



# Who can receive the pass? A computational model for quantifying availability in soccer

Uwe Dick<sup>1</sup> · Daniel Link<sup>2</sup> · Ulf Brefeld<sup>3</sup> 

Received: 19 March 2021 / Accepted: 12 February 2022  
© The Author(s) 2022

## Abstract

The paper presents a computational approach to Availability of soccer players. Availability is defined as the probability that a pass reaches the target player without being intercepted by opponents. Clearly, a computational model for this probability grounds on models for ball dynamics, player movements, and technical skills of the pass giver. Our approach aggregates these quantities for all possible passes to the target player to compute a single Availability value. Empirically, our approach outperforms state-of-the-art competitors using data from 58 professional soccer matches. Moreover, our experiments indicate that the model can even outperform soccer coaches in assessing the availability of soccer players from static images.

**Keywords** Sports analytics · Football · Elite sports · Tracking data · Graph networks · MDNs · Ball dynamics · Movement models · Passing behavior

## 1 Introduction

Passes are the most frequent event in soccer (Pappalardo et al. 2019). By changing ball possession from player to player, teams play the ball closer to the opponents' goals

---

Responsible editor: Ulf Brefeld and Albrecht Zimmermann.

---

✉ Ulf Brefeld  
brefeld@leuphana.de  
Uwe Dick  
uwe.dick@deltatre.com  
Daniel Link  
daniel.link@tum.de

<sup>1</sup> Sportec Solutions AG, Münchener Str. 101B, 85737 Ismaning, Germany

<sup>2</sup> Performance Analysis and Sports Informatics, TU Munich, Georg-Brauchle-Ring 60-62, 80992 Munich, Germany

<sup>3</sup> Machine Learning Group, Leuphana University of Lüneburg, Universitätsallee 1, 21335 Lüneburg, Germany

and try to create chances for scoring. The quality of passing affects the team's success and represents an important category for evaluating collective and individual match performances (Hughes and Franks 2005). It is easy to accept that simple indicators, such as the number of completed passes, are not very meaningful when it comes to assessing passing skills (Mackenzie and Cushion 2012). A simple pass backwards to a non-marked teammate has to be assessed differently than a long through ball between the defending lines, which may lead to a scoring opportunity. Therefore, research suggests using more advanced metrics e.g. the risk of a pass in relation to its potential effort (Goes et al. 2019; Power et al. 2017; Bransen et al. 2019b).

When coaches analyze match performance, they use – explicitly or implicitly – the concept of *Availability*. For this paper, we want to define Availability as the probability of which a selected player can receive the ball from a teammate. Availability is related to the risk of a pass and does not refer to a specific pass only, but asks for the aggregated success probability of all (hypothetical) passes, which could be received in a worthwhile zone. This concept is highly useful for understanding the performance of the passing player as well as for the receiver. The passing player has to choose from several target players (Steiner 2018), and has to adjust the kicking movement precisely with regard to the ball's trajectory and speed. The Availability of the teammates is an important factor for assessing both, the quality of the decision for a target player and the motor-technical skills for passing. On the other hand, Availability is also important for evaluating the tactical behavior of potential pass receivers. In elite soccer, players have to move in an unpredictable and dynamic way to create free spaces in which they can receive the ball (Bangsbo and Peitersen 2004). If players fail to create this space, the tactical options in the attacking game are limited and the defending team can be more successful (Fernández and Bornn 2018). Against this background, an objective quantification of Availability would be very useful for match analysis in soccer.

This paper presents a probabilistic machine-learning approach for computing Availability based on spatiotemporal tracking data. From a computational perspective, quantifying Availability is a challenging problem, since there are many factors to consider including that (i) players are constantly in motion and may receive the ball at many different positions on the pitch, (ii) there are many possible trajectories of the ball to the same position resulting in different transition times of the ball, (iii) opponents may intercept the ball, (iv) the technical skill levels of pass giver and the (v) time for controlling the ball influence successful completion of the pass. Our approach addresses all these aspects: For every moment  $t$ , we compute probabilities at which player  $r$  can receive a pass at the position  $x$ , played with multiple speeds. After this, we aggregate those probabilities for many  $x$  to an overall probability of a successful pass, which we call Availability of  $r$  at the moment  $t$ . Figure 1 shows an example and the corresponding availability of the receiver, which serves as a running example for the rest of the paper.

The remainder is organized as follows. Section 2 reviews related work. We present our Availability model including sub-models for interception, player skills and reception in Sect. 3, followed by the sub-models for ball and player movement in Sect. 4. The experimental validation in Sect. 5 compares computed Availability scores with observed receptions and interception of real passes in 58 matches of German Bun-



**Fig. 1** Running example: Scene from a Bundesliga match where the blue player with the black circle denoting the ball passes to the blue player below the orange zone. Left: The receiving player will receive the ball at the location marked with the light black circle. The orange zone marks the area in which the player can receive the ball, where darker orange indicates higher success probability. The number denotes the overall availability of the player. Note that the dark black circle indicating the position of the ball does not align perfectly with the actual position of the ball. This is due to small errors in the provided positional data. Right: The moment of the first touch of the receiving player. Incidentally, the receiver is not able to control the ball and will lose the ball immediately. For better visibility, a larger version of this figure is shown in Appendix C

desliga. Additionally, we compare computed Availabilities to human observer ratings. Section 6 highlights different application areas and Sect. 7 concludes.

## 2 Related work

Data-driven analyses of sports and soccer in particular are manifold in the literature. Existing approaches cover different aspects of the game including tactical constructs, estimating outcome of matches or quantifying the probability of goal scoring, see Goes et al. (2021) for an overview. For the purpose of this paper we aim to focus on approaches dealing with passing and the movement of players in a narrow sense.

Estimating the likelihood of successful passes has been investigated in Spearman et al. (2017) and Peralta Alguacil et al. (2020), where the authors propose a physics-based approach to predict the time until a player can reach a certain position. The time component can be computed by solving the player's *equation of motion*. In Peralta Alguacil et al. (2020) for example, the model is based on Fujimura and Sugihara (2005) for solving the equation of motion and augmented with an additional logistic distribution model to define an overall reachability. The authors also employ a physics based ball dynamics model. Overall, the approach is similar to ours, though we show in our experiments that by using our fully learned player dynamics and ball dynamics, we are able to better predict dynamics and, ultimately, the Availability of players. Alternative approaches for solving the equation of motion are provided in Taki et al. (1996) and Brefeld et al. (2019). Note that Peralta Alguacil et al. (2020) also identify potential runs of attacking players that maximize a combination of pass probability, pitch impact and pitch control. For an attacking player, based on the chosen combination, an optimal position is computed and compared to the actual observed position.

Other work on the analysis of passes in soccer include Power et al. (2017), who compare the risk of a pass (probability of an intercepted pass) versus its reward, the likelihood that the attacking team will take a shot at goal within 10 sec after the pass. Goes et al. (2019) estimate the effectiveness of passes by measuring how much defensive players have to move and how much their defensive organization reduces following a pass, while Bransen et al. (2019a) use event data of passes and model the reward by measuring the impact of these pass on the goal scoring probability.

Other publications rate general game states according to different measures and can also be used to rate the effectiveness of passes. Spearman (2018) develops a model that combines scoring probabilities from a certain point on the pitch with a team's control at that point and the probability that the ball will reach the point. Fernández et al. (2019) measure pitch control of teams and players and pitch value, which was estimated to correlate with positions that defenders aim to occupy. Recently, many publications use machine learning approaches to predict which player will receive the next pass (Vercruyssen et al. 2016; Fournier-Viger et al. 2018; Hubáček et al. 2018; Dauxais and Gautrais 2019; Li and Zhang 2019; Fernández et al. 2021). Although these approaches aim to model tactical decisions of the passing player, the Availability approach presented in this paper asks for the success probability of such a pass.

Estimating future positions of soccer players is another aspect that has been widely investigated. A general problem when learning coordinated movements of several agents, like players in a team, is that trajectories come as unordered sets of individuals. When learning from several games incorporating different teams and players, a model has to work without a natural ordering of the agents. Le et al. (2017b, a) learn future positions of players by estimating the roles of players in a given episode and using the role assignments to predict future movements using the then ordered set of players. Other publications also use role assignments to predict future positions of players using a variational recurrent neural networks (Zhan et al. 2019, 2018; Felsen et al. 2018). Yeh et al. (2019), on the other hand, study graph representations to model interactions of all agents including the ball. They leverage graph neural networks (GNN) which are naturally suited to model coordinated behavior because of their invariance to permutations in the input and propose a graph variational recurrent neural network to predict future positions of soccer and basketball players. Hoshen (2017) and Kipf et al. (2018) deploy graph-related attention mechanisms to learn trajectories for soccer and basketball players, respectively. GNNs have been widely used to model structured or relational data, Battaglia et al. (2018) provide an overview. In cases where data is sequential in nature, graph recurrent neural nets (GRNN) have been widely used, starting e.g. with Sanchez-Gonzalez et al. (2018) who mix graph representations with recurrent layers, such as gaited recurrent units (GRUs, Cho et al. 2014).

Due to the complex nature of soccer players' movements, one can expect the distribution of future points of a player to be multi-modal, which a probabilistic model that predicts future points should reflect. Thus, (conditional) variational models (CVM) with Gaussian emission functions are frequently deployed to account for multi-modality in the data (Zhan et al. 2019, 2018; Yeh et al. 2019; Felsen et al. 2018). However, Graves (2013) show that combining RNNs with mixture density networks (MDNs, Bishop 1994), with a Gaussian mixture model (GMM) as output distribution yields accurate predictive results for spatiotemporal tasks. In fact, Rudolph et al.

(2020) provide empirical evidence that combining GMM emission with recurrent graph networks works on par with using more complex CVM models.

### 3 Modeling availability

#### 3.1 Preliminaries

Our approach uses spatiotemporal data including  $xy$  positions of the players and  $xyz$  position of the ball. Data is recorded at 25Hz by a semiautomatic optical tracking system (TRACAB<sup>®</sup>, ChyronHego), which consists of up to 24 cameras around the pitch. The system uses computer vision methods to detect objects in the video stream. Tracking loss and identity swaps are eliminated manually after the matches. We make also use of a manually logged ball status flag, which indicates whether the ball is in play or not. The data is provided by German Professional Soccer League (Bundesliga) for the purpose of this paper. Reliability and validity of the system for measuring soccer specific movements is verified in Linke et al. (2020).

Computational analyzes that build upon tracking data, that is, sequences of player positions, movement directions, require some kind of formal representation of that data. To not clutter notation unnecessarily, we informally define the *state* of player  $i$  at time  $T$  by  $\mathcal{S}_T^i$ . The state  $\mathcal{S}_T^i$  contains the player's *position* in  $xy$ -coordinates and *velocity*, as well as *team* and *ball possession indicators*. Superscript 0 is reserved to index the state of the ball  $\mathcal{S}_T^0$  with its position, velocity, and additionally its  $z$ -coordinate at time  $T$ . We sometimes aggregate states of all players and ball at time  $T$ , denoted by  $\mathcal{S}_T^{0:N}$ , where usually  $N = 22$ , as well as time windows of interest by  $\mathcal{S}_{T_1:T_2}^i$ . For simplicity, we define  $\mathcal{S}$  as the entire history of states of the game until the current point in time.

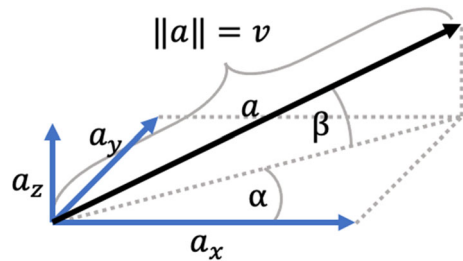
The goal of this paper is to establish a model for Availability. In other words, we aim to devise a model that computes the likelihood that a pass, irrespectively of whether it is a footed and headed pass, played from position  $b = [b_x, b_y, b_z]$  with initial direction  $a = [a_x, a_y, a_z]$  and speed  $\|a\|$ , can be reached by the intended receiver  $r$  without being intercepted by any opposing player. However, there are many passes with (slightly) different directions or velocities that may reach the target player. A solution thus needs to aggregate likelihoods for *all* possible passes to receiver  $r$  and aggregate them into an Availability value.

We assume that the passer chooses the best passing direction vector for passing to  $r$  but may not be able to execute the pass optimally. That is, the actual ball trajectory may differ from the intended one, as defined by the initial direction  $a' = a + \Delta a$ . Instead of working directly with vectors  $a$  and residual  $\Delta a$ , we will instead make use of horizontal and vertical angles  $\alpha$  and  $\beta$ , respectively, as well as ball speed  $v$ . The re-parameterization is given by

$$a(\alpha, \beta, v) = [\cos(\beta) \cos(\alpha), \cos(\beta) \sin(\alpha), \sin(\beta)] \cdot v$$

such that  $a' = a(\alpha + \Delta\alpha, \beta + \Delta\beta, v + \Delta v)$  with residual  $\Delta a = [\Delta\alpha, \Delta\beta, \Delta v]$ . The re-parameterization allows to formulate non-optimal executions of passes in

**Fig. 2** Visualization of initial ball direction in vector representation  $a = [a_x, a_y, a_z]$  and re-parameterized angle  $\theta = (\alpha, \beta, v)$



terms of differences in horizontal angle  $\Delta\alpha$ , vertical angle  $\Delta\beta$ , and the initial ball speed  $\Delta v$ . For ease of notation, we will often use tupled parameters  $\theta = (\alpha, \beta, v)$  and  $\Delta\theta = (\Delta\alpha, \Delta\beta, \Delta v)$  and  $a(\theta)$ ; Figure 2 provides a visualization. Using the re-parameterization, the Availability model consists of two parts: (i) computing likelihoods  $p_a^r(\theta)$  of successful passes to target player  $r$  along vertices  $a(\theta)$ , and (ii) aggregating these likelihoods into a single Availability value  $A^r(\psi)$  where  $\psi$  denotes the *skill* of the pass giver to determine the expected deviation of the actual pass from the intended one.

In the next section, we introduce models for ball dynamics that are used to predict ball trajectories from initial ball directions  $a(\theta)$ . Together with a predictor of whether a player can reach certain positions on the pitch in time, Section 3.3 derives a model to estimate the probability that a pass along an initial direction can be successful. Finally, Sect. 3.4 aggregates those values over a variety of initial directions into an availability value. The previously mentioned player reachability model is slightly more involving and introduced in Sect. 4.

### 3.2 Ball dynamics

Assume that the ball is played with initial movement vector  $a(\alpha, \beta, v)$  and moves along a straight line on the  $xy$ -plane. That is, we ignore curve balls for a moment. Naturally, the velocity of the ball decreases over time due to air and ground friction and so does its  $z$ -speed and position due to gravity and rotation. Physics implies that the deceleration curve of the ball depends strongly on the initial movement vector  $a(\alpha, \beta, v)$ , in particular on  $z$ -angle  $\beta$  and initial speed  $v$ . The reason behind this is two-fold. First, the ball is either flying (air friction, quadratic in speed) or rolling (mainly ground friction, approximately linear in speed). Secondly, depending on the intended distance and speed of the ball, the ball is played with more or less (backward) rotation, which changes its dynamics significantly. Note that rotation is not directly observable in tracking data.

The same holds for the acceleration in  $z$ -direction. While gravity force is constant, the observed acceleration varies significantly between passes, due to unobservable ball rotation. It becomes obvious that an Availability model cannot be computed in absence of ball movement. In fact, we capture ball dynamics with three distinct models that are also listed in Table 1.

**Table 1** Models for Ball Dynamics

$t(d; \beta, v)$	Time until ball reaches distance $d$
$u(d; \beta, v)$	Speed of ball at distance $d$
$p(z(t)   \beta, v)$	Distribution of height of ball at time $t$

- The first model is denoted by  $t(d; \beta, v)$  and estimates the time until the ball reaches a certain distance  $d$  after it has been kicked with initial  $z$ -angle  $\beta$  and velocity  $v$ . Function  $t$  is learned by a ridge regression with a polynomial kernel from historic data where  $\beta$  and  $v$  are estimated for every pass by taking the difference of the first two frames.
- The second model  $u(d; \beta, v)$  estimates ball velocity at a certain distance  $d$  with initial angle  $\beta$  and velocity  $v$ . We learn function  $u$  using a ridge regression with a polynomial kernel on historic data.
- The third model informs about the height  $z$  of the ball at a given distance. Ignoring air friction and ball rotation, the ball's  $z$ -coordinate dynamics would be determined by

$$z(t) = 1/2gt^2 + \sin(\beta)vt + z_0$$

where  $g \approx -9.81$  is the gravity force. However, the “observed gravitational force”, or in other words the observed acceleration  $\hat{g}$ , deviates strongly from frictionless acceleration  $g$ . In fact, as can be seen in Fig. 3, left, the observed acceleration  $\hat{g}$  depends on the initial  $z$ -angle  $\beta$ . We therefore learn a probabilistic model of “gravitation” from historic ball data by assuming that  $\hat{g}$  follows a Gaussian distribution when mean  $\mu(\beta, v)$  and variance  $\sigma(\beta, v)$  are linear functions. We thus have

$$p(z(t) | \beta, v) = \mathcal{N}\left(z(t) \mid \frac{1}{2}\mu(\beta, v)t^2 + \sin(\beta)vt + z_0, \frac{1}{2}\sigma(\beta, v)t^2\right). \quad (1)$$

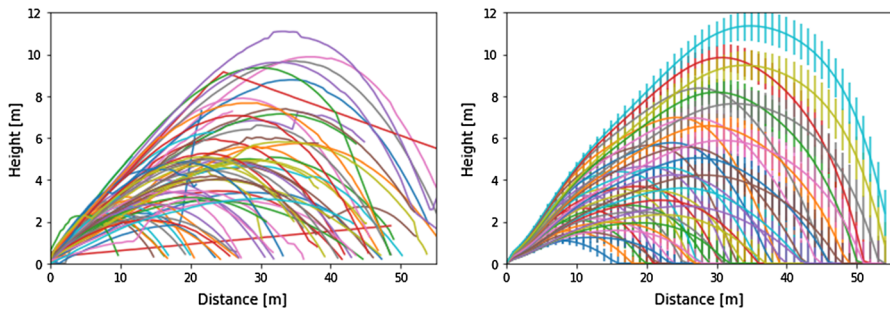
for  $z(t) \geq 0$  and  $p = 0$  otherwise and learn by minimizing the negative log-likelihood of the data.

Figure 3, right, shows sampled simulated ball trajectories for varying initial ball speeds  $v$  and  $z$ -angles  $\beta$ . Each simulated ball trajectory corresponds to a specific  $v, \beta$  pair and the height  $z$  at distance  $d$  is estimated by first computing time  $t_d = t(d; \beta, v)$  until the ball reaches distance  $d$  using the first model and then computing the distribution  $p(z(t_d) | \beta, v)$  using the third model and Eq. (1).

### 3.3 Quantifying the likelihood of passes

We begin the derivation of Availability by quantifying the likelihood that the intended receiver  $r$  reaches the ball when it is played from position  $b$  with parameters  $\theta = (\alpha, \beta, v)$  along the initial direction  $[\cos(\beta)\cos(\alpha), \cos(\beta)\sin(\alpha), \sin(\beta)]$  with initial speed  $v$ . We denote this likelihood by  $p_A^r(\theta, S)$ .





**Fig. 3** Left: Sample ball trajectories from the data (data contains artifacts). Right: Sampled ball trajectories using learned models  $t(d; \beta, v)$  and  $p(z(t) | \beta, v)$  (cf. Table 1). Error bars show standard deviations according to the distribution  $p(z(t))$

To proceed, we first derive the probability that the receiver can reach the ball at any point along the line given by  $\theta$ , we then derive the probability that any defender can intercept the ball before any of those points and in a last step aggregate those probabilities into a single value. In the course of this section, we will make use of movement models  $p_R^i(m, t; \mathcal{S})$  that quantify the probability that player  $i$  reaches position  $m$  in time  $t$ . We will introduce the model properly in Sect. 4 to not clutter this section unnecessarily.

### 3.3.1 Low passes

Let us assume that  $\beta = 0$  before we turn to the general case. We thus focus on low passes starting at the current position of the ball with arbitrary angle  $\alpha$  and velocity  $v$  while  $\beta = 0$ . Let  $m$  be an arbitrary  $xy$ -position on the pitch.

Still assuming only straight passes, the probability  $p_\alpha(m | \theta, \mathcal{S})$  that the ball passes through position  $m$  is a point measure that is 1 only if there exists  $c > 0 : m = b + c \cdot a(\alpha, 0, v)$ ,

$$p_\alpha(m | \theta, \mathcal{S}) = \begin{cases} 1 & : \exists c > 0 : m = b + c \cdot a(\alpha, 0, v) \\ 0 & : \text{otherwise.} \end{cases}$$

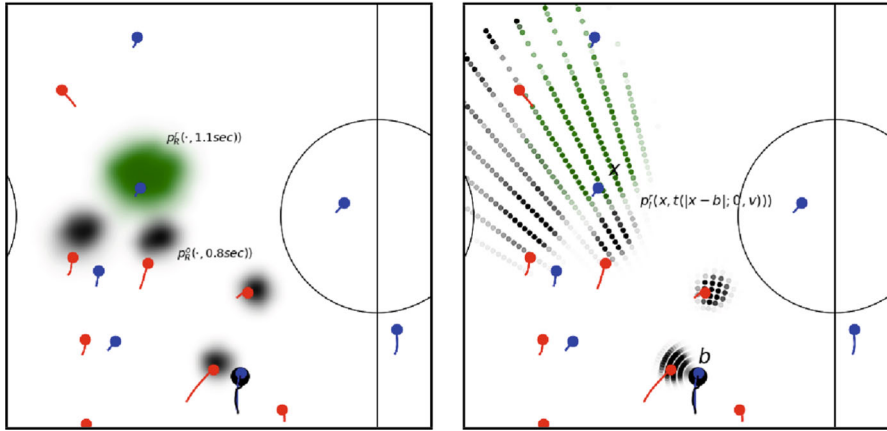
Furthermore, the probability  $p_I^i(m; \theta, \mathcal{S})$  that a low pass can be *intercepted* at position  $m$  by player  $i$ , equals the likelihood that player  $i$  can reach position  $m$  *before* the ball (which is currently at position  $b$ ), that is

$$p_I^i(m; \alpha, 0, v, \mathcal{S}) = p_R^i(m, t(\|m - b\|; 0, v); \mathcal{S}),$$

where  $t$  is the time-to-position function of the ball. Figure 4 shows a visualization. Analogously, the likelihood that player  $i$  can intercept the pass anywhere on the passing line *before* position  $m$  is given by

$$p_{CI}^i(m; \theta, \mathcal{S}) = \max_{c \in [0, 1]} p_I^i(b + c(m - b); \theta, \mathcal{S}).$$





**Fig. 4** Running example: The blue player at the bottom has possession of the ball (large black circle). Left: Reachability likelihoods  $p_R^r(\cdot, 1.1 \text{ sec}; S))$  of the ball receiver  $r$  (green shade) and of defenders (black shades). The depicted reachability probabilities correspond to estimated times of arrival of the ball with  $v = 25 \text{ m/s}$ . That is, the ball reaches the receiver in 1.1 sec according to model  $t(d; \beta, v)$  (cmp. Table 1). Right: Interception likelihoods  $p_I^r(x, t(\|x - b\|; 0, v); S))$  of receiver  $r$  (green) and defenders (black) with  $v = 25 \text{ m/s}$

In other words, the player will attain the position where the interception probability is highest. Figure 5, left, shows interception probabilities for the running example.

Putting everything together yields the probability that a low pass to player  $r$ , starting at position  $b$  along trajectory  $a(\alpha, \beta = 0, v) = a(\theta)$  and ending in position  $m$ , is successful is given by (i) the probability that position  $m$  lies on the trajectory of the ball, (ii) the probability that player  $r$  can intercept the ball exactly at position  $m$ , and (iii) the probability that no opponent  $o$  will intercept the ball before it reaches position  $m$ ,

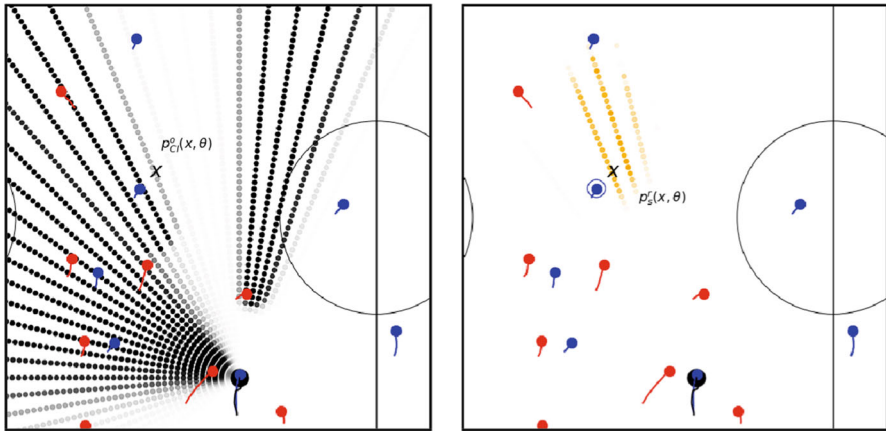
$$p_s^r(m; \alpha, \beta = 0, v, S) = \underbrace{p_\alpha(m | \theta, S)}_{\text{ball passes through } m} \cdot \underbrace{p_I^r(m; \theta, S)}_{\text{target player } r \text{ reaches ball in } m} \prod_o \underbrace{(1 - p_{CI}^o(m; \theta, S))}_{\text{no opponent intercepts}}. \quad (2)$$

Figure 5, right, shows examples of those probabilities. The likelihood that a low pass along direction  $a(\theta)$  is successful can be written as

$$p_A^r(\theta, S) = \max_{m \in \{b + c \cdot a(\theta); c > 0\}} p_s^r(m; \theta, S) \quad (3)$$

### 3.3.2 Generalization to all passes

We now extend the concept to all passes by including high passes for which  $\beta > 0$ . High passes are slightly more involving since the ball may be too far up to be reachable



**Fig. 5** Running example: Blue player at bottom has possession of the ball (large black circle). Left: Cumulative interception likelihoods  $p_{CI}^o(x; \theta)$  with  $\theta = (\alpha, 0, 25 \text{ m/s})$ . Right: Success likelihood  $p_s^r(x, \theta)$  with  $\theta = (\alpha, 0, 25 \text{ m/s})$

for a player. We thus make use of the ball model  $p(z | d; \beta, v)$  in Eq. (1) that estimates the density of the height  $z$  of the ball at a given distance  $d$  and velocity  $v$ .

The idea is to incorporate the notion of  $z$ -reachability into the interception probability  $p_I$ . Let  $p_z(z < h | d; \beta, v)$  be the cumulative distribution that the ball is lower than height  $h$  at distance  $d$  when a pass was played with initial parameters  $\beta$  and  $v$ . Furthermore, let  $h_I^i$  be the maximum height at which a ball can be intercepted by player  $i$ .<sup>1</sup> It follows that the interception probability of the  $i$ -th player can be written as a product of the  $xy$ -reachability given by the movement model of player  $i$  and the  $z$ -reachability, given by

$$p_I^i(m; \theta, S) = p_R^i(m; t(\|m - b\|; \beta, v), S) \cdot p_z(z < h_I^i | \|m - b\|; \beta, v).$$

So far, the probability of a successful pass along  $a$  does not take into account whether the receiver can control the ball. Consider a pass over 10 m that is played with 30 m/s and reaches the receiver at a height of 1.5 m. This pass is reachable but certainly not controllable. Therefore we introduce a control-likelihood

$$p_C(x; \beta, v) = f(u(\|x - b\|; \beta, v)) \cdot p_z(z < h_C | \|x - b\|; \beta, v)$$

that is a function of the predicted speed of the ball  $u$  when it reaches the receiver and the likelihood that the ball is below  $h_C = 0.5 \text{ m}$ . Putting everything together, the likelihood that a pass with initial parameters  $\theta$  can be successfully received by player  $r$  at position  $m$  is

<sup>1</sup> For simplicity, we use  $h_I^i = 1.9 \text{ m}$  for all players in this paper, however, individual values can be used to distinguish e.g., short vs large players, athletic and jumping skills, etc.

$$p_s^r(m; \theta, \mathcal{S}) = p_\alpha(m | \theta, \mathcal{S}) \cdot p_C(m; \beta, v) \cdot p_I^r(m; \theta, \mathcal{S}) \prod_o (1 - p_{CI}^o(m; \theta, \mathcal{S})). \quad (4)$$

### 3.4 Full availability

The previous sections showed how to compute likelihoods of successful passes along vectors  $a(\theta)$ . However, while a player may intend to hit the ball with certain initial parameters  $\theta = (\alpha, \beta, v)$ , there will generally be a (possibly slight) deviation  $\Delta\theta = (\Delta\alpha, \Delta\beta, \Delta v)$  from the intended pass trajectory, determined by the individual skill and circumstances such as pressure on the ball carrier or running speed. We will now present our model  $A^r(\psi)$  that determines the overall likelihood of a successful pass to  $r$  using uncertainty parameters  $\psi = (\sigma_\alpha, \sigma_\beta, \sigma_v)$  which can be understood as *skill* parameters. We assume that deviations are drawn from a normal distribution with mean 0 and diagonal covariance matrix

$$(\Delta\alpha, \Delta\beta, \Delta v) \sim \mathcal{N}(0, \text{diag}(\sigma_\alpha, \sigma_\beta, \sigma_v))$$

It follows that the expected success for intended initial pass parameters  $\theta$  is

$$\begin{aligned} p_\Delta^r(\theta, \mathcal{S}) &= \int_{\Delta\theta} p_a^r(\theta + \Delta\theta, \mathcal{S}) \mathcal{N}(\Delta\theta | 0, \text{diag}(\psi)) d\Delta\theta \\ &= \mathbb{E}[p_a^r(\theta + \Delta\theta, \mathcal{S}) | \Delta\theta \sim \mathcal{N}(\Delta\theta | 0, \text{diag}(\psi))] \\ &= \mathbb{E}_{\Delta\theta}[p_a^r(\theta + \Delta\theta, \mathcal{S})]. \end{aligned}$$

Using the expected success we compute the final availability score as

$$A^r(\psi, \mathcal{S}) = \max_{\theta} p_\Delta^r(\theta; \mathcal{S}).$$

In other words, the passer chooses the best intended option to pass to player  $r$ .

## 4 Modelling players

We now present the player reachability model  $p_R^i(m, t; \mathcal{S})$  that determines the likelihood that player  $i$  can reach position  $m$  in time  $t$ . The idea is to derive the reachability model from an underlying motion or movement model that estimates the future density of the position of a player, conditioned on the game state  $\mathcal{S}$ . Several different movement models have been proposed in the literature, for example using simplified physics (Taki et al. 1996; Fujimura and Sugihara 2005) or frequency statistics (Brefeld et al. 2019).

However, as we will show below, the reachability model  $p_R^i(m, t; \mathcal{S})$  estimates reachability based on a cumulative distribution function of moving to position  $m$  in time  $t$ . Computing cumulative distribution functions can be computationally very expensive if one has to rely on sampling in order to approximate the true cumulative distribution

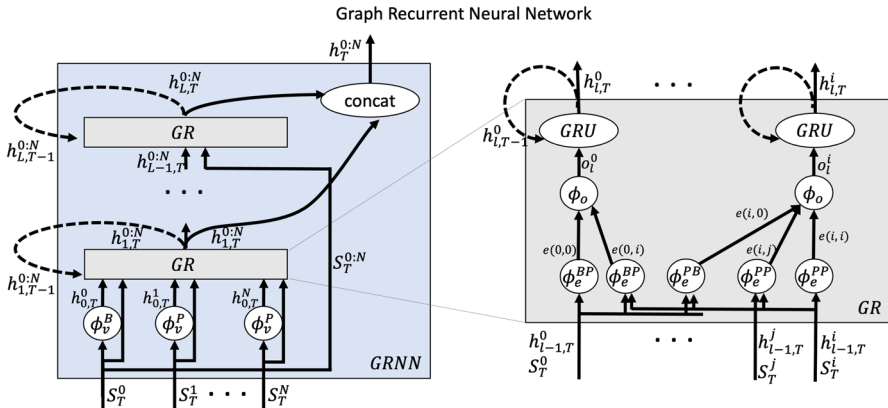


Fig. 6 Left: GRNN model. Right: One layer of GRNN

function (CDF). We thus follow a different approach and represent  $p_M^i(m | t, S)$  as a graph recurrent mixture density network (GRMDN) for which the desired function can be computed in closed form given the parameters of the distribution. GRMDNs are built on two blocks, graph recurrent neural networks (GRNNs, Yeh et al. 2019; Battaglia et al. 2018) which allow us to model interactions between players and ball and mixture density networks (MDNs, Bishop 1994) that represent movement distributions of players as Gaussian mixtures.

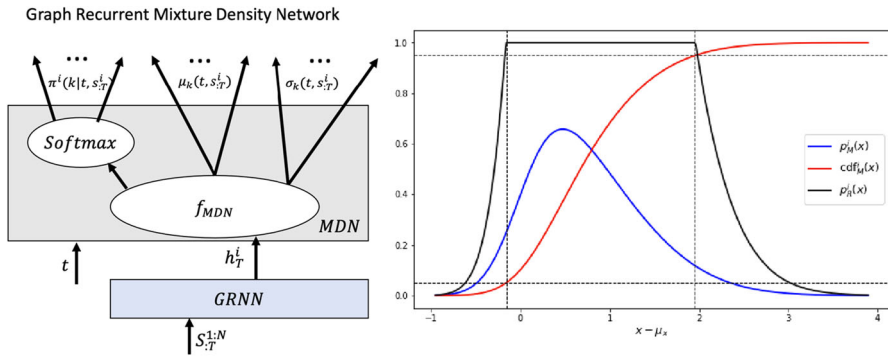
#### 4.1 Player and ball interactions

We use *GRNNs* to model the interactions between players and ball using a fully connected graph structure. Players and ball correspond to nodes in that graph and edges represent their relations. This part of the model is depicted in Fig. 6, left, and consists of several layers. One layer or block *GR* of the model is shown in Fig. 6, right. To describe such a layer, or block, of the graph network, recall that state  $S_T^i$  contains the position and speed of player/ball  $i$  at time  $T$ , as well as team and ball possession indicators.

The  $\ell$ -th block of the graph network takes as input the states  $S_T^i$  for all  $0 \leq i \leq 22$  as well as the outputs of layer  $\ell - 1$  given by feature vectors  $h_{\ell-1,T}^i$ . Since the graph is fully connected, every player/ball  $i$  is connected to any other players/ball  $j$  via typed edges  $e_{\ell}^{type}(i, j)$  with  $type \in \{PP, BP, PB\}$  representing directed edges either between two players (*PP*), between ball and player (*BP*), or between player and ball (*PB*). Edge features  $\phi_e^{type}$  are computed via attention functions  $\alpha_{\ell}^{type}(\cdot; \theta_{\ell})$ , depending again on the edge *type* and per-node-functions  $f_v$ , which are fully connected subnets, such that

$$e_{\ell}^{type}(i, j) = \phi_e^{type} \left( S_T^i, S_T^j, h_{\ell-1,T}^j; \theta_{\ell} \right) = \alpha_{\ell}^{type} (S_T^i - S_T^j; \theta_{\ell})^{\top} f_v(h_{\ell-1,T}^j)$$

$$o_{\ell}^i = \phi_o(\{e(i, j) : j \in \{0, \dots, N\}\}) = \sum_j e_{\ell}(i, j).$$



**Fig. 7** Left: GRMDN model. Right: Visualization of reachability likelihood  $p_R^i(m, t; S)$  and its derivation from  $p_M^i(m | m', S, t)$ . Cutoff value  $c_{CO} = 0.05$  is visualized with dotted lines

The intermediate representations  $o_\ell^i$  are fed into standard gaited recurrent units (GRU, Cho et al. 2014) to model time-dependent behavior and compute the output of the layer as  $h_{\ell,T}^i = \text{GRU}(h_{\ell,T-1}^i, o_{\ell,T}^i)$ . To sum up, the layer of the GRNN shown in Fig. 6, right, is denoted by

$$h_{\ell,T}^{0:22} = GR(h_{\ell-1,T}^{0:22}, h_{\ell,T-1}^{0:22}, S_T^{0:22}).$$

The full GRNN, displayed in Fig. 6, left, concatenates  $L$  such layers and is given by the following equations,

$$h_T^{0:22} = (h_{1,T}^{0:22}, \dots, h_{L,T}^{0:22}) = GRNN(h_{0,T}^{0:22}, h_{T-1}^{0:22}, S_T^{0:22}),$$

where the inputs  $h_{0,T}^i$  of the first layer are computed by a single layer fully connected network  $\phi_v(S_T^i)$ . The inputs  $S_T^i$  to the network are detailed in Appendix B.

## 4.2 Movement model

The distribution of future positions  $m$  that can be attained in time  $t$  are represented as a mixture model with  $k$  Gaussian mixtures,

$$p_M^i(m | t, S) = \sum_{k=1}^K \pi^i(k | t, S^i) \mathcal{N}(m | \mu_k^i(t, S^i), \sigma_k^i(t, S^i) \cdot \mathbb{I}) \quad (5)$$

and is realized by a *mixture density network* (MDN, Bishop 1994). The MDN takes the  $v_T^i$  outputs of the GRNN for player  $i$  and the time horizon  $t$  as the input to a single layer fully connected subnet  $f_{MDN}$ . Categorical mixture distribution  $\pi^i$  is computed from the outputs of  $f_{MDN}$  with a standard softmax. Gaussian means  $\mu_k^i(t, S^i) \in \mathbb{R}^2$  and variances  $\sigma_k^i(t, S^i) \in \mathbb{R}^2$  are predicted using linear and exponential, resp., activation functions, where  $\mathbb{I}$  is the diagonal identity matrix. Figure 7, left, shows an overview.

The two described building blocks GRNN and MDN form a joint *graph recurrent mixture density network* and are trained simultaneously.

### 4.3 Player reachability

Having computed the movement model  $p_M^i(m | t, \mathcal{S})$ , we are now ready to derive the reachability distribution  $p_R^i(m, t; \mathcal{S})$ . While the movement model describes where the average player will be in time  $t$ , the reachability model estimates which positions *can be reached* in time  $t$ . Reachability is modeled using the movement model by defining a (pseudo) cumulative distribution function of positions and using a *cdf* cutoff parameter that defines which positions are reachable with probability 1. All positions that lie outside that *cdf* cutoff are reachable with probability below one. Figure 7, right, shows a visualization of that approach in 2D.

Let the expected position of player  $i$ , exactly  $t$  seconds into the future, given by

$$\mu_m = \mathbb{E}_{p_M^i(m | t, \mathcal{S})}[m].$$

We define the (pseudo) cumulative distribution function at position  $m$  as the *cdf* of the one-dimensional distribution defined on the line that goes through the mean  $\mu_m$  and  $m$ .

$$\text{cdf}_M^i(m | t, s:T) = \frac{1}{Z} \int_{c=-\infty}^1 p_M^i(\mu_m + c(m - \mu_m) | t, \mathcal{S}) dc$$

where  $Z$  is the partition function given by

$$\text{with } Z = \int_{c=-\infty}^{\infty} p_M^i(\mu_m + c(m - \mu_m) | t, \mathcal{S}) dc$$

Based on the *cdf*, we assume that player  $i$  can reach position  $m$  in time  $t$  with probability 1 if its  $\text{cdf}_M^i(m | t, \mathcal{S})$  is between cutoff-values  $c_{co}$  and  $1 - c_{co}$ . Otherwise, the reachability likelihood is scaled with  $1/c_{co}$  to guarantee a smooth reachability surface, as shown below and in Fig. 7, right,

$$p_R^i(x, t; s:T) = \begin{cases} 1, & \text{if } \text{cdf}_M^i(x | t, s:T) \in [c_{co}, 1 - c_{co}] \\ \text{cdf}_M^i(x | t, s:T) / c_{co}, & \text{if } \text{cdf}_M^i(x | t, s:T) < c_{co} \\ (1 - \text{cdf}_M^i(x | ts:T)) / c_{co}, & \text{if } \text{cdf}_M^i(x | t, s:T) > 1 - c_{co}. \end{cases}$$

The cutoff parameter is tuned on data in order to maximize observed reachability of pass receivers and minimize reachability of defenders that did not intercept the ball. That is, we define a binary classification problem such that for each observed pass in the data we create one positive example (pass receiver intercepts the ball at position  $m$ ) and one negative example for each defender that did not intercept the ball along the ball trajectory.

## 5 Experimental evaluation

We evaluate our model on passes extracted from 58 Bundesliga games from the 2017/18 season. The data comes in form of tracking and event logs. The tracking data is sampled at 25 fps and contains positions of all players and the ball at each frame/timestamp. Pass information is extracted from corresponding event data. This comprises the passing player, the time of the pass, the target position of the pass as well as the receiving player. However, the receiving player could be an opponent who intercepts the pass. In this case, data does not contain ground-truth about the intended receiver. We overcome this problem by identifying the most likely team mate according to the initial direction of the ball at the time of the pass as described in Appendix A. The entire data consists of 38,851 passes with 33,561 successful and 5290 intercepted passes. This sums up to an average success rate of 0.86.

Model selection is conducted via a five-fold cross-validation on 54 games. We report average results on an independent test set consisting of the remaining 4 games. The optimized architecture of the GRNN consists of two  $GR$  layers where both layers have a GRU width of 1024 and the attention function  $\alpha_l^X$  as well as function  $f_v$  possess a single hidden layer with 256 units. Input functions  $\phi_v$  have a single hidden layer with 64 units. The MDN consists of a Gaussian mixture with  $k = 20$  components and function  $f_{MDN}$  has a single hidden layer with 256 hidden units. The cutoff parameter of the player reachability model is set to  $c_{co} = 0.005$  and the uncertainty or skill parameters  $\psi = (\sigma_\alpha, \sigma_\beta, \sigma_v)$  are set to  $(3^\circ, 6^\circ, 2m/s)$ . The ball dynamics model  $t(d, \beta, v)$  uses a polynomial kernel of degree 7 and the speed model  $u(d, \beta, v)$  a kernel of degree 2.

**Baseline Approach** We compare our approach to Peralta Alguacil et al. (2020) that grounds on the movement model by Fujimura and Sugihara (2005) and the work by Spearman et al. (2017) and considers players as physical objects whose dynamics are described by an equation of motion with internal and external forces. We also use the proposed logistic distribution to estimate final player movement probabilities. Our approach thereby differs from Peralta Alguacil et al. (2020) in that our model learns the movement distributions from observed player data only instead of using a solely physics-based model based on approximated properties of soccer players. We also test against the ball dynamics model as described in the appendix in Peralta Alguacil et al. (2020) which again is a model that describes the ball movement based on its approximated physical properties. In contrast, our model learns the ball dynamics from observed ball trajectory data. All parameters of the baseline are set according to the values proposed by the authors in Peralta Alguacil et al. (2020).

**Movement Models** We begin the empirical evaluation by comparing our player reachability model with the motion model of the baseline. To do so, we compare the likelihoods that the true receiver of a pass can actually reach the observed receiving position. We take both, successful and unsuccessful, passes into account, such that the true receiver could be an intercepting opponent player. Thus, we compute interception probabilities for all opponents and report the maximum over all opponents. Thus, an accurate model assigns higher probabilities to the true receiver and lower probabilities to non-intercepting opponents. Table 2 shows the results. Our approach



**Table 2** Average interception likelihoods of observed receivers and uninvolved opponents on observed passes

	True receiver	Uninvolved opponents	AUC
Our approach	0.850	0.202	0.913
Peralta Alguacil et al. (2020)	0.772	0.100	0.892

**Table 3** Comparison of ball dynamics of our approach and the baseline approach to observed passes

	Observed passes	Our approach	Peralta Alguacil et al.
Duration of pass (sec)	1.128±0.024	1.135±0.026	0.947±0.018
Max height of pass (m)	0.52±0.05	0.51±0.06	0.36±0.04

distinguishes by much higher average likelihoods for true receivers but also assigns higher likelihoods to uninvolved defenders who did not intercept the ball.

The table also shows AUC values for both methods where we compare the ability to predict whether a pass is going to be successful or not. We therefore compute the pass success probability  $p_s^r(m)$  in Eq. (4) at the observed reception position  $m$  and, instead of using a ball model, compute the cumulative interception probabilities  $p_{CI}^o(m)$  along the observed ball trajectory. The computed success probabilities are compared to the observed outcomes of the pass (success or interception) to give the AUC of the prediction. The resulting AUCs support the previous outcome, our reachability model is in fact more accurate in predicting whether a player can reach a certain position in time.

**Ball Dynamics** We now compare our ball dynamics model to observed passes and the baseline model to instantiate whether our models predict realistic passes. Table 3 shows experimental results that compare our models for ball dynamics to the baseline. The evaluation is performed on all observed passes in our data. We test model  $t(d; \beta, v)$  that predicts the time it takes a ball to reach a certain distance  $d$  given the initial  $z$ -angle  $\beta$  and speed  $v$ .  $\beta$  and  $v$  are estimated over the first 5 frames of observed ball trajectories and the error is computed between observed and predicted time at the observed ball reception positions. The first row in Table 3 shows that our model is better in predicting the expected time the ball needs to cover a certain distance than the baseline. The second row shows results for predicting the maximum height of a pass given  $\beta$  and  $v$  where we took the mean of distribution  $p(z(t) | \beta, v)$  in Table 1. Again, our model outperforms the baseline.

**Successful Passes** Next, we compare predictions about whether a pass will be successful or not. For every pass, the first four frames are used to estimate the initial direction and speed of the ball. On the basis of these estimates, the models then predict the ball trajectory. In addition to the full baseline (Peralta Alguacil et al. 2020), we also compare against a hybrid model that uses the physics-based reachability model from Peralta Alguacil et al. (2020) but our ball model. We evaluate the three models by comparing their predictions with the true outcome. Again, we use AUC to measure predictive performance. Table 4 shows the results.

**Table 4** Predicting the outcome of a pass

	AUC	Avg. success prob. of successful pass	Avg. success prob. unsuccessful pass
Our Model	0.886	0.830	0.181
Hybrid	0.879	0.862	0.297
Peralta Alguacil et al.	0.879	0.846	0.267

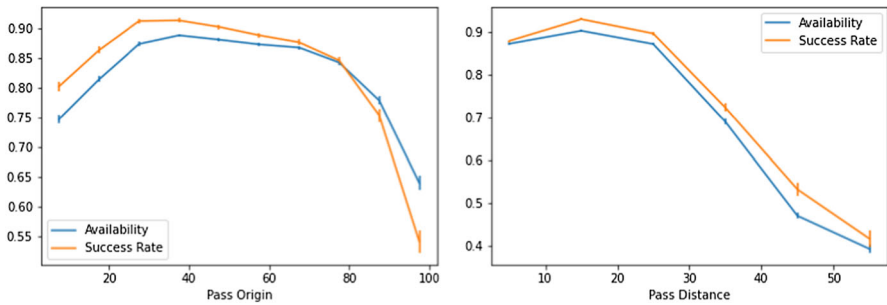
Our model outperforms both baselines in terms of AUC. However, in contrast to Table 3, which shows that ball dynamics of the baseline are clearly inferior to our proposed approach, this result does not imply inferior performance when predicting outcomes of passes. Note that the baseline performs, with a predictive accuracy of about 0.880, significantly better in our experimental setup than in Spearman et al. (2017) who report an accuracy of 0.819. Naturally, this effect may come (in parts) from using different data. For example, our Bundesliga data has an average rate of successful passes of 0.86 while Spearman et al. (2017) reports only 0.789.

**Receiving Position** The pass reception probability  $p_s^r(m; \theta)$  describes the likelihood that a pass to receiver  $r$  can be successfully completed at position  $m$ . We empirically quantify the estimate by comparing the true positions of passer and receiver to the predicted success rate at those positions. Again, we consider successful as well as intercepted passes. Consider the example in Fig. 1 which serves as our running example in the paper. The black circle in the left part shows the actual pass reception position of the pass and its corresponding color-coded pass reception likelihood while the right figure shows the instance of the first touch of the receiver. Our evaluation on all available pass data shows that for successful passes, the average success likelihood and standard deviation of the end position is  $0.789 \pm 0.328$  whereas the average likelihood of bad passes is  $0.169 \pm 0.269$ . In other words, the model is able to predict reliably, whether passes to certain positions can be successful. This is also highlighted by a corresponding AUC value of 0.913.

**Availability** After having shed light on different aspects of the proposed approach, we now turn to evaluate the main contribution of the paper, the predicted Availability scores  $A^r(\psi)$ . For every pass in the data, we compare the (expected) availability of the (intended) receiver to the true outcome of the pass. In our evaluation, the average Availability score and standard deviation of successful and unsuccessful passes are  $0.884 \pm 0.23$  and  $0.572 \pm 0.24$ , respectively. Moreover, measuring AUC on Availability and true outcome yields a score of 0.870. This allows to draw the conclusion that the computed Availability scores correlate highly with success of passes. Note that the overall success rate of observed passes of 0.86 is almost matched by an average predicted Availability of 0.84.

Figure 8, left, shows average success rates and Availability scores for different pass origins,<sup>2</sup> while the right part of the figure shows those quantities w.r.t. pass distance.

<sup>2</sup> The  $x$ -axis of the figure is identical to the  $x$ -coordinate of the position of the pass giver at the time of the pass. The attacking team plays always from left to right so that 0 is identical to the own goal while 100 corresponds to the opponent's goal.

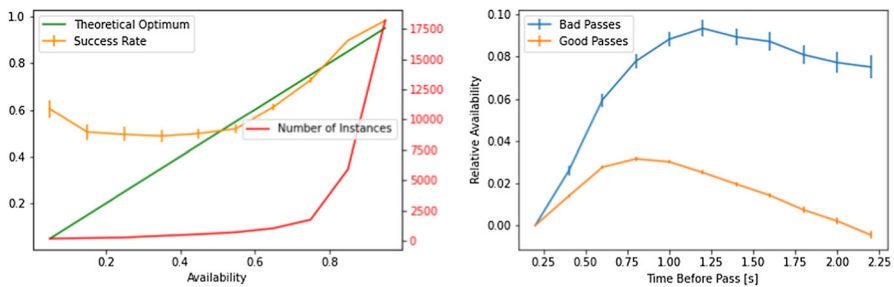


**Fig. 8** Evaluation of passes. Left: origin of the pass in meters from goal line. Right: distance of pass in meters

Both figures show that average success rates and Availabilities line up nicely. However, a limitation of our model can also be clearly identified in the left part of the figure: since we do not incorporate the notion of “pressure” on the pass giver into the model, areas closer to the own goal have higher success rates than the estimated ones. On the other hand, the true success rates are smaller than the predictions in areas closer to the opponent’s goal. We credit this finding to the expected pressure on pass giver at the time of a pass. The closer the passer is to the opponent’s goal, the more pressure is issued by the opponents and the player has to act in smaller spaces and in shorter time windows to play a pass.

Note that the pressure argument does not translate to the pass distance as Fig. 8, right, demonstrates. A reason for this can be observed in Fig. 9, left. The figure shows the mean success rates for bins of Availability scores. An optimal scoring function would yield the diagonal green line where Availability scores and average success rates would perfectly align. However, as observed in Fig. 9, left, this holds only for Availability scores above 0.5. The observed success rates for smaller values exceed the expectations significantly. An in-depth analysis of the results shows that this is mainly due to imperfections in the recorded data (cf., Linke et al. 2020) for a quantitative evaluation of tracking data) and the small number of passes with low Availability, as shown in red in the figure. As an example, consider Fig. 10, left, where the recorded ball position (thick black circle) suggests that the ball is located at the player’s heels. However, in reality, the ball is positioned half a meter to the left. The computed Availability uses the recorded position to come to the conclusion that the player at the bottom of the image can only receive a high pass to the left of him with probability 0.26. However, the real ball reception is at the light circle, following a low pass. In the right part of the figure, the recorded ball position is slightly below the player’s position, whereas the real ball position is about a meter above the recorded position. Because of that, computed success probabilities  $p_s^r(x, \theta)$  are non-zero mainly below the receiver (to the right in playing direction) because according to the data the ball was one meter below the real position. The actual ball reception position is unreachable according to the data.

**Availability over Time** We now investigate how Availability changes over time prior to a pass. We differentiate between good and bad passes and measure relative



**Fig. 9** Left: Success rate vs. predicted Availability. Right: Availabilities over time before a pass



**Fig. 10** Pass situations where real ball position and recorded data position differ significantly. Thick black circle indicates recorded ball position, light black circle indicates ball reception position. Orange shades indicate computed potential receiving positions. Left: The recorded ball position suggests that the ball is at the positions of the player's heels. In reality, the ball is half a meter to the left. Computed availability indicates that according to recorded position the player at the bottom of the image can only receive a high pass to the left of him with probability 0.26. However, the real ball reception is at the light circle, following a low pass. Right: Recorded ball position is slightly below the player's position. Real ball position is about a meter above the recorded position. For better visibility, a larger version of this figure is shown in Appendix C

Availabilities by subtracting the score at the time of the pass from the respective score  $s$  seconds before the pass. Figure 9, right, shows the resulting relative Availabilities over time. For all passes, relative Availabilities decline rapidly before the actual pass. This is explained by passing players turning their bodies into the direction of the pass before the pass is being played, which in turn means that opponents read their intentions from their posture and try to close the passing window. This effect is stronger for bad passes, which partly explains why they are unsuccessful.

**Comparison to Expert Ratings** To evaluate validity against an external criterion, we compare computed Availability scores to those estimated by human observers. Four soccer coaches rated the availability of players in 60 situations. We included situations only, in which the ball possessing player had full ball control (ball flat, orientation to opponent's goal, at least one second ball possession, low pressure). To consider different types of tactical constellations, we distributed the 60 situations evenly on build-ups, transitions, and situations in the box. The coaches are presented

**Table 5** Correlations between experts and our method

	Obs1	Obs2	Obs3	Obs4	Model
Obs1	1.000	0.806	0.740	0.800	0.769
Obs2	0.806	1.000	0.813	0.842	0.708
Obs3	0.740	0.813	1.000	0.840	0.638
Obs4	0.800	0.842	0.840	1.000	0.701
Model	0.769	0.708	0.638	0.701	1.000

**Table 6** Average rating of successful (+1) and unsuccessful (−1) passes

	Obs1	Obs2	Obs3	Obs4	Model	#
+1	−0.25	0.25	0.12	0.12	0.12	8
−1	1.33	1.46	1.24	1.32	1.77	48
AUC	0.826	0.814	0.729	0.787	0.889	

an image from the video and rated the Availability of every team mate of the ball possessing player in that situation with a score in the set  $\{-2, -1, 0, 1, 2\}$ , where  $-2$  corresponds to lowest Availability and  $+2$  to highest.

To provide a common understanding of the ratings, the experts were instructed before the experiment. In short, for a rating of  $+2$ , a player should be able to receive the ball safely if the passing player does not make a terrible mistake. A rating of  $+1$  represents a situation in which a player has a good chance to receive the ball, but there is maybe a small interception chance for an opponent. A rating of  $0$  indicates a 50:50 chance to get the ball, and so on. The experts were encouraged to rate the situations based on their personal understanding of soccer. Since Availability is a probability and naturally ranges between zero and one, we binned the scores evenly in intervals of length  $0.2$  so that an Availability within  $[0, 0.2)$  corresponds to a rating of  $-2$ , interval  $[0.2, 0.4)$  is mapped to  $-1$ , and so on.

Table 5 shows correlations between the expert ratings and the binned Availabilities. The table shows a strong positive correlation between the coaches and our computational model. Differences between the observers indicate that it may not be possible to entirely objectify Availability. This is quite typical for non-trivial tactical concepts and has also been reported for other metrics (Link et al. 2016). However, correlations between experts are generally higher than the ones between expert and model with Obs1 being the only exception. Though the finally outcomes are comparable, this result suggests that our model rates Availabilities slightly different than the experts.

Interestingly, Table 6 shows that our algorithm in fact rates Availabilities slightly better than the experts. The table summarizes average completion rates of passes as an indicator. That is, we compare average ratings of successful and non-successful passes as follows. A mean of  $0.12$  for non-successful passes is comparable to the rating of the experts. For successful passes, however, an average rating of  $1.77$  is significantly higher than those of the experts. This result also becomes obvious when comparing AUCs, where our algorithm significantly outperforms all experts.

## 6 Scenarios for application

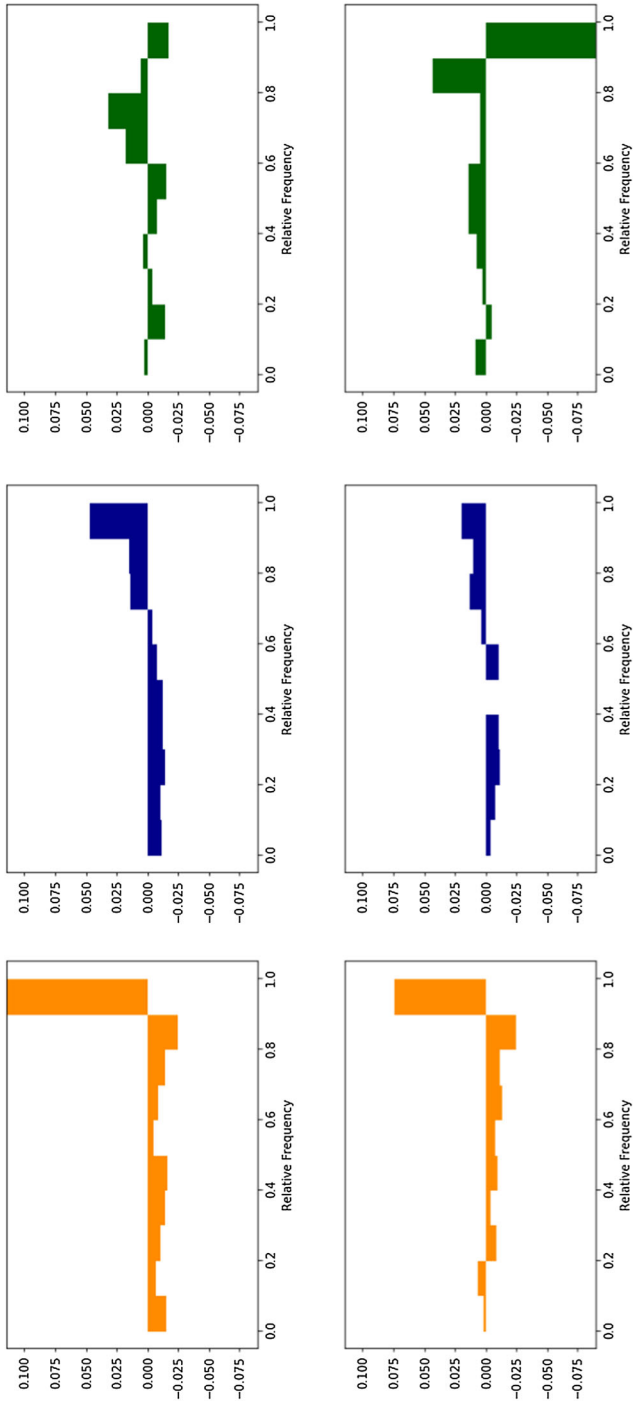
From the perspective of performance analysis, our model presents a set of interesting applications. As an example, coaches and other experts can use Availability to characterize the passing tactics of players, that is, does a player only try easy passes or is the proportion of difficult passes noticeably high? The model can also be used to rate the actual passing ability of players. While coaches have a very good understanding of their players' passing capabilities, by seeing them in training and matches on a daily basis, quantifying those abilities is still a hard task. While average passing statistics are readily available in a wide variety of sources<sup>3</sup> the raw pass count and average success rates do not show the full picture. E.g. attacking players generally have a lower pass percentage than defensive players, simply because they operate in tighter spaces and have fewer, if any, available passing options. Therefore, it should prove beneficial to compare a player's observed success rates to the expected one in order to more objectively quantify a player's passing ability.

Figure 11 and Table 7 show preliminary results for both use cases. In Fig. 11, we compare pass selections of defenders, midfielders, and forwards of Bayern Munich. The results show that the position in which a player generally operates has a significant influence on the kinds of passes a player attempts. Defenders take fewer risks when passing, for one because the outcome of an unsuccessful pass could more likely result in a scoring chance by the opponent but also because they have more available passing options. On the other hand, forwards take more risks, either because the potential reward is higher, or because simple passes are not possible due to high pressure of defenders. Table 7 shows the top 10 ranked players in our data w.r.t. their expected versus observed pass percentage. We would like to note that our evaluation data consists of only 58 games and several players did not have enough passes in the data to validate reliable analyses. We only considered players with at least 350 passes which left us with 65 players from FC Bayern Munich, Hamburger SV, TSG 1899 Hoffenheim, FC Schalke 04, and SG Eintracht Frankfurt. Still, the top 10 exhibits an impressive over-representation of Bayern Munich players. Sebastian Rudy, Joshua Kimmich, Arturo Vidal, David Alaba, Corentin Tolisso, and Niklas Süle all played for Bayern that season with only 2 Schalke players (Daniel Caligiuri, Benjamin Stambouli) and 2 Hamburg players (Kyriakos Papadopoulos, Gotoku Sakai) in the list.

## 7 Conclusion

We presented and evaluated a data-driven approach to estimating Availability of soccer players. The investigated model leverages graph recurrent neural networks to predict whether players can intercept the ball in a given time to compute the probability of a successful pass along a ball trajectory. By computing all possible ball trajectories using trained models for ball dynamics, we showed how to aggregate those potential passes into a single value that represents the overall Availability of a player. Experimental

<sup>3</sup> e.g. Opta Sports data via e.g. <https://www.whoscored.com/Regions/81/Tournaments/3/Seasons/7405/Stages/16427/PlayerStatistics/Germany-Bundesliga-2017-2018>



**Fig. 11** Comparison of pass selections of players from Bayern Munich via histograms over pass success probabilities of observed passes. Histograms show frequency distances compared to the average frequencies over all players. Orange: two forwards Robert Lewandowski (top) and Arjen Robben (bottom). Green: two midfielders Mats Hummels (top) and David Alaba (bottom). Blue: two defenders Thiago (top) and Sebastian Rudy (bottom). Pass selections differ notably depending on position. Defenders play more passes with low risk (rightmost bar), whereas forwards play significantly less of those



**Table 7** Player Analysis

Player	Availability	Success	Count	Difference
S. Rudy	0.898	0.944	412	0.046
K. Papadopoulos	0.889	0.935	367	0.046
J. Kimmich	0.907	0.945	855	0.038
A. Vidal	0.912	0.946	405	0.034
D. Caligiuri	0.861	0.894	500	0.033
D. Alaba	0.921	0.950	515	0.029
C. Tolisso	0.921	0.950	458	0.029
G. Sakai	0.874	0.902	540	0.028
N. Süle	0.937	0.965	510	0.028
B. Stambouli	0.888	0.916	392	0.028

evaluation showed that this overall model outperforms the state-of-the-art approach on 58 professional soccer matches. Additionally, our experiments indicate that the model can even outperform soccer coaches in assessing the Availability of soccer players.

**Acknowledgements** We would like to thank Hendrik Weber and Sportec Solutions / Deutsche Fussball Liga (DFL) for providing the tracking data.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

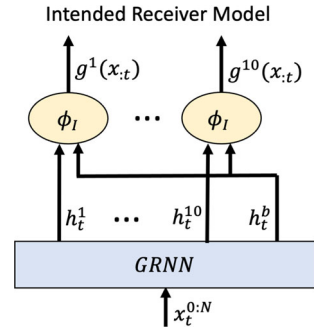
**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A Intended receiver model

The presented model allows to estimate the probability that a player can successfully receive a pass. To learn a general model that can deal properly with intercepted passes, we need to know the intended receiver of an intercepted pass. This information is usually not contained in commercial data and manual annotation tedious and expensive. We thus propose a way to estimate the intended receiver from data. Note that this model is not part of the presented model in the main part so that training and application of this approach needs to be done *before* training the model in Sect. 4. By doing so, the model in Sect. 4 can take the provided labels into account.

A bad pass may have two possible outcomes: It is either intercepted by an opposing player or the ball goes out of bounds. To be able to observe the initial trajectory of the ball, which is vital for predicting the receiver, we only consider passes where the ball travels at least 0.24 sec (6 frames) before it is either intercepted or goes out of bounds.

**Fig. 12** Model for predicting intended receivers



Our approach is to learn from data of successful passes, where the intended receiver is also the observed receiver, and simply use that model to predict intended receivers of unsuccessful passes.

Let  $t_p$  be the recorded time of a successful pass and  $y_{t_p}$  the true receiver of that pass. The pass is turned into a training example by computing the initial direction of the pass on the first 6 frames and using the positions and trajectories of players and ball in the subsequent frame as the input. The output is simply the true receiving player, e.g., by a one-hot encoding.

Figure 12 shows an overview of the model. Let  $\phi_I(v_T^k, v_T^b)$  be a score function of outputs  $h_T^k$  of potential receiver  $k$  and passer  $h_T^b$  of the GRNN model as described in Sect. 4. The model minimizes the cross-entropy loss between real labels and scores and thus outputs a softmax distribution over all possible receiving players.

## Appendix B Input to GRNN

In our approach, a player is represented by the  $(x, y)$  coordinates of her position on the pitch and her speed in  $x$  and  $y$  direction given by  $s_x$  and  $s_y$ , respectively. Additional indicator variables and flags inform the model about

- the  $x$ -direction of the opponent's goal ( $\rho_d \in \{-1, +1\}$ )
- whether the player's team has ball possession ( $\rho_t \in \{0, 1\}$ )
- whether player has ball possession ( $\rho_{bp} \in \{0, 1\}$ ) and
- whether player is a goal keeper ( $\rho_{gk} \in \{0, 1\}$ )

The representation for the  $i$ -th player at time  $T$  is thus given as a feature vector

$$S_T^i = (x^i, y^i, s_x^i, s_y^i, \rho_d^i, \rho_t^i, \rho_{bp}^i, \rho_{gk}^i)^\top$$

where we omitted the time index  $T$  for readability. The input to the GRNN is now given by the set of feature vectors of all players and the ball, where the latter is the same as for players but all flags and indicators are set to zero, i.e.,

$$S_T^{ball} = (x^i, y^i, s_x^i, s_y^i, 0, 0, 0, 0)^\top.$$



**Fig. 13** Top: Figure 1 (left part). Bottom: Figure 1 (right part)

## Appendix C Figure 13 and 14

For better visibility, we show larger sized versions of Figures 1 and 10 here.



**Fig. 14** Top: Figure 10 (left part). Bottom: Figure 10 (right part)

## References

- Akira F, Kokichi S (2005) Geometric analysis and quantitative evaluation of sport teamwork. *Syst Comput Jpn* 36(6):49–58
- Bangsbo J, Peitersen B (2004) Offensive soccer tactics. Human Kinetics Publishers, Champaign
- Battaglia PW, Hamrick JB, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malinowski M, Tacchetti A, Raposo D, Santoro A, Faulkner R et al (2018) Relational inductive biases, deep learning, and graph networks. [arXiv:1806.01261](https://arxiv.org/abs/1806.01261)
- Bishop CM (1994) Mixture density networks. Technical Report NCRG\_94\_004, Aston University
- Bransen L, van Haaren J, van de Michel V (2019) Measuring soccer players' contributions to chance creation by valuing their passes. *J Quant Anal Sports* 15(2):97–116

- Brefeld U, Lasek J, Mair S (2019) Probabilistic movement models and zones of control. *Mach Learn* 108(1):127–147
- Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the conference on empirical methods in natural language processing*
- Dauxais Y, Gautrais C (2019) Predicting pass receiver in football using distance based features. In: Van Davis J, Haaren J, Brefeld Zimmermann AU (eds) *Machine learning and data mining for sports analytics*. Springer, New York City, pp 145–151
- Felsen P, Lucey P, Ganguly S (2018) Where will they go? Predicting fine-grained adversarial multi-agent motion using conditional variational autoencoders. In *Proceedings of the European conference on computer vision*
- Fernández J, Bornn L, Cervone D (2021) A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Mach Learn* 110:1389–1427
- Fernández J, Bornn L (2018) Wide open spaces: a statistical technique for measuring space creation in professional soccer. In *Proceedings of the Sloan sports analytics conference*
- Fernández J, Bornn L, Cervone D (2019) Decomposing the immeasurable sport: a deep learning expected possession value framework for soccer. In *Proceedings of the MIT Sloan sports analytics conference*
- Fournier-Viger P, Liu T, Chun-Wei Lin J (2018) Football pass prediction using player locations. In: Van Davis J, Haaren J, Brefeld Zimmermann AU (eds) *Machine learning and data mining for sports analytics*. Springer, New York City, pp 152–158
- Goes Floris R, Matthias K, Meerhoff Laurentius A, Lemmink Koen APM (2019) Not every pass can be an assist: a data-driven model to measure pass effectiveness in professional soccer matches. *Big Data* 7(1):57–70
- Goes FR, Meerhoff LA, Bueno MJO, Rodrigues DM, Moura FA, Brink MS, Elferink-Gemser MT, Knobbe AJ, Cunha SA, Torres RS, Lemmink KAPM (2021) Unlocking the potential of big data to support tactical performance analysis in professional soccer: a systematic review. *Eur J Sport Sci* 21(4):481–496
- Graves Alex (2013) Generating sequences with recurrent neural networks. [arXiv:1308.0850](https://arxiv.org/abs/1308.0850)
- Heng L, Zhiying Z (2019) Predicting the receivers of football passes. In: Van DJ, Haaren J, Zimmermann A, Brefeld U (eds) *Machine learning and data mining for sports analytics*. Springer, New York City, pp 167–177
- Hoshen Yedid (2017) Vain: attentional multi-agent predictive modeling. In *Advances in Neural Information Processing Systems*
- Hubáček O, Sourek G, Zelezný F (2018) Deep learning from spatial relations for soccer pass prediction. In: Van Davis J, Haaren J, Brefeld Zimmermann AU (eds) *Machine learning and data mining for sports analytics*. Springer, New York City, pp 162–169
- Hughes M, Franks I (2005) Analysis of passing sequences, shots and goals in soccer. *J Sports Sci* 23:509–14
- Kipf Thomas, Fetaya Ethan, Wang Kuan-Chieh, Welling Max, Zemel Richard (2018) Neural relational inference for interacting systems. In *Proceedings of the international conference on machine learning*
- Le HM, Carr P, Yue Y, Lucey P (2017a) Data-driven ghosting using deep imitation learning. In *Proceedings of the MIT Sloan sports analytics conference*
- Le HM, Yue Y, Carr P (2017b) Coordinated multi-agent imitation learning. In *Proceedings of the international conference on machine learning*
- Link D, Lang S, Seidenschwarz P (2016) Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLoS One* 11(12):e0168768
- Linke D, Link D, Lames M (2020) Football-specific validity of tracab’s optical video tracking systems. *PLoS One* 15(3):e0230179
- Lotte B, Van HJ, Van de Michel V (2019) Measuring soccer players’ contributions to chance creation by valuing their passes. *J Quant Anal Sports* 15(2):97–116
- Luca P, Paolo C, Alessio R, Emanuele M, Paolo F, Dino P, Fosca G (2019) A public data set of spatio-temporal match events in soccer competitions. *Sci Data* 6(1):236
- Peralta Alguacil F, Fernández J, Arce Pablo P, Sumpter D (2020) Seeing in to the future: using self-propelled particle models to aid player decision-making in soccer. In *Proceedings of the MIT sloan sports analytics conference*
- Power Paul, Ruiz Hector, Wei Xinyu, Lucey Patrick (2017) Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*, pp1605–1613

- Rob M, Christopher C (2012) Performance analysis in football: a critical review and implications for future research. *J Sports Sci* 31:639–676
- Rudolph Yannik, Brefeld Ulf, Dick Uwe (2020) Graph conditional variational models: too complex for multiagent trajectories? I Cannot Believe It's Not Better, In *NeurIPS Workshop*
- Sanchez-Gonzalez I, Heess N, Springenberg JT, Merel J, Riedmiller M, Hadsell R, Battaglia P (2018) Graph networks as learnable physics engines for inference and control. In *Proceedings of the international conference on machine learning*
- Silvan S (2018) Passing decisions in football: introducing an empirical approach to estimating the effects of perceptual information and associative knowledge. *Front Psychol* 9:361
- Spearman William (2018) Beyond expected goals. In *Proceedings of the MIT Sloan sports analytics conference*, pp1–17
- Spearman W, Basye A, Dick G, Hotovy R, Pop P (2017) Physics-based modeling of pass probabilities in soccer. In *Proceedings of the MIT Sloan sports analytics conference*
- Taki T, Hasegawa J, Fukumura T (1996) Development of motion analysis system for quantitative evaluation of teamwork in soccer games. In *Proceedings of the IEEE international conference on image processing* 3, pp 815–818
- Vercruyssen V, Raedt LD, Davis J (2016) Qualitative spatial reasoning for soccer pass prediction. In *Proceedings of the ECML PKDD workshop on machine learning and data mining for sports analytics*
- Yeh Raymond A, Schwing Alexander G, Huang J, Murphy K (2019) Diverse generation for multi-agent sports games. In *Proceedings of the IEEE conference on computer vision and pattern recognition*
- Zhan E, Zheng S, Yue Y, Sha L, Lucey P (2018) Generative multi-agent behavioral cloning. In *Proceedings of the international conference on machine learning*
- Zhan E, Zheng S, Yue Y, Sha L, Lucey P (2019) Generating multi-agent trajectories using programmatic weak supervision. In *Proceedings of the international conference on learning representations*

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.