



TF-IMAGES/GETTY

Belgium (in red) beat Brazil in a quarter-final match during the 2018 World Cup.

HOW BIG DATA IS CHANGING FOOTBALL

As the FIFA World Cup kicks off, researchers are showing their skills to help soccer coaches develop players and tactics. **By David Adam**

The scowl on Cristiano Ronaldo's face made international headlines last month when the Portuguese superstar was pulled from a match between Manchester United and Newcastle with 18 minutes left to play. But he's not alone in his sentiment. Few footballers agree with

a manager's decision to substitute them in favour of a fresh replacement.

During the upcoming football World Cup tournament in Qatar, players will have a more evidence-based way to argue for time on the pitch. Within minutes of the final whistle, tournament organizers will send each player a detailed breakdown of their performance.

Strikers will be able to show how often they made a run and were ignored. Defenders will have data on how much they hassled and harried the opposing team when it had possession.

It's the latest incursion of numbers into the beautiful game. Data analysis now helps to steer everything from player transfers and the intensity of training, to targeting opponents

and recommending the best direction to kick the ball at any point on the pitch.

Meanwhile, footballers face the kind of data scrutiny more often associated with an astronaut. Wearable vests and straps can now sense motion, track position with GPS and count the number of shots taken with each foot. Cameras at multiple angles capture everything from headers won to how long players keep the ball. And to make sense of this information, most elite football teams now employ data analysts, including mathematicians, data scientists and physicists plucked from top companies and labs such as computing giant Microsoft and CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

In return, insights from analysts are altering how the game is played: strikers shoot less frequently from a distance, wingers pass to a teammate rather than cross the ball and coaches obsess about winning possession high up the pitch – tactical shifts all backed up with hard evidence to support a coach's intuition.

"Big data has ushered in a new era of football," says Daniel Memmert, a sports scientist at the German Sport University Cologne. "It has changed the philosophy and behaviour of teams, how they analyse opponents and the way they develop talent and scout players."

Covering all bases

One of the best known cases of how data is changing sports comes from a different game. In his 2003 book *Moneyball*, Michael Lewis detailed how Oakland Athletics' manager Billy Beane relied on player statistics to deliver a winning baseball team on a shoestring budget in 2002. Beane recruited players on the basis of detailed data about their performance, including previously undervalued measurements, such as how often a batter made it to base.

Beane had an advantage over those trying to repeat the trick in football. "Football is much more complex than baseball," Memmert says. Baseball is a natural stop-start game in which only one team at a time is trying to score, and baseball statistics had been collected routinely and studied on a large scale for decades. By contrast, football is a fluid and low-scoring 'invasion' game (one in which territory is regularly gained and surrendered), and it's much harder to record who does what and how it affects the outcome. For decades, football statisticians tended to focus on goals scored and conceded, and to find a way to model them to make predictions.

Variants of this method are still used today to predict the outcomes of matches. A mathematical model that assumes goals scored and conceded are distributed around a mean value, developed by epidemiologists at the University of Oxford, UK, correctly predicted that Italy would beat England in the Euro 2020 international tournament. It also correctly called six of the eight quarter-finalists¹.

Such success is not unusual. Statistical match predictions are more accurate than many people realize, says Matthew Penn, a PhD student at Oxford, who developed the Euro 2020 model.

"You want to give each team an offensive and a defensive strength, and you work that out from the total number of goals that each team has scored and the relative difficulty of their opponents," he says. "You end up with this big set of equations to solve for these two sets of strengths, and then it becomes really quite easy to predict each match." For the upcoming Qatar World Cup, Penn's model suggests that Belgium (priced at a generous 14/1 as *Nature* went to press) has the highest chances of raising the famous trophy, followed by Brazil (see 'Who will win the World Cup?').

Pitch perfect

What's more interesting to coaches is information about events on the pitch and how players influenced them.

Football analysts have long been recording this type of information. Most successfully, a former Royal Air Force accountant called Charles Reep spent much of the 1950s watching matches in England and made basic observations of factors such as pitch positions and passing sequences. Reep even used his data to analyse team performances and suggest strategy and tactics. At Wolverhampton Wanderers Football Club (FC), he helped to introduce a direct and incisive playing style that frowned on sideways passes and won three league championships in five years.



YOU CAN START TO ASK QUESTIONS ABOUT TACTICS OR COUNTERFACTUALS."

Modern technology makes such data so much easier to obtain and analyse that most top clubs, and many national teams, started to employ data analysts more than a decade ago. And its use extends deep into the football pyramid. As well as studying for his PhD, Penn works as a part-time data analyst for Oxford City, a semi-professional club that plays in the National League South, the sixth tier of the English system.

Many analysts attribute some of the recent success of the London club Brentford FC to an in-house algorithm that rates players across different leagues and helps the team to recruit undervalued stars. Liverpool FC's data team,

which includes physicists formerly at CERN and the University of Cambridge, UK, has built a model that can assess whether a player's actions on the pitch make a goal more likely. And in a partnership with the Spanish giants FC Barcelona, sports scientists at the University of Lisbon, Portugal, last year published an analysis of how long opportunities last for different types of pass in a match².

"I think the most useful thing we're doing [at Oxford City] is the pre-match reports," Penn says. "We look at the attributes of the players for the other team and then produce some graphs to show how they're playing and how they're moving through possession. And then I'll suggest some tactical tips or changes." Ahead of a recent match against a previously unbeaten side, Penn's analysis identified that the left back had poor heading stats. "So the suggestion was our big striker stands on our right-hand side of the pitch," he says. Oxford won the game.

That's also the kind of insight readily available to the naked eye of an experienced scout. But, Penn says, "the data is going to be less biased than someone's opinion".

Clubs don't have to generate the raw data for these kinds of tactical analysis themselves. Instead, they can buy the information from commercial companies that code filmed matches to record the outcome of some 3,000 major in-game events, including dribbles, passes and tackles. At first, such data were recorded manually, but it's now usually done using a type of artificial intelligence (AI) called computer vision. Often, these data come with summary statistics, such as each player's pass-completion rate.

Working with Penn at Oxford City, Joanna Marks, a mathematics undergraduate at the University of Warwick, UK, developed a model earlier this year to use those raw data to assess the passing strength of all players in Oxford's league – the kind of detailed analysis not usually available in the raw data supplied by the companies.

"You need to take into account what kind of pass they attempt. You can't just take the completion ratio because some passes are much more difficult," Marks says. "The model helps to prepare the team, because if you know an opponent is passing very well from some area of the pitch, then you know what to watch out for."

Ravi Ramineni worked as a data analyst at Microsoft before transferring in 2012 to a similar job at his local US Major League Soccer (MLS) club, the Seattle Sounders in Washington. One of his first tasks was to use GPS data on how far the players ran, to optimize their training and preparation sessions. "Collecting this data during training, you can tell maybe today the training was too much or too little. You do that to try to prevent injuries."

Did it work? "We had some really good seasons when we applied the methods. But I don't know. The hardest thing to quantify here is if

Feature

an injury doesn't happen."

The lack of certainty raises an issue with all claims about the role of data in sporting success: there's no control experiment to check on efficacy. Still, Ramineni says, the coaches at Seattle were open to his analyses, both in training and later when judging the strengths of players.

"I was given access to the coaches, and I could even go directly talk to players," he says. "At other clubs it's not the same. Sometimes the coach doesn't even interface with the data guy."

Off-camera action

Analysts are now increasingly paying attention to what happens when players don't have the ball.

"One thing you would hear in football analytics all the time is that we need to know what the player does off the ball," says Ramineni.

That's more difficult and expensive, because it requires dedicated cameras that don't simply track the main action, but also keep an eye on players who are not directly involved and tag their locations some 25 times per second. Companies that supplied this kind of data tended to sign exclusive deals with national leagues, Ramineni says, which made access difficult for outsiders.

"If I was scouting an international player from South America or Europe for the MLS, I wouldn't know their off-the-ball metrics," he says.

In recent years, a more powerful technique has emerged that harnesses AI to predict the movements of players in filmed matches, even when they are not captured directly by cameras. This means that data companies can use broadcast footage of games (available with no restrictions) to produce comprehensive on- and off-the-ball analytics for players anywhere in the world.

One such predictive model has been developed by a partnership between researchers at DeepMind, the Google-owned AI company based in London, and the data team at Liverpool FC³.

"With that sort of application, you can start to ask questions about tactics or counterfactuals," says Ian Graham, Liverpool FC's director of research, who quit a postdoc in polymer physics at the University of Cambridge to start working in football statistics.

"For a specific incident in the match, the model can produce thousands of different simulations about what could have happened instead. So, you can start to say something about how well an attacking move went in that period of play."

Clubs' data teams tend not to share information on the specifics of what they are doing or how well it works, so publishing this work was an uncomfortable step for Liverpool. But it was a condition of working with Deep Mind.

"Liverpool has got one of the biggest and most-developed analytics departments in

WHO WILL WIN THE WORLD CUP?

A statistical 'double Poisson' model that considers the attacking and defending strengths of each men's team ranks Belgium as having the highest odds of winning the World Cup, whereas Brazil tops the FIFA rankings.

Model's ranking	Chance of winning (%)	FIFA ranking position
1. Belgium	13.88	1. Brazil
2. Brazil	13.51	2. Belgium
3. France	12.11	3. Argentina
4. Argentina	11.52	4. France
5. Netherlands	9.65	5. England
6. Germany	7.24	6. Italy*
7. Spain	6.37	7. Spain
8. Switzerland	5.29	8. Netherlands
9. Portugal	3.78	9. Portugal
10. Uruguay	3.36	10. Denmark
11. Denmark	3.17	11. Germany
12. England	2.56	12. Croatia
13. Poland*	2.33	13. Mexico
14. Croatia	1.46	14. Uruguay
15. Mexico	0.67	15. Switzerland

*Ranked 26th by FIFA; *Did not qualify for World Cup.

football, and we don't have anything like the resources that we need to build these models ourselves," Graham says. This reassures him that no other club can either.

Like other data analysts, Graham is reluctant to take direct credit for success on the pitch. "Football is a high-variance game, so teams quite often lose when they shouldn't lose and they win when they shouldn't win," he says. "In many ways, our job's easier when the team's doing badly, because our analysis quite often shows that we played fine. And if we keep play-



DATA-ANALYTICS PEOPLE HAVE STARTED SAYING: 'WHY ARE YOU SHOOTING FROM THERE?'"

ing like that, we will win the expected number of games this season."

Karl Tuyls, a computer scientist at DeepMind, says the off-camera modelling work is the first step towards creating a virtual, AI-driven assistant coach that uses real-time data to guide decision-making in football and other sports. "You can imagine the AI looking at the first-half performance and suggesting a change in formation

that might do better," he says.

The approach could also be useful away from the pitch, for tasks such as modelling the trajectories of self-driving cars and pedestrians on a busy city street, Tuyls adds.

What's next? Like all good scientists, the experts involved in football data stress that more research is still needed. Sarah Rudd, a former Microsoft data scientist who left Arsenal FC last year after nearly a decade working on analytics for the London club, covets the masses of telemetry data produced by a racing car that helps support teams to tweak and improve its performance.

"We always look at Formula One and say it would be great to have that level of data," she says. "There's still a lot of stuff in football that isn't being measured, or it's being measured but we haven't figured out how to derive insights from it."

The next advance could be data that show players' orientation, and even how they shift their weight. "The tracking data is still maybe not at the granularity that people want," Rudd says. "You're not yet picking up that small stutter step or shift in weight that a player does to throw the defender off balance, or to give the goalkeeper a little bit of a pause."

Even Liverpool's AI-driven analytics can be bamboozled by incomplete knowledge of a player's position. "The model might say this player did a bad thing because he should have started running at this point and he didn't," Graham says. "But that might be because he was just tripped over and was lying on the floor."

As modern football drowns in data, how have numbers changed the game?

"I think recruitment is probably where you get the biggest bang for buck," Ramineni says. Another area, however, is in strategies for set pieces, when a team gets a free kick after play is paused.

One clear lesson that has emerged from data analysis is that players shouldn't shoot when they're far from the goal. "If you look at any league in the world, the distance from where players have taken shots was much higher ten years ago," Ramineni says. "That all happened because data-analytics people have started saying: 'Why are you shooting from there? It's only a 2% chance!'"

Many teams also now discourage players from attempting long crosses towards the penalty area, he adds, after statistics showed most were pointless.

And as the volume of data generated continues to grow, so will the job opportunities, says Ramineni. "I think the data footprints are all over the sport now and there's no going back."

David Adam is a freelance journalist in London.

1. Penn, M. J. & Donnelly, C. A. *PLoS ONE* **17**, e0268511 (2022).
2. Gómez-Jordana, L. I., Amaro e Silva, R., Milho, J., Ric, A. & Passos, P. *Sci. Rep.* **11**, 9792 (2021).
3. Omidshafiei, S. et al. *Sci. Rep.* **12**, 8638 (2022).

SOURCE: REF. 1