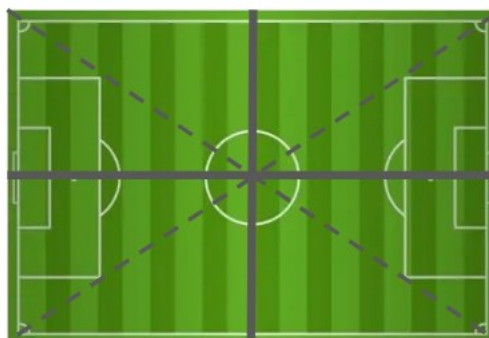


Figure1. Soccer field size.

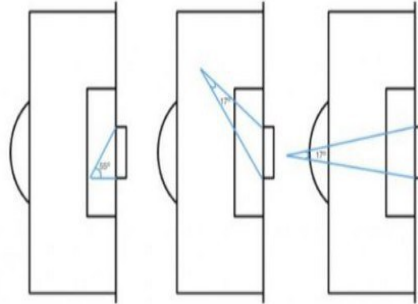


Figure 2. from “Fitting Your Football XG Model · Dato Futbol.”

2.2. Specific data for two players

Table 1 and Table 2, these two tables detail the particular data of Messi and Ronaldo.

Table 1. Messi’s data.

X	Y	Distance	accurate	V.Angle	accurate	Body Part	Half	Duration	Counter	Note
42.5	-12.2	15.7747	15.77	18.96188	18.96	LF	2nd	47	n	
39.3	-9.6	16.3218	16.32	22.86638	22.86	RF	1st	30	n	
38.8	-6.4	15.1212	15.12	27.26662	27.26	LF	2nd	69	n	
37.8	-6.6	16.1137	16.11	25.76472	25.76	RF	1st	37	n	
34.2	-3.9	18.711	18.71	23.66266	23.66	LR	2nd	88	y	
32.8	-4.4	20.1854	20.18	21.93	21.92	LR	1st	12	n	
41.5	2.5	11.2805	11.28	38.34439	38.34	RF	1st	12	n	
42.1	3.9	11.1072	11.10	37.77177	37.77	RF	2nd	58	n	
39.0	3.7	13.9979	13.99	30.97218	30.97	RF	2nd	48	n	
38.6	4.3	14.5499	14.54	29.60596	29.60	LR	1st	44	n	
44.5	12.3	14.6728	14.67	17.8039	17.80	RF	2nd	91	n	
40.7	11.1	16.2003	16.20	20.9588	20.95	RF	2nd	87	y	
39.9	10.2	16.2111	16.21	22.21647	22.21	LF	1st	31	n	
48.5	-2.9	4.94065	4.94	75.27495	75.27	RF	1st	41	n	
47.9	-0.4	4.61736	4.61	81.77401	81.77	H	2nd	68	n	
49.1	0.2	3.40588	3.40	99.18884	99.18	RF	1st	14	n	
49.3	1.6	3.57771	3.57	97.12502	97.12	H	2nd	78	n	
49	3.5	4.94975	4.94	73.11321	73.11	H	2nd	88	n	
48.3	3.8	5.66392	5.66	64.42556	64.42	H	1st	37	y	
45.8	-1.8	6.93758	6.93	59.05983	59.05	RF	1st	23	n	
45.5	-0.5	7.01783	7.01	59.30028	59.30	H	1st	9	n	
44.6	0.5	7.91581	7.91	53.56188	53.56	H	1st	25	n	
44.7	3.1	8.39345	8.39	48.89218	48.89	H	2nd	45	n	
40.2	0	12.3	12.30	36.02939	36.02	RF				Penalty
40.2	0	12.3	12.30	36.02939	36.02	RF				Penalty
40.2	0	12.3	12.30	36.02939	36.02	RF				Penalty
40.2	0	12.3	12.30	36.02939	36.02	RF				Penalty
40.2	0	12.3	12.30	36.02939	36.02	RF				Penalty
40.2	0	12.3	12.30	36.02939	36.02	RF				Penalty
40.2	0	12.3	12.30	36.02939	36.02	RF				Penalty

Table 2. Ronaldo’s data.

X	Y	Distance	(accurate)	V.Angle	(accurate)	Body Part	Half	Duration	Counter	Note
42.5	-12.2	15.77466323	15.77	18.96188001	18.96	LF	2nd	47	n	
39.3	-9.6	16.32176461	16.32	22.86638098	22.86	RF	1st	30	n	
38.8	-6.4	15.1211772	15.12	27.26661941	27.26	LF	2nd	69	n	
37.8	-6.6	16.1136588	16.11	25.76472454	25.76	RF	1st	37	n	
34.2	-3.9	18.71095936	18.71	23.66265616	23.66	LR	2nd	88	y	
32.8	-4.4	20.18539076	20.18	21.92999706	21.92	LR	1st	12	n	
41.5	2.5	11.28051417	11.28	38.34439354	38.34	RF	1st	12	n	
42.1	3.9	11.10720487	11.10	37.77176743	37.77	RF	2nd	58	n	
39.0	3.7	13.99785698	13.99	30.97218498	30.97	RF	2nd	48	n	
38.6	4.3	14.54991409	14.54	29.60596025	29.60	LR	1st	44	n	
44.5	12.3	14.67276388	14.67	17.80390475	17.80	RF	2nd	91	n	

Table 2. (continued).

40.7	11.1	16.20030864	16.20	20.95879666	20.95	RF	2nd	87	y	
39.9	10.2	16.21111073	16.21	22.21647373	22.21	LF	1st	31	n	
48.5	-2.9	4.940647731	4.94	75.27494797	75.27	RF	1st	41	n	
47.9	-0.4	4.617358552	4.61	81.77401391	81.77	H	2nd	68	n	
49.1	0.2	3.405877273	3.40	99.18883777	99.18	RF	1st	14	n	
49.3	1.6	3.577708764	3.57	97.12501801	97.12	H	2nd	78	n	
49	3.5	4.949747468	4.94	73.11321012	73.11	H	2nd	88	n	
48.3	3.8	5.663920903	5.66	64.42555633	64.42	H	1st	37	y	
45.8	-1.8	6.937578828	6.93	59.0598301	59.05	RF	1st	23	n	
45.5	-0.5	7.017834424	7.01	59.30027846	59.30	H	1st	9	n	
44.6	0.5	7.915806971	7.91	53.56187754	53.56	H	1st	25	n	
44.7	3.1	8.393449827	8.39	48.89217517	48.89	H	2nd	45	n	
40.2	0	12.3	12.30	36.02938845	36.02	RF				Penalty

3. Approach

The datasets, including Time, Visible range of the goal, and Shot Distance, are numerical and can potentially be made into some curve-fitting models. Noticing that the data are dimensional, applying a traditional polynomial regression is impossible. “Probability distribution models are useful tools for modeling soccer data, particularly for analyzing the likelihood of scoring in certain game situations” (Wunderlich, 2019) [9]. So we use a probabilistic model. “Probability distribution models can provide a more accurate and nuanced understanding of soccer data, particularly when dealing with complex, multi-dimensional datasets” [10]. To construct such a model, which can summarize the possibilities of something happening in a particular situation, we decide to split the x-axis intervals into equal-length subintervals and define the density of each subinterval to be the number of goals over the total goals. So firstly, define

$$[0, a_n] = [0, a_1) \dots [a_{n-1}, a_n] \quad (1)$$

Denote G_i as the goals within each interval; we then have:

$$1 = \sum_{i=1}^n \frac{G_i}{G_{\text{total}}} \quad (2)$$

The function is then:

$$f(x) = \frac{G_i}{G} \text{ for } a_i \leq x \leq a_{i+1} \quad (3)$$

The area of each “box” is:

$$A = l_i \frac{G_i}{G} \quad l_i = \text{length of each interval} \quad (4)$$

After constructing the function, we found that the resulting graph is not smooth, contradicting our goals to make a model of comparison. It is continuous, though, so we can still compare through the work by integrating the functions and comparing them with the number. But we still want a visualizable comparison model. The basic idea of density is unchanged, but after researching online, we found a way to smooth the graph -- kernel smoothing.

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad [11]$$

Here $K()$ is the kernel smoothing function, n is the sample size, x_i is the values of the samples, and h is the bandwidth. The new model is still looking for densities, but it solves the problem of unsmooth by dropping off the term “for.” By assigning weights on different distances through $K()$, the function enables x to be calculated with the original value of itself. Algebraically the role treats each x as the kernel, which means the distance of the samples to x does not vary from left or right, and the space is the critical part of kernel smoothing. The kernel smoothing function is our desired numeric model, as

shown in the figure. Finally, thanks to Matlab, there is a convenience tool called density. The graphs will be made by this tool through Matlab.

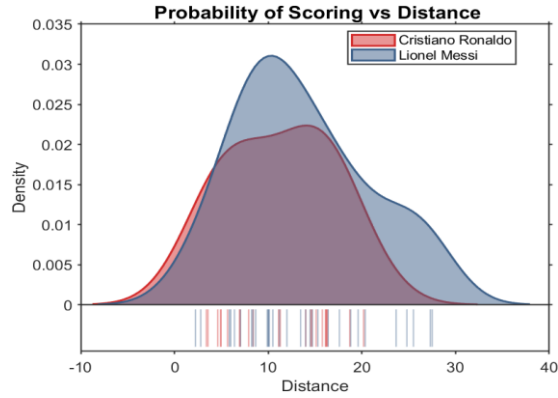


Figure 3. Comparison chart about Shooting Distance.

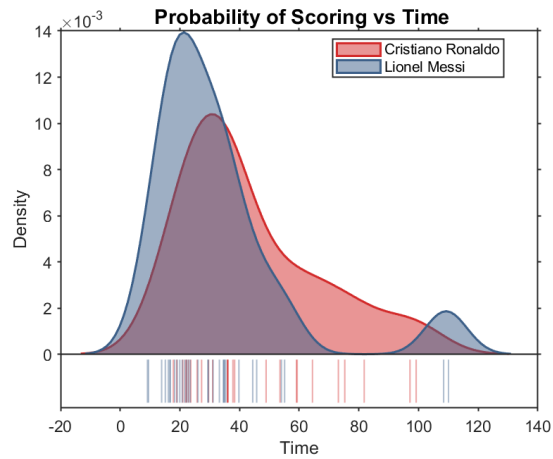


Figure 4. Comparison chart about Game Time.

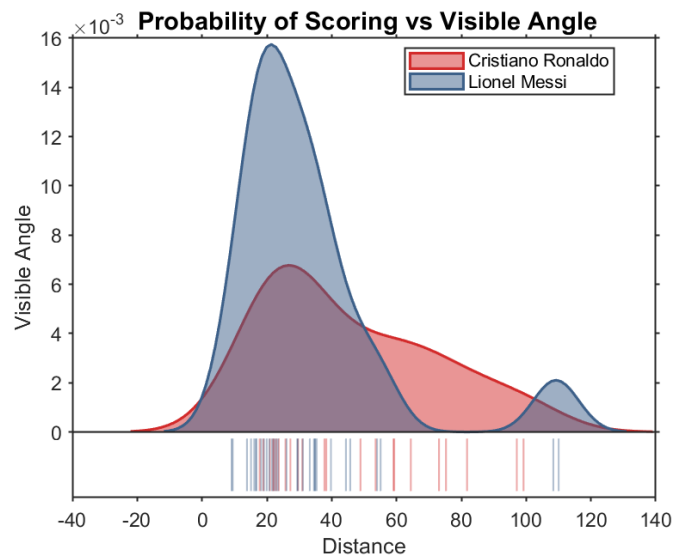


Figure 5. Comparison chart about Shooting visual Angle.

4. Explanation Analysis

Figures 5, 6, and 7 showed several meaningful results. In Figure 5, Lionel Messi shows a distinct advantage compared to Cristiano Ronaldo when the shot is over 20 meters long. In Figure 6, Cristiano Ronaldo shows his strong ability to attack the goal from 60 minutes to 100 minutes of the game. In Figure 7, Cristiano Ronaldo again shows his comprehensive methods to win scores. Also, in the last two figures, we can still see that Lionel Messi is dominantly more decisive than Cristiano Ronaldo when the game comes to his familiar situations. The text only shows the goal preferences of Messi and Crosby in different situations; this model applies to the comparison of any player.

5. Conclusion

In conclusion, this study analyzed the goal-scoring abilities of Lionel Messi and Cristiano Ronaldo based on the 2020-2021 season data and expected goals (xG). Messi demonstrated strength in long-range shots, while Ronaldo displayed versatility in scoring from various situations. The findings provide valuable insights into their respective scoring preferences. This methodology can be extended to compare other players, contributing to player evaluation in soccer. Future research can expand the dataset and employ advanced statistical techniques to enhance the understanding of goal-scoring abilities. Overall, this study contributes to the ongoing debate about Messi and Ronaldo's scoring abilities and demonstrates the application of numerical models for player comparison in soccer.

Reference

- [1] Baker, M. (2021). The impact of Video Assistant Referee (VAR) on the fairness of soccer. *Journal of Sports Sciences*, 39(5), 483-490.
- [2] Kang, S. K., & Lee, S. Y. (2020). The impact of technology on soccer performance analysis. *Journal of Physical Education and Sport*, 20(1), 95-100.
- [3] Rathke, A. (2017). An examination of expected goals and shot efficiency in soccer. *J. Hum. Sport Exer.* 12, 514–529. doi: 10.14198/jhse.2017.12. Proc2.05
- [4] Alexander, Duncan. *How Soccer Analytics Works*. Penguin Random House, 2021.
- [5] Fernández, Javier, and Luke Bornn. "Wide Open Spaces: A Statistical Technique for Measuring Space Creation in Professional Soccer." *Journal of Quantitative Analysis in Sports*, vol. 12, no. 3, 2016, pp. 139-150.
- [6] Ismael Gómez, et al. "Fitting Your Own Football XG Model · Dato Futbol." DATO FUTBOL, 14 Apr. 2020, <https://www.datofutbol.cl/xg-model/>.
- [7] Lucey, Patrick, et al. "A Multi-Scale Approach to Predicting Goals in Soccer." *Journal of Quantitative Analysis in Sports*, vol. 12, no. 4, 2016, pp. 159-168.
- [8] Xu, Qingyang, et al. "Learning to Score in the Wild." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5455-5463.
- [9] Wunderlich, F., & Memmert, D. (2019). Data science and soccer: identifying interesting variables through machine learning techniques. *Current Opinion in Psychology*, 34, 155-159.
- [10] Bialkowski, A., Lucey, P., Carr, P., & Matthews, I. (2014). Probabilistic event forecasting in soccer. In *Proceedings of the 23rd international conference on World Wide Web* (pp. 557-562). <https://www.mathworks.com/help/stats/ksdensity.html>