FOOTBALL DATA ANALYSIS: THE PREDICTIVE POWER OF EXPECTED GOALS

Marco Roccetti
Dept. of Computer Science and Engineering
University of Bologna, Italy
marco.roccetti@unibo.it

Filippo Berveglieri
Dept. of Computer Science and Engineering
University of Bologna, Italy
filippo.berveglieri@studio.unibo.it

Giuseppe Cappiello
Dept. of Management
University of Bologna, Italy
giuseppe.cappiello@unibo.it

KEYWORDS
Football data analytics, Performance football indicators, Expected goals, Data science, Predictive intelligence.

ABSTRACT

Major football events, such as the World Cup, have popularized the European football (i.e., soccer) beyond its traditional geographical borders, for example with a growing number of Americans who have begun to incorporate European football into their daily sport consumption.
Many video game series based on this sport have further contributed to this increasing development. Not only that, but many performance indicators inspired by football computer games have become metrics that are now part of the normal conversation about real football, being used to analyse tactics, teams and players. Among these indicators, of prominent importance is the role of expected goals (xG) that measure the quality of a shot by calculating the likelihood that it will be scored from a particular position during a particular phase of play.
Often, someone talks about xG leading to imprecise conclusions and to a distorded point of view on this performance indicator. In this paper, using a non-parametrical Kolmogorov Smirnov hypothesis testing method, we show that there is no statistically significant difference if, in a given football tournament, the final scores are awarded to the participating teams either with real goals or based on xGs, thus demystifying some wrong assumptions about this indicator. Data for this study come from all the matches played during the 2020 UEFA European Football Championship.

INTRODUCTION

This paper is about the concept of *expected goals* (xG, for short) utilized in football (i.e., soccer) data analysis. More specifically, we developed a study to better understand the role and implications of using this performance indicator for analyzing tactics, players, and even teams, during a football match.
xG measure the quality of a chance before a player shoots, by calculating the likelihood that they will be scored from a given position during a particular phase of play. xG are measured on a scale between 0 and 1, where 0 means that it is impossible to score and 1 represents the fact that a player is expected to score that goal for sure. Many factors have an influence on the value associated to xG and they may vary, depending, not only on the adopted statistical model, but also on metrics, like distance, angle, quality of the player and many others.
While there is some debate about the precise origin of the term "expected goals" (Barnett and Hilditch 1993; Ensum et al. 2004), there are few doubts that xG, as a concept, is grounded in the culture of video game series devoted to football, like for example those discussed in (Parkin 2021).
With the passage of time, the problem, with this football performance indicator, is that it has become part of the normal conversation about football, up to a point where the majority of analysts do not use it appropriately or do not fully understand its implications. For example, it is a common opinion of many football analysts that xG measure the quality of a chance, but there is no agreement on the fact that xG may also have the capacity to predict the expected outcome of a match (Mead et al. 2023; Tweedale 2024).
Our hypothesis, instead, is that, while it is true that on a given match a team could have overperformed or underperformed their xG, nonetheless xG can tell us a lot about the long term performance of that team. In the long run, in fact, teams which are underperforming their xG, typically start to convert more chances into real goals, with the opposite situation for teams which are temporarily overperforming.
We have tried to provide evidence in support of this hypothesis by studying both real goals and xG of the 2020 UEFA European Football Championship which ran from 11 June to 11 July 2021, with 51 footbal matches played by 24 national teams. At the time of writing this paper, this tournament was the one with the most goals per game in a European Championship since the introduction of the group stage, totaling 142 goals, for an average of 2.78 goals per match.
To conduct our analysis, we have first converted into a unique ranking the results of all the matches played by all the 24 national teams, considering the group stage phase and the knockout phase of the tournament all together.
Then, we have used the non-parametrical Kolmogorov Smirnov statistical procedure of hypothesis testing to compare the two rankings, obtained using either real goals or xG. In particular, we conducted a series of eight different Kolmogorov Smirnov tests, showing that there is no statistically significant difference between the two score distributions, obtained, respectively, with real goals and with xG.
In the end, we have got a quite typical result which only a scientific perspective on big data can put in evidence (Roccetti et al. 2021; Roccetti et al. 2008), that is that a meaningful

To cite as: Roccetti, M., Berveglieri, F., Cappiello, G., In Proc. 25th International Conference on Intelligent Games and Simulation, GAME-ON 2024, 2024, Istanbul, pp. 20–24

football performance indicator, like xG, can have, in the long run, a good predictive power (for example, with respect to the overall performance of a team).
The remainder of this paper is structured as follows. In the next Section, we provide information useful to understand what data we have used and how we have elaborated them. In the subsequent Section, we present our results and briefly discuss them along with some limitations. The Conclusions Section finally terminates this paper.

DATA AND METHODS OF ANALYSIS

We here provide details on data and methods utilized in our analysis.

Sources of data

To begin, it is worth revealing the source of data we have used for our study. They were drawn by the 51 footbal matches played by the 24 national teams participating in the 2020 UEFA European Football Championship, commonly referred to as Euro2020, the quadrennial international men's football championship of Europe, which ran from 11 June to 11 July, 2021 (it was postponed to 2021 due to the Covid pandemics).
The participating national teams were: Austria, Belgium, Croatia, Czech Republic, Denmark, England, Finland, France, Germany, Italy, Hungary, Netherlands, North Macedonia, Poland, Portugal, Russia, Scotland, Slovakia, Spain, Sweden, Switzerland, Turkey, Ukraine, Wales.
The tornament was organized based on two subsequent phases: a group stage phase and a final knockout phase (going through the following steps: a round of 16, quarter finals, semifinals and final).
The teams admitted to the Round of 16 were: Austria, Belgium, Croatia, Czech Republic, Denmark, England, France, Germany Italy, Netherlands, Portugal, Spain, Switzerland, Sweden, Ukraine and Wales.
The teams admitted to the Quarter finals were: Belgium, Czech Republic, England, Denmark, Italy, Spain, Switzerland, Ukraine.
Those admitted to the Semifinals were: England, Denmark, Italy, Spain. To the final: England and Italy.
The tournament (won by the Italian national team) was the one with the most goals per game in a European Championship since the introduction of the group stage, and with only two goalless matches. More details on the tornament can be found at the following Web reference: (Euro 2020).
The football data utilized in this study come, instead, from an international football data repository from which all types of relevant data, including goals and xG, have been scraped.
The repository is Statsbomb, with its relative Web reference: (Statsbomb 2024).
In general, all xG models provide their own statistical framework to systematically evaluate the value of each shot and determine how much likely one is to result in a goal. For our study, we have decided to build upon the model provided by Statsbomb, as the correspondent xG model is well known to be one the best available, capable of using more contextual events and better quality data than many others.
Interested readers can find more readings on StatsBomb's xG model from the StatsBomb website, at the following Web reference: (Statsbomb-xG 2024).

Analysis

The initial data were subsequently elaborated by the authors of the present study to obtain what is described in the following. For more details, see also (Berveglieri 2024).
First, since the tournament was comprised of two different phases (group stage and knockout), a unique ranking was drawn up, with all the teams and relative scores, by considering the results of all the matches played by each team during the tournament, based on the conversion rules listed in Table 1. With this conversion, we got the ranking represented as an histogram in Figure 1.
Then, we arranged an alternative ranking with the results we would obtain in all the 51 matches if we replace real goals with xG. This alternative standing is shown in Figure 2.
The research hypothesis to validate becomes now if there is a statistically significant difference between the two scores distributions displayed in Figures 1 and 2.
Further, to make our investigations more precise, we have also divided all the teams of the tournament into three different groups: A, B and C, as listed in Tables 2, 3 and 4, below.
Group A is that of the teams admitted to the quarter finals.
Group B is that of the teams eliminated after the round of 16.
Finally, Group C is that of the teams which did not qualify after the group stage phase.
For each different team of groups A, B and C, we have computed (and reported in Table 5): i) the scores obtained considering respectively real goals (second column) and xG (third column), and the corresponding ii) real goals (fourth column) and xG (fifth column). The relative average values on a per group basis are displayed in grey in Table 5.

Table 1: Rules for awarding scores: group stage + knockout

Group stage	Score	Knockout	Score
win	3	win	3
loss	0	loss	0
draw	1	win (in extratime)	3
		loss (in extratime)	0
		win (on penalties)	2
		loss (on penalties)	1

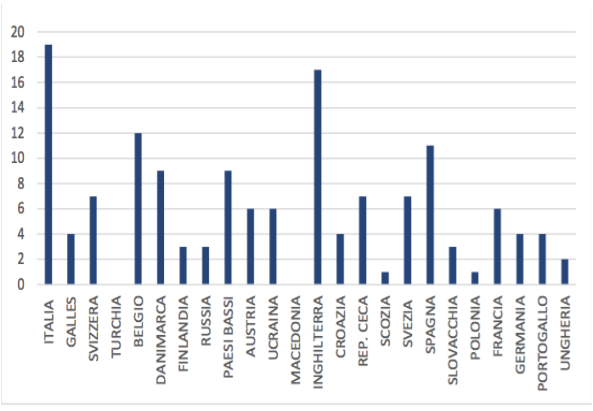

Figure 1: All-teams ranking using real goals (Names of teams in Italian)

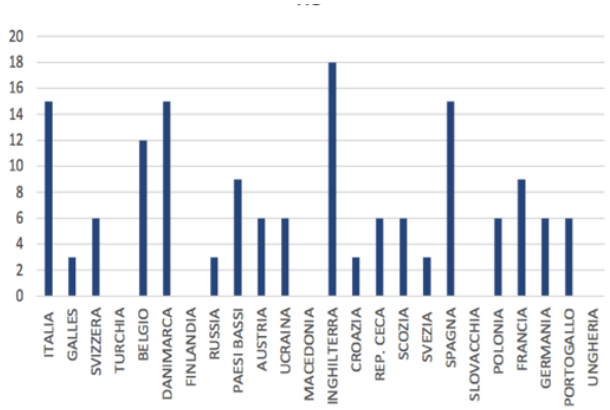

Figure 2: All-teams ranking using xG (Names of teams in Italian)

Table 2: Group A: teams admitted to the quarter finals

Team	
1	Belgium
2	Czech Republic
3	England
4	Denmark
5	Italy
6	Spain
7	Switzerland
8	Ukraine

Table 3: Group B: teams eliminated after the round of 16

Team	
9	Austria
10	Croatia
11	France
12	Germany
13	Netherlands
14	Portugal
15	Sweden
16	Wales

Table 4: Group C: teams which did not qualify after the group stage phase

Team	
17	Finland
18	Hungary
19	North Macedonia
20	Poland
21	Russia
22	Scotland
23	Slovakia
24	Turkey

At this point, our generic research hypothesis about the difference between goals and xG can be translated into two sets of null hypotheses, as described in the following.
The former set of four null hypotheses is about the difference between the two score distributions (i.e., rankings), based on the two following cases: i) when all the teams are considered together, and ii) when the teams are split over the three different groups A, B and C mentioned above. These four null hypotheses are displayed in Table 6 below.
The latter set of four null hypotheses, instead, is relative to difference between the distributions of goals and xG, again considering respectively: all the teams together and each group separately. They are shown in Table 7 below.

Table 5: Scores with real Goals (second column) and xG (third column); number of Goals (fourth column) and xG (fifth column); average values in grey

Team/Group	Score w/G	Score w/xG	G	xG
Belgium/A	12	12	7	3.58
Czechia/A	7	6	3	3.12
England/A	17	18	2	2.20
Denmark/A	9	15	5	5.90
Italy/A	19	15	7	4.96
Spain/A	11	15	6	7.76
Switzer./A	7	6	4	3.88
Ukraine/A	6	6	4	4.24
Average/A	**11**	**11.62**	**4.75**	**4.45**
Austria/B	6	6	4	3.34
Croatia/B	4	3	4	2.11
France/B	6	9	4	3.64
Germany/B	4	6	6	5.05
Netherl./B	9	9	8	7.99
Portugal/B	4	6	7	6.66
Sweden/B	7	3	4	5.05
Wales/B	4	3	3	3.69
Average/B	**5.5**	**5.62**	**5**	**4.69**
Finland/C	3	0	1	1.49
Hungary/C	2	0	3	1.41
N. Maced./C	0	0	2	3.57
Poland/C	1	6	4	3.60
Russia/C	3	3	2	2.60
Scotland/C	1	6	1	3.58
Slovakia/C	3	0	2	1,00
Turkey/C	0	0	1	2.70
Average/C	**1.625**	**1.87**	**2**	**2.49**

Table 6: Null hypotheses: no statistically significant difference between score distributions (w/ goals and xG); All together, Group A, Group B and Group C

$H_S(0)_{Tog}$	Score distributions (w/ goals and xG) are not statistically significantly different, teams all-together
$H_S(0)_A$	Score distributions (w/ goals and xG) are not statistically significantly different, Group A
$H_S(0)_B$	Score distributions (w/ goals and xG) are not statistically significantly different, Group B
$H_S(0)_C$	The two score distributions (w/ goals and xG) are not statistically significantly different, Group C

To cite as: Roccetti, M., Berveglieri, F., Cappiello, G., In Proc. 25th International Conference on Intelligent Games and Simulation, GAME-ON 2024, 2024, Istanbul, pp. 20–24

Table 7: Null hypotheses: no statistically significant difference between the distributions of goals and xG; All together, Group A, Group B and Group C

$H_G(0)_{Tog}$	Distributions of goals and xG are not statistically significantly different, teams all-together
$H_G(0)_A$	Distributions of goals and xG are not statistically significantly different, Group A
$H_G(0)_B$	Distributions of goals and xG are not statistically significantly different, Group B
$H_G(0)_C$	Distributions of goals and xG are not statistically significantly different, Group C

Each null hypothesis of the eight above has been finally tested using the non-parametric Kolmogorov-Smirnov testing procedure (KS tests, for short), and a statistical significance value equal to 5% (Pearson 1972). Obviously, if our procedure finds significant statistical differences, the null hypotheses above are to be rejected in favor of alternative hypotheses which would say that one distribution (either goals or xG) is superior to the other.
Instead, if there is no significant statistical difference, the null hypothesess cannot be rejected. This would not automatically mean that the null hypotheses are true. Nonetheless, in the case where eight different null hypotheses, all focusing on the differences between goals and xG, are not rejected, we would have enough strength to conclude that no strong evidence has been found in favor of the general alternative hypothesis of a difference between goals and xG.
In the end, to conduct our analysis, we decided to use a two-sample Kolmogorov Smirnov statistic, as it lies at the basis of one of the most powerful non-parametric procedure for comparing the empirical distribution function of two samples, which was exactly our case. In the next Section, we have reported the results we have obtained with our two-tailed KS tests, conducted to validate the null hypotheses.

RESULTS AND LIMITATIONS

Tables 8 and 9 below show the results we have obtained with the Kolmogorov Smirnov hypothesis testing procedure mentioned above.
In all the eight examined cases the null hypothesis was not rejected, as the corresponding *p*-values (rightmost columns in Tables 8 and 9) were always much larger than the corresponding significance threshold of 0.05. In conclusion, these results, as a whole, are in favor of the general hypothesis according to which a strong difference between goals and xG cannot be found.
Nonetheless, there are some imitations to these results which go as follows.
First, from a technical viewpoint it should be noticed that the statistical significance, tested with the Kolmogorov Smirnov method, can be strongly affected by the number of observations (Vrbik 2018), and so even small discrepancies can be considered significant for very large sample sizes, while very large discrepancies are required to reject the null hypothesis for small sample sizes, as it may be the case here (with just 51 matches and 142 goals from a single tournament).
Not only that, but the tournament itself has gone through two steps (i.e., group stage and knockout) where not all the teams have had the possibility to play the same number of matches (each of similar duration in time, consider: extratimes and penalties). This has required to use the conversion rules of Table 1 to draw up a unique ranking comprising all the teams participating in Euro 2020.
In some sense, it should be noted that this has been a kind of special distorsion of the big picture. Along this line of sense, it goes without saying that the extension to more football tournaments (based on a single group stage phase), and coming from different geographies, could result into more robust statistical results.

Table 8: Score distributions (goals *vs* xG): KS results - significance level=0.05

Hypothesis testing	Average score per team w/ goals	Average score per team w/ xG	KS test: *p*-value
All Teams	6.042	6.375	0.9941
Group A	11	11.625	0.9639
Group B	5.5	5.625	0.6272
Group C	1.625	1.875	0.6272

Table 9: Goal distributions (goals *vs* xG): KS results - significance level=0.05

Hypothesis testing	Average goals per team	Average xG per team	KS test: *p*-value
All Teams	3.917	3.882	0.4490
Group A	4.75	4.458	0.9801
Group B	5	4.691	0.6601
Group C	2	2.496	0.6601

Limitations are evident also from a pure football analysis viewpoint.
For example, converting the two different phases of the tournament (group stage and knockout) into a unique ranking has determined some inversions in the ranking. Among others: Belgium, Denmark and Spain have almost reached the top of the standing based on xG, very close to the real finalists: England and Italy.
Nonetheless, this should be considered as a secondary effect of the unique ranking, obtained merging the two different phases of the tournament, rather than a direct consequence of the use of xG.

CONCLUSIONS AND FUTURE RESEARCH

xG is a football metric that has roots in the computer gaming culture and it is unanimously considered by football analysts as a powerful tool to measure the likelihood of finalizing a given shot (Goldman 2024; Kloke 2024; Pollard 2004).
There is, instead, more debate about the idea that it can be a good predictor of the performances of the attack/defence of a teams, as a whole.
We are in favor of the hypothesis that, in a longer period, this performance indicator may also have a predictive power with

respect to the overall performance of a football team. To provide evidence in support to this hypothesis, we have developed a statistical study, using the non-parametrical Kolmogorov Smirnov hypothesis testing method with data from the 2020 UEFA European Football Championship (source: Statsbomb), that has shown that there is no statistically significant difference if the final scores of that tournament were awarded to the participating teams with either real goals or based on xGs, thus demystifying some wrong assumptions about this football indicator.What is also interesting in our study is that we have obtained this kind of result by viewing at things from the perspective of big data, as only many data are able to put in evidence the peculiarity of these phenomena, as already noticed in the technical literature (Casini et al. 2020; Roccetti et al. 2024; Roccetti et al. 2019; Marchetti et al. 2018; Marfia et al. 2011; Marfia et al. 2010; Roccetti et al. 2007; Salomoni et al. 2007; Palazzi et al. 2006; Corradini et al. 1997). Finally, future researches should extend this type of analysis to more football tournaments and to different geographies.

ACKNOWLEDGEMENTS

M.R. and F.B conducted the research. M.R. wrote the manuscript. G.C. contributed to the discussion. All authors reviewed the manuscript. M.R. thanks the University of Bologna for supporting his research under a RFO 2023 grant.

REFERENCES

Barnett, V. and S. Hilditch. 1993. "The Effect of an Artificial Pitch Surface on Home Team Performance in Football (Soccer)". *Journal of the Royal Statistical Society. Series A (Statistics in Society).* 156 (1): 39–50. doi:10.2307/2982859.

Berveglieri, F. 2024. "Il concetto di Expected Goals nella Football Data Analysis: uno studio comparativo relativo a Uefa Euro 2020 (in Italian)". *Laurea Thesis in Management and Information Science,* University of Bologna.

Casini, L., Delnevo, G.,. Roccetti, M., Zagni, N. and G. Cappiello. 2020. "Deep water: Predicting water meter failures through a human-machine intelligence collaboration". *Advances in Intelligent Systems and Computing.* 1018:688-694. doi: 10.1007/978-3-030-25629-6_107.

Ensum, J., Pollard, R. and S. Taylor. 2004. "Applications of logistic regression to shots at goal in association football: calculation of shot probabilities, quantification of factors and player/team". *Journal of Sports Sciences.* 22 (6): 504.

Corradini, F., Gorrieri, R. and M. Roccetti. 1997. "Performance preorder and competitive equivalence". *Acta Informatica.* 34 (11): 805–831. doi: 10.1007/s002360050107.

Goldman, S. 2024. "Comparing public expected goal models: How they work and what we should take away from them". *The Athletic*. 30 April 2024.

Kloke, J. 2024. ""Understanding expected goals and how they impact Toronto FC". *The Athletic*. 30 April 2024.

Mead, J., O'Hare, A. and P. McMenemy. 2023. "Expected goals in football: Improving model performance and demonstrating value". *Plos One.*18 (4). doi: 10.1371/journal.pone.0282295.

Marchetti, N., Angelini, I., Artioli, G., Benati, G., Bitelli, G., Curci, A., Marfia, G. and M. Roccetti. 2018. "NEARCHOS. Networked Archaeological Open Science: Advances in Archaeology Through Field Analytics and Scientific Community Sharing". *Journal of Archaeological Research*. 26 (4):447-469. doi: 10.1007/s10814-017-9112-4.

Marfia, G., Roccetti, M., Amoroso, A., Gerla, M., Pau, G. and J.-H. Lim. 2011. "Cognitive cars: Constructing a cognitive playground for VANET research testbeds". In *ACM International Conference Proceeding Series - 4th International Conference on Cognitive Radio and Advanced Spectrum Management.* doi: 10.1145/2093256.2093285.

Marfia, G. and M. Roccetti. 2010. "TCP at last: Reconsidering TCP's role for wireless entertainment centers at home". *IEEE Transactions on Consumer Electronics.* 56 (4):233-2340. doi: 10.1109/TCE.2010.5681095.

Palazzi, C.E., Ferretti, S., Roccetti, M., Pau, G. and M. Gerla. 2006. "What's in that magic box? The home entertainment center's special protocol potion, revealed". *IEEE Transactions on Consumer Electronics*. 52 (4):1280-1288, doi: 10.1109/TCE.2006.273146.

Parkin, S. 2016. ""Fifa: the video game that changed football". *The Guardian*. 21 December 2016.

Pearson, E.S. and H.O. Hartley. (Eds.) 1972. "*Biometrika Tables for Statisticians*". Vol. 2. Cambridge University Press. 117–123, ISBN 978-0-521-06937-3.

Pollard, R., Ensum, J. and S. Taylor. 2004. "Estimating the probability of a shot resulting in a goal: The effects of distance, angle and space". *International Journal of Soccer and Science.* 2 (1): 50–55.

Roccetti, M., Tenace, M. and G. Cappiello. 2024. "Prescient Perspectives on Football Tactics: A Case with Liverpool FC, Corners and AI". In Proceedings of the 2024 International Symposium on Foundations and Applications of Big Data Analytics (FAB/ASONAM'24 - Sept 2024, Cosenza Italy).

Roccetti, M., Delnevo, G., Casini, L. and S. Mirri. 2021. "An alternative approach to dimension reduction for pareto distributed data: a case study". *Journal of Big Data*. 8 (1). doi: 10.1186/s40537-021-00428-8

Roccetti, M., Delnevo, G., Casini, L. and G. Cappiello. 2019. "Is bigger always better? A controversial journey to the center of machine learning design, with uses and misuses of big data for predicting water meter failures". *Journal of Big Data*. 6 (1). doi: 10.1186/s40537-019-0235-y.

Roccetti, M., Ferretti, S. and C.E. Palazzi. 2008. "The Brave New World of Multiplayer Online Games: Synchronization Issues with Smart Solutions". In Proceedings of *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC).* Orlando, FL, USA, 2008, 587-592, doi: 10.1109/ISORC.2008.17.

Roccetti, M., Gerla, M., Palazzi, C.E., Ferretti, S. and G. Pau. 2007. "First responders' crystal ball: How to scry the emergency from a remote vehicle". In Proceedings of *2007 IEEE International Performance, Computing, and Communications Conference (IPCC).* 556 – 56. doi: 10.1109/PCCC.2007.358940.

Salomoni, P., Mirri, S., Ferretti, S. and M. Roccetti. 2007. "Profiling learners with special needs for custom e-learning experiences, a closed case?". In Proceedings of *ACM International Conference Proceeding Series.* Volume 225, 84 – 92. *International Cross-Disciplinary Conference on Web Accessibility.* doi: 10.1145/1243441.1243462.

Vrbik, V. 2018. "Small-Sample Corrections to Kolmogorov–Smirnov Test Statistic)". *Pioneer Journal of Theoretical and Applied Statistics.* 15 (1-2): 15–23.

WEB REFERENCES

Euro 2020. Online from: https://en.wikipedia.org/wiki/UEFA_Euro_2020

Statsbomb 2024. Online from: https://statsbomb.com/what-we-do/soccer-data/360-2/

Statsbomb's xG 2024. Online from: https://statsbomb.com/soccer-metrics/expected-goals-xg-explained/

Tweedale 2024. Online from: https://www.coachesvoice.com/cv/expected-goals-xg-explained