

# Optimally Disrupting Opponent Build-ups

Paper Track

Maaïke Van Roy, Pieter Robberechts, Jesse Davis

KU Leuven, Dept. of Computer Science; Leuven.AI, B-3000 Leuven, Belgium

first.lastname@kuleuven.be

## Introduction

Build-up play refers to the phase of play in a soccer match when a team has possession of the ball and tries to score while the opponent is in an organized defense. It is one of the most structured phases of play in soccer, often starting with a defender in possession and all other players in front of the ball [2]. This well-defined initial set-up allows teams to develop controlled passing sequences and movements [6] that aim to break the opposition's first line of pressure and progress play into the next phases. These sequences are often well-practiced and geared towards the team's strengths or deliberately executed to exploit the opponent's weaknesses. Therefore, the set of patterns that teams typically use and their relative efficiencies vary [4].

While the attacking team attempts to execute the patterns that have the highest likelihood of yielding a goal-scoring chance, the defending team's goal is to optimally disrupt these patterns in order to diminish their threat. This can be achieved by forcing the opponent to play the ball in an unfamiliar or inefficient way. Therefore, knowing a team's preferred locations and how direct their style of play is can support tactical analyzes and help formulate tactical plans to stop the opponent's attacks in the initial stage.

Our research proposes a novel way to identify a team's most effective build-up patterns and the corresponding defensive positioning strategies to disrupt them. That is, we attempt to answer questions such as:

- Should the defending team press high or sit low to avoid long balls?
- Should the defending team force the left center back to initiate the build-up?
- Is it more optimal to direct play towards the left flank or the right flank?

These are difficult questions to answer. First, identifying an opponent's most effective build-up patterns requires generalizing over the multitude of possible variations in build-up play and the inherent uncertainty of actions succeeding. Second, understanding the effect of different

defensive strategies necessitates counterfactual reasoning about what could have happened if a different strategy had been followed [7].

To address these challenges, we build further on our previously proposed framework to reason about decision-making questions [10,11] and defensive game plans [12] in soccer. This framework involves a two-step process that combines learning a model of a team's typical behavior from historic data and using probabilistic model checking to reason about properties of the learned model. Specifically, we first identify the start of build-up situations, such as those in which a center-back has possession without any offensive pressure on the ball. Next, we model the attacking team's build-up strategies by learning a Markov Decision Process (MDP) on their passing sequences following an initial build-up situation with reaching a favorable situation (i.e., the final third) as the final state. When designing the MDP, we encode rich information about the positions of the defending players in the MDP's states by leveraging the unique added context available in StatsBomb's 360 event stream data. Finally, reasoning techniques such as probabilistic model checking [3] can be applied to the MDP to identify how the defending team should optimally organize themselves to block the most dangerous build-up patterns.

The proposed framework provides actionable insights on the build-up patterns of opponents and can be used to provide invaluable insights to coaches when they are planning their defensive strategies against them. To summarize, this paper makes the following contributions:

1. We propose a novel MDP for modeling the build-up strategies of professional soccer teams.
2. We show how probabilistic model checking can reason about the optimal defensive strategies to disrupt the initial phase of a team's build-up.
3. We apply our framework to the top ten teams in the German Bundesliga and the Spanish LaLiga and illustrate how it can be applied during training when preparing for the next match.

## **A Markov Model for Build-Ups**

Our goal is to identify defensive strategies that can disrupt a team's most effective build-up patterns. Therefore, we require a general model of how a team behaves during their build-up play and how they adapt to different positioning strategies of the defending team.

Markov Decision Processes (MDPs) are a popular formalism to model a team's build-up behavior [8,9,13,10–12]. In a nutshell, these MDPs model whether a player possessing the ball will either shoot, or move the ball from one location on the field to another location. For each team, we learn one such MDP per defensive set-up of the opponent. Each MDP thus captures how a team behaves

during build-up possession sequences against a specific defensive set-up. This is a first key difference with the existing MDP's in soccer that attempt to learn a single universal model of a team's behavior. Second, we introduce an alternative state space for the MDPs that extends the traditional location-based states with information about a center back's passing options during the initiation of the build-up. This enables analyzing how a team reacts when players in certain areas of the field are blocked by opponents.

Determining the defensive set-ups and reachability of teammates is made possible by the new added context in StatsBomb's 360 event stream data. This type of data is extracted from broadcast video and contains snapshots of player positioning for each action. As long as a player is visible in the video, information about his location and relationship to the ball carrier (i.e., teammate or opponent) are included. In addition, the data contains the regular event stream data which describes all on-the-ball actions such as passes, dribbles, and shots.

## 2.1 Defining the Markov model

Typically, an MDP consists of the following five parts: the state space, the action space, the transition function, the policy, and the reward function. In what follows, we define each part.

### State space

The state space of our MDP includes the different states or situations that the team can enter and, possibly, exit during a build-up possession sequence. The state space consists of two types of states: absorbing and transient states. Absorbing states can never be left once entered and therefore denote the end of a possession sequence. Similar to our earlier work [10–12], we define three absorbing states which denote the different ways of ending a possession sequence:  $s_{loss\_possession}$ ,  $s_{goal}$ , and  $s_{no\_goal}$ . This means that the possession sequence can either end by losing possession of the ball after a failed move action, by scoring a goal, or by missing a shot, respectively. Transient states can be exited again once a team has entered the state. In this work, we define two different transient state spaces, which allow us to analyze the defensive set-ups both at a higher and a more fine-grained level. The transient state spaces are defined as follows:

**A location-based state space.** The first transient state space consists solely of states that denote a particular location on the field. The 30 states that make up this transient state space are depicted in Figure 1. We use more fine-grained states in a team's own half as this facilitates the analysis of the initial build-up situation, and more coarse-grained states in the opponent's own half.

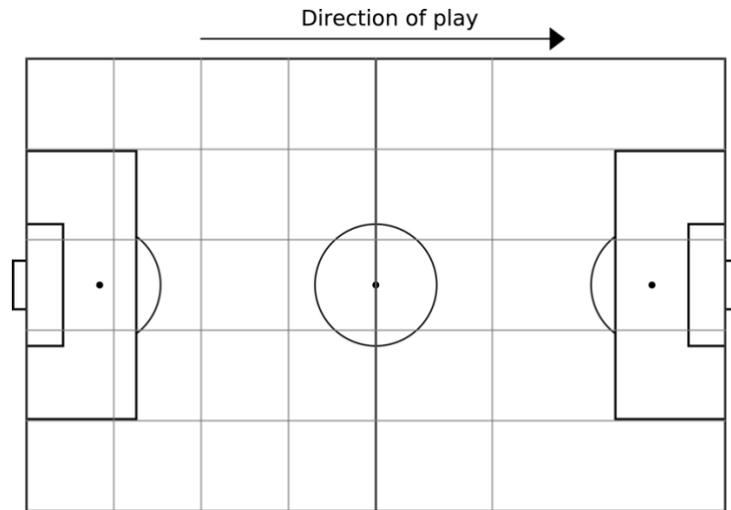


Figure 1: Illustration of the location-based transient states.

When using this transient state space, one such MDP is learned per defensive set-up (cf. Section 2.2.1). This allows us to analyze how a team behaves over the field during their build-up when a certain defensive strategy is in place.

**An extended state space using 360 data.** The second transient state space extends the location-based states with additional information about the locations of other players that is uniquely available for event data in StatsBomb’s 360 data. More specifically, each state is now also defined by four additional binary features indicating whether or not there is a reachable teammate in each of the colored patches in Figure 2. This results in a transient state space containing 480 states. This additional information allows for analyzing how a team reacts when players in certain areas of the field are blocked by opponents.

Whether or not a teammate is reachable in each patch is determined by the proximity of opponents to each teammate as well as the proximity of opponents to passing lanes. Specifically, a teammate is reachable if the following conditions are met:

1. There is no opponent within a radius of 1.5 meters of the teammate.
2. There is no opponent within one meter of the passing lane between the ball carrier and the teammate.

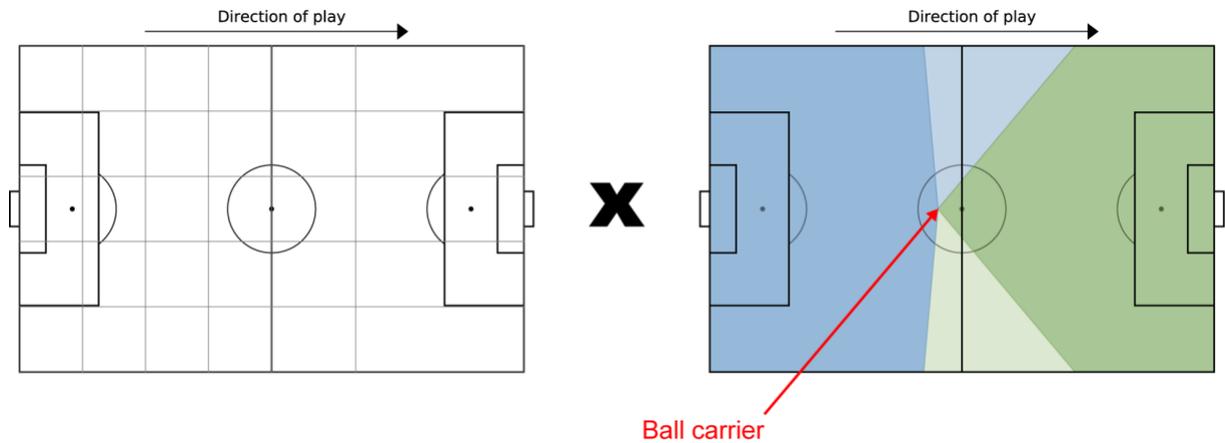


Figure 2: Illustration of the extended transient states. Each location-based state (left) is augmented with four binary features signifying the reachability of a teammate in each of the four colored patches (right). The patches are created relative to the ball carrier and thus differ for each action.

### Action space

In each of the transient states, the MDP can transition to the next state using the following two types of actions: *move\_to(s)* and *shoot*. The *move\_to(s)* action denotes the deliberate ball-moving action to state  $s$ . This action thus encompasses all event stream actions that can move the ball between two locations on the field, i.e., passes, dribbles, and crosses. The *shoot* action denotes shooting on goal. Each action can succeed or fail, and the probability of it doing so is determined by the transition function.

### Transition function

The probability of transitioning from one state to another is defined by the transition function. For an absorbing state, it is not possible to transition to another state, thus only a self-loop with probability one is possible. For a field state  $s$  and a certain time step  $t$  at which action  $a$  takes place, the transition function is defined as:

$$P(s, a, s') = \Pr[S_{t+1} = s' \mid S_t = s, A_t = a]$$

where *Pr* stands for “probability”. As can be seen, the next state  $s'$  only depends on the current state and the action taken (also called the Markov property). This formula can be summarized for each type of action as follows:

- $P(s, \text{move\_to}(s'), s')$  is equal to the probability of successfully moving the ball from state  $s$  to state  $s'$ , and  $s'$  is another field state.

- $P(s, move\_to(s'), s_{loss\_possession})$  is equal to the probability of unsuccessfully moving the ball from state  $s$  to state  $s'$ . The possession sequence ends in the  $s_{loss\_possession}$  absorbing state.
- $P(s, shoot, s_{goal})$  is equal to the probability of scoring a goal from state  $s$ .
- $P(s, shoot, s_{no\_goal})$  is equal to the probability of missing the shot from state  $s$ . The possession sequence ends in the  $s_{no\_goal}$  absorbing state.
- Other combinations of actions and states result in a transition probability of zero.

## Policy

The policy defines the probability distribution over the actions in each state. For an action  $a$  and a state  $s$ , it is defined as:

$$\pi(a | s) = Pr[A = a | S = s]$$

It essentially defines how each team chooses to perform the different actions in each state.

## Reward function

The reward function determines which transitions in the MDP yield a reward and which do not. We mimic the real rewards that can be obtained during a soccer match and only assign a reward of 1 when a goal is scored from one of the field states. All other transitions in the MDP do not yield a reward.

## 2.2 Learning the Markov model from data

Once the MDP is defined, it can be learned from the observed data. Specifically, we use StatsBomb's 360 data to learn separate models for each of the top ten ranked teams in the 2020-2021 season of the German Bundesliga and the Spanish LaLiga. The different steps needed to learn these MDPs are outlined below.

### 2.2.1 Retrieving build-up sequences against defensive set-ups

As a first step, we retrieve all build-up possession sequences from the data. These sequences are defined as starting with a defender in possession of the ball and without being under any pressure from the opposing team. The build-up sequences end when the team loses possession of the ball or takes a shot (regardless of whether it is successful or not). We retrieve approximately 1700 ( $\pm 160$ ) build-up sequences for each team.

Next, the set of build-up sequences of a team is split into four sets based on the defensive set-up they were performed against. One MDP (either with the location-based or extended state space) can then be learned per defensive set-up, allowing a global analysis of the defensive positioning.

When using the extended state space, a more fine-grained analysis of player positioning is also possible.

We consider two pairs of alternative defensive set-ups:

1. a high / low block: A situation in which all visible defensive players are positioned higher / lower up the field than five meters past the halfway line on their opponent's / own half.
2. a left / right-forcing block: A situation in which the defending team's players are higher up the field on the attacking team's right / left side and lower up the field on the attacking team's left / right side. Additionally, the attacking team should move the ball to their own left / right flank in the first two actions.

To illustrate these definitions, Figure 3 shows the starting positions of Real Madrid's build-ups against each type of block. It can be seen that Real Madrid (playing left to right) starts their build-ups lower in their own half against a high block, than when playing against a low block. Furthermore, they start their build-ups more often on their right flank when playing against a left-forcing block and more often on their left flank against a right-forcing block. Splitting the build-up sequences based on these definitions, we obtained about 850 build-up sequences against each type of block for each team in the dataset.

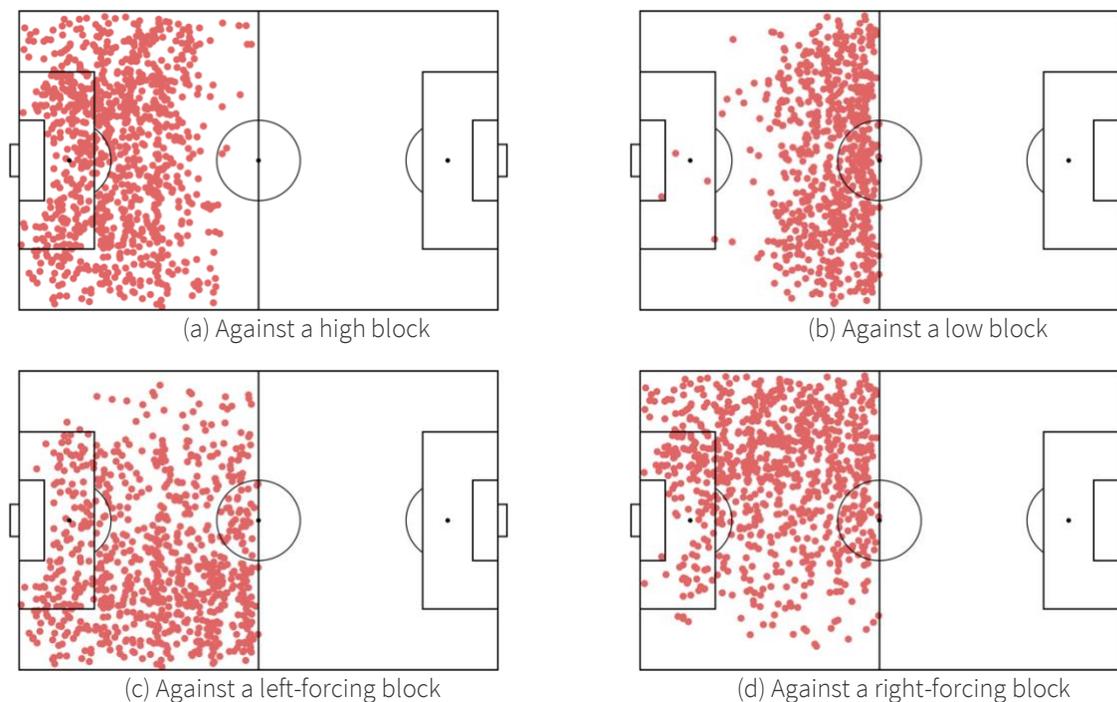


Figure 3: The starting positions of Real Madrid against (a) a high block, (b) a low block, (c) a left-forcing block, and (d) a right-forcing block during the 2020-2021 Spanish LaLiga season.

### 2.2.2 Predicting the intended end state of actions

Once the different sets of build-up sequences have been identified, the ball-moving actions (i.e., pass, dribble, cross, shot) that are part of these sequences can be used to learn the models. As discussed in our previous work [10–12], the intended end location of each action needs to be known in order to learn the transition probabilities and policy from the data. To this end, we employ the same approach as in our previous work, but improve it by using a second post-processing step that takes into account the possible reachable players in each state. This information is not available in regular event stream data, but is now included in the snapshots of the StatsBomb 360 data.

In short, our previous approach predicts for each failed pass, dribble and cross a probability distribution over possible intended end states. For passes and crosses, it does this by learning a gradient boosted trees ensemble on a team’s successful executions of each of these action types. The ensembles are then used to predict a probability distribution over the intended end locations, which is further post-processed using domain knowledge [10]. For dribbles, we assume that the intended end state is simply the state the dribble ended in.

In this work, we add another post-processing step that exploits the reachable teammate information available in the 360 data. As it is more likely that a state is the intended end state when there is a reachable teammate in that state, we boost those states by making them twice as likely to be the intended targets compared to the other states. The distribution is renormalized afterwards. Using this additional post-processing step slightly improves the performance for both state spaces (Table 1).

Table 1: Brier scores [1] for the end location prediction algorithm, with and without the additional post-processing step using reachable teammate information, averaged over all teams. Lower Brier score values indicate better models. For both proposed state spaces, including the post-processing step results in better predictions.

	Without post-processing	With post-processing
MDP with location-based states		
Pass	0.41816	0.41597
Cross	0.13019	0.13014
MDP with extended states		
Pass	0.59445	0.57548
Cross	0.31643	0.29467

### 2.2.3 Learning the transition function and policy

Finally, the transition probabilities and policy can be learned from the extracted build-up sequences and predicted end location distributions.

For the models using the location-based transient state space, all probabilities are computed by counting the occurrences of the specific actions (cf. Section 2.1) in the data. Gaussian smoothing is applied afterwards to ensure the policy and transition probabilities are spatially coherent.

For the models using the extended transient state space, a smoothed prior for the policy (transition probabilities) is first learned over all other teams in the same league using the same approach as for the location-based transient state space. Afterwards, a Bayesian average in combination with this prior is used to compute the policy (transition probabilities) of a specific team. Specifically, for each team and given a certain state  $s$ , the probabilities in the model are calculated as follows:

$$\pi(\text{shoot} | s) = \frac{C_1 m_1 + \#shots\ in\ s}{C_1 + \#actions\ in\ s}$$

$$\pi(\text{move\_to}(s') | s) = \frac{C_2 m_2 + \#attempted\ moves\ from\ s\ to\ s'}{C_2 + \#actions\ in\ s}$$

$$P(s, \text{move\_to}(s'), s') = \frac{C_3 m_3 + \#successful\ moves\ from\ s\ to\ s'}{C_3 + \#attempted\ moves\ from\ s\ to\ s'}$$

$$P(s, \text{shoot}, s_{goal}) = \frac{C_4 m_4 + \#successful\ shots\ in\ s}{C_4 + \#shots\ in\ s}$$

Here,  $m_1$  is the prior probability of shooting in state  $s$ ,  $m_2$  is the prior probability of moving to state  $s'$  from state  $s$ ,  $m_3$  is the prior probability of successfully moving to state  $s'$  from state  $s$ , and  $m_4$  is the prior probability of scoring a goal from state  $s$  (all calculated over all other teams). All  $C_i$  are constants denoting the number of data points of the corresponding value  $m_i$  that are added. In case there are less than 30 actions in a state, all  $C_i$  for that state are set to  $9/m_i$  ensuring a tight 99% confidence interval around the prior. Otherwise, all  $C_i = 0$  and thus no prior information is used. The minimum number of actions needed to use the prior was determined empirically by analyzing the number of visits per state. Using the outlined approach mitigates issues with data sparsity in states where a limited amount of data is available.

## Optimally Disrupting Build-Ups

Using the constructed models, we would like to analyze each team’s behavior during build-up and identify the optimal ways a defending team should position themselves to disrupt the attacking team’s build-up sequences.

To this end, we first analyze a team’s most likely build-up sequences to identify how they generally behave during their build-up. This can already give coaches and managers a general understanding of their next opponent’s build-up play. Next, the main part of our analysis focuses on analyzing the *build-up efficiency* of the attacking team when they play against different defensive set-ups (e.g., a high vs. a low block). We define a team’s *build-up efficiency* as their chances of reaching the final third of the pitch during a build-up sequence and use probabilistic verification<sup>1</sup> to compute this probability. Reaching the final third of the pitch can be seen as having reached a desired and threatening situation from which the next phases of the game can begin. Naturally, a defending team will want to decrease their opponent’s build-up efficiency. Thus, by comparing their opponent’s build-up efficiencies when playing against different set-ups, we can identify which defensive set-ups are most effective.

In what follows, we will analyze a team’s build-up efficiency under three alternating circumstances:

- *When starting the build-up from their left or right flank.* This allows us to identify if a team is less effective when they decide to start their build-up from a particular side of the pitch.
- *When playing against a left-forcing or a right-forcing block.* This allows us to identify the side to which the defending team should force their opponent to play the ball in order to diminish their efficiency.
- *When playing against a defending team that actively blocks players to prevent the first pass going to the left or right side.* This illustrates the added value of the teammate reachability information and allows us to analyze which players should be blocked when the center back is in possession.

### 3.1 Most likely patterns

As a first analysis, we extract the most likely build-up sequences according to the models for each team. Visualizing these sequences can provide teams with a general idea of how each opponent

---

<sup>1</sup> Probabilistic verification allows evaluating user-defined properties (e.g., reaching the final third, performing a shot, scoring a goal) against the MDP and computing the probability that the property holds in the MDP. We use PRISM v4.7: <https://www.prismmodelchecker.org/>.

behaves during their build-up and therefore help teams to formulate plans to disrupt their opponent's build-ups.

Figure 4 shows Barcelona's, Real Madrid's and Bayern Munich's ten most likely build-up sequences in 2020-2021. The most likely sequence is shown in red. It is clearly visible that Barcelona tends to focus on central combinations using Busquets as the main base upon which their possession game is built. In contrast, Real Madrid tends to use the flanks immediately. Their buildup often takes a left-sided focus with Kroos' dropping in the left halfspace and with the creative Marcelo also operating on this side. In the German Bundesliga, we see that Bayern Munich has the same preference for the left flank. Against a high block, Alphonso Davies usually pushes high on the left while Benjamin Pavard at right-back usually acts less aggressively.

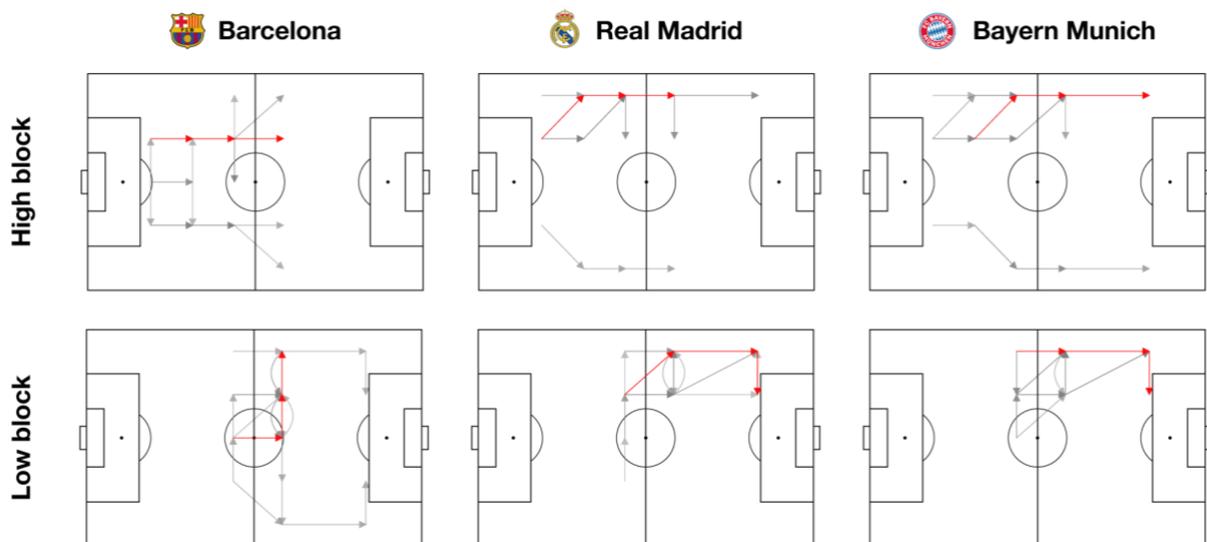


Figure 4: The most likely build-up sequences of Barcelona, Real Madrid, and Bayern Munich when playing against a high and a low block.

### 3.2 Starting the build-up from the left or right flank

Next, we analyze a team's build-up efficiency when playing against different defensive set-ups to identify the set-up with the biggest impact. As a first analysis of build-up efficiency, we investigate the attacking team's left/right symmetry. That is, we investigate the efficiency of the attacking team when starting their build-ups from their left vs. right flank. The resulting information can then be used by the defending team to identify from which side the attacking team is less efficient and thus where they should position themselves before the build-up starts. We will evaluate this efficiency separately for when the team plays against a high and a low block.

Figure 5 shows the results for the top 10 teams in the 2020-2021 Spanish LaLiga and German Bundesliga. Regardless of the block structure, it is best to let Real Madrid start their build-ups from their right side since these are about 2.6% less efficient against a high block and 3.8% less efficient against a low block compared to their left-side build-ups. This corresponds to the previously identified most likely build-up sequences, which were mostly down their left flank. For some teams like Granada, we see that the optimal side to disrupt switches when defending in a high and a low block. However, in general, needing to switch sides is not common. If we require a significant decrease of more than 0.5% in build-up efficiency when disrupting opposite sides against a high vs. a low block, we find that switching is needed only for Granada, Real Betis, and Sevilla. Finally, teams like Atlético Madrid and Bayern Munich are equally efficient at building up from both sides.

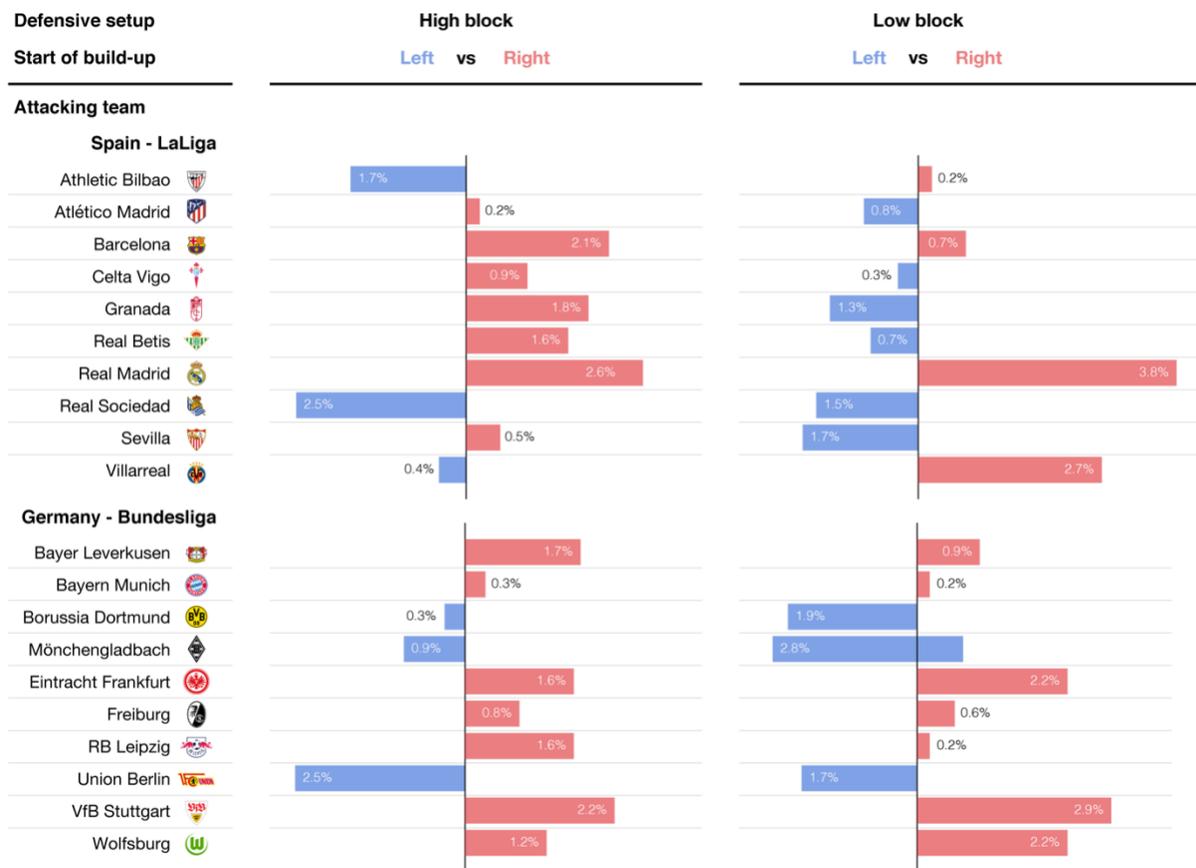


Figure 5: Percentage decrease in the build-up efficiency of a team when starting their build-up from their left vs. right flank, both against a high and a low block, for each team in the top 10 of the 2020-2021 Spanish LaLiga and German Bundesliga.

### 3.3 Forcing a team to play to their left or right flank

Besides letting a team start their build-ups on one side or the other, it might also be interesting to know what happens if the defending team starts to actively force the opponent to play the ball to a certain side. In this second analysis, we will therefore investigate the effect of the previously defined left- and right-forcing blocks.

Figure 6 shows the results for the top 10 teams in the 2020-2021 Spanish LaLiga and German Bundesliga. We see that it is best to force most of the teams in the Spanish LaLiga to their right side, whereas there is about a 50/50 split in the German Bundesliga. For example, for Atlético Madrid it is advantageous to force them to their left side, away from the creative Kieran Trippier who finished the season with six assists from the right-back position. For Barcelona, there is no real difference in forcing them to either side.

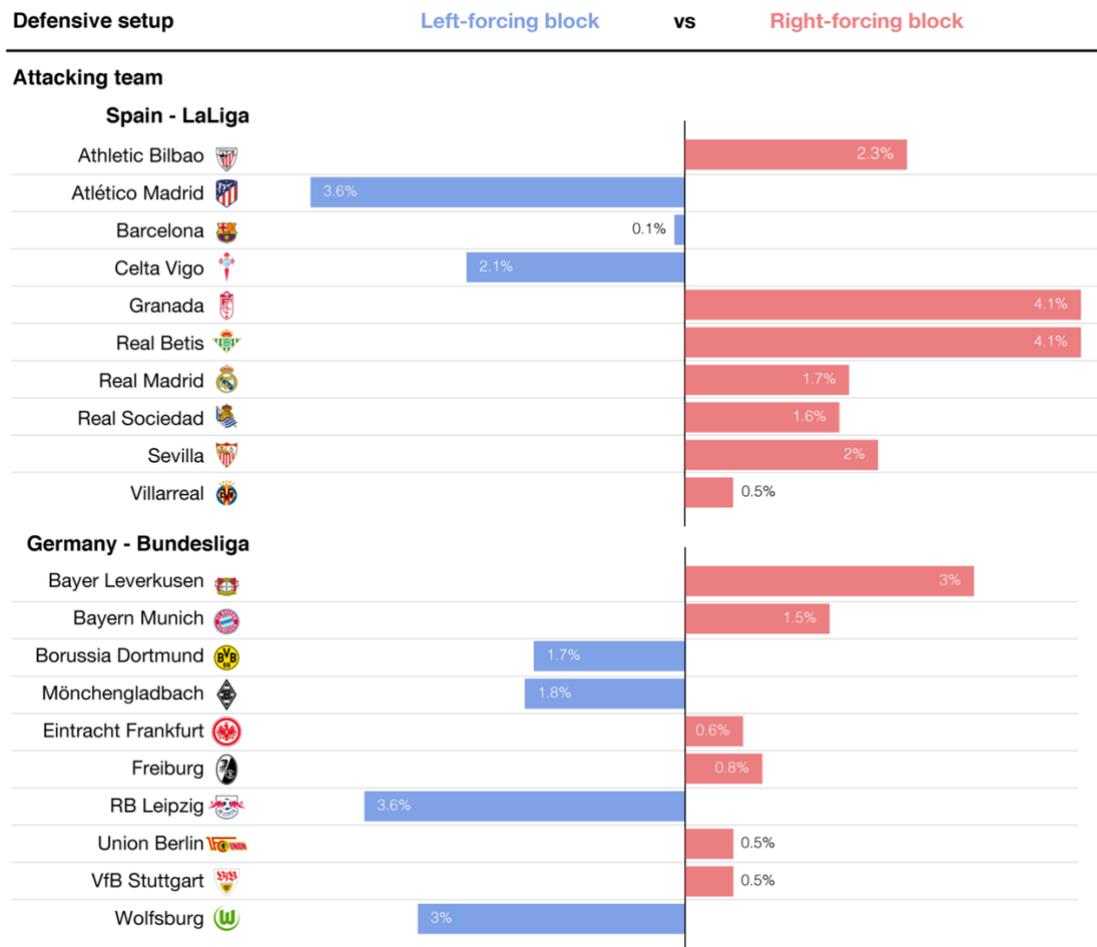


Figure 6: Percentage decrease in the build-up efficiency of a team when forced to play the ball towards a certain side for each team in the top 10 of the 2020-2021 Spanish LaLiga and German Bundesliga.

### 3.4 Blocking players

In this final analysis, we use the extended state space to evaluate whether it is best to block all the players on a center back’s left or right side during the first action of the build-up (Figure 7). Effectively, we measure the change in build-up efficiency when blocking passes to the left/right-back or the central midfielder dropping into the left/right halfspace.

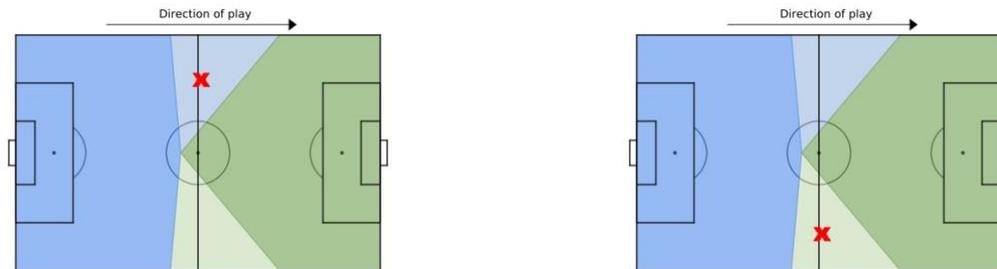


Figure 7: Illustration of blocking all players on a certain side of the center back. The red cross indicates the patch in which no reachable teammate is allowed (i.e., all teammates are blocked by the defending team). The left (right) figure illustrates that we do not allow a reachable player on the left (right) side of the center back.

Figure 8 shows the results of this analysis for the top 3 teams in both the 2020-2021 Spanish LaLiga and German Bundesliga. When employing a high block, it is generally best to block the players on the CB’s right side for all six teams, with Atlético incurring a clear decrease in their efficiency. On the other hand, when using a low block, we see that for Dortmund and Atlético it is best to block the players on the CB’s left side.

Defensive setup	High block		Low block	
	Block left	vs Block right	Block left	vs Block right
<b>Attacking team</b>				
Bayern Munich		0.6%		0.9%
Borussia Dortmund		0.4%	1.1%	
RB Leipzig		0.8%		1%
Atlético Madrid		4.1%	1.5%	
Barcelona		1.3%		1.1%
Real Madrid		0.9%		0.4%

Figure 8: Percentage decrease in the build-up efficiency of a team when all players are blocked on the center back’s left or right side, both against a high and low block, for each team in the top 3 of the 2020-2021 Spanish LaLiga and German Bundesliga.

## Related Work

Markov models have been used extensively to model ball possession sequences in soccer. Their most prominent use is to objectively quantify a player's contributions during a match. This idea was introduced by Rudd [8] in 2011, and later refined by Yam [13] and Singh [9]. Conceptually, our MDP is similar to these existing models, but is tailored towards analyzing the build-up from the back by encoding information about defensive set-ups and reachable teammates in the transient states. Another key difference with the aforementioned approaches is the inclusion of the intended end location of actions in our formalism. This extension was proposed by Van Roy et al. [10] and refined in the current work using the StatsBomb 360 data.

We have proposed the framework based upon probabilistic verification of an MDP in earlier work to identify the best strategy regarding long-distance shooting [11], rating a player's passing riskiness [10] and inform a defensive game plan [12]. The current work further establishes this framework as a general approach to analyze a team's playing style and for providing tactical advice. Furthermore, this work differs from the above-mentioned use cases by the inclusion of contextual information apart from the ball's location in the game state representation.

Both Gurpinar-Morgan [2] and Michalczyk [4] have used a clustering approach to identify a team's most frequent initiating passes when playing out from the back. Michalczyk also identifies a team's most frequent pass chains that reach the final third for each cluster of initiating passes. However, the sparsity of data did not allow them to evaluate the efficiency of these patterns. Moore [5] addressed this by ignoring the intermediate locations of pass chains that start at a center-back and end in the final third. In contrast, a key advantage of our model-based approach is the ability to analyze the efficiency of specific build-up patterns.

## Conclusions

This paper proposed a model for identifying the different build-up strategies that are used by teams and incorporated how each team reacts to a specific defensive set-up in the model. In our analysis, we showed how the model can be used to reason about the optimal defensive strategies to disrupt a specific team's build-ups. The insights that our framework provides can immediately be applied during training when preparing for the next match.

## Acknowledgments

This work was supported by the Research Foundation – Flanders under EOS No. 30992574 and the KU Leuven Research Fund (C14/17/070, C14/18/062). We thank StatsBomb for providing the data used in this research.

## References

- [1] Glenn W Brier. 1950. Verification of Forecasts Expressed in Terms of Probability. *Monthly weather review* 78, 1 (1950), 1–3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2)
- [2] Will Gurpinar-Morgan. 2019. Passing Out at the Back. *Stats Perform*. <https://www.statsperform.com/resource/passing-out-at-the-back/>
- [3] Marta Kwiatkowska, Gethin Norman, and David Parker. 2011. PRISM 4.0: Verification of Probabilistic Real-Time Systems. In *Proceedings of the 23rd International Conference on Computer Aided Verification (CAV'11)* (LNCS), Springer, Berlin, Heidelberg, 585–591. [https://doi.org/10.1007/978-3-642-22110-1\\_47](https://doi.org/10.1007/978-3-642-22110-1_47)
- [4] Kuba Michalczyk. 2019. A Glance at Building Out From the Back. In *Opta Pro Forum*. London. <https://kubamichalczyk.github.io/2019/04/05/A-glance-at-building-out-from-the-back.html>
- [5] Jamon Moore. 2018. Pass Chain Analysis: Should Teams Build From The Back? *American Soccer Analysis*. <https://www.americansocceranalysis.com/home/2018/7/3/gdomsoqkfxu27tatucddmu2a55u2pw>
- [6] John Muller. 2020. The Big Book of Buildups, Day 1: Rotation. *space space space*. <https://spacespacespaceletter.com/the-big-book-of-buildups-day-1-rotations/>
- [7] Judea Pearl and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect* (1st edition ed.). Basic Books, New York.
- [8] Sarah Rudd. 2011. A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains. In *New England Symposium on Statistics in Sports (NESSIS)*.
- [9] Karun Singh. 2019. Introducing Expected Threat (xT). <https://karun.in/blog/expected-threat.html>
- [10] Maaïke Van Roy, Pieter Robberechts, Wen-Chi Yang, Luc De Raedt, and Jesse Davis. 2021. Learning a Markov Model for Evaluating Soccer Decision Making. In *Reinforcement Learning for Real Life (RL4RealLife) Workshop at ICML 2021*, Vienna, Austria (virtual).
- [11] Maaïke Van Roy, Pieter Robberechts, Wen-Chi Yang, Luc De Raedt, and Jesse Davis. 2021. Leaving Goals on the Pitch: Evaluating Decision Making in Soccer. In *Proceedings of the 13th MIT Sloan sports analytics conference*, Boston, USA (virtual). <http://arxiv.org/abs/2104.03252>
- [12] Maaïke Van Roy, Wen-Chi Yang, Luc De Raedt, and Jesse Davis. 2021. Analyzing Learned Markov Decision Processes Using Model Checking for Providing Tactical Advice in Professional Soccer. In *Proceedings of the AI for sports analytics (AISA) workshop at IJCAI 2021* (AISA), Montreal, Canada (virtual).
- [13] Derrick Yam. 2019. Attacking Contributions: Markov Models for Football. *StatsBomb*. <https://statsbomb.com/2019/02/attacking-contributions-markov-models-for-football/>