Strategy Optimization in Sports via Markov Decision Problems

Susanne Hoffmeister and Jörg Rambau

Abstract In this paper, we investigate a sport strategic question related to the Olympics final in beach volleyball: Should the German team play most aggressively to win many direct points or should it rather play safer to avoid unforced errors? This paper introduces the foundations of our new two-scale approach for the optimization of sport strategic decisions. We present possible answers to the benchmark question above based on a direct approach and the presented two-scale method. A comparison shows the benefits of the new paradigm.

1 Introduction

Picture yourself in a beach volleyball match. The score is 14:13 for you in the third set. One more point and you win the Olympic gold medal. Your team's repertoire contains a very risky and a slightly safer style. The risky style gives you more chances to directly win the point, at the expense of a larger probability of a direct failure. For the safer style, it is more likely that the rally will continue. How should you play? This will be the *benchmark question* for the concepts in this paper. Does the best style depend on the score? Does it depend on who serves first?

We approach this type of principle strategic decision problems in sports games with mathematical models based on *Markov Decision Problems (MDPs)* [18]. Whereas *Markov chains* have been used extensively, mainly to identify which skills are pivotal in sports games [13], MDPs, and even more so, *Markov games*, are much less prominent.

An MDP consists of several (temporal) stages. In all stages there are states, actions, rewards, and transition probabilities. In each stage, the decision maker can

Susanne Hoffmeister
University of Bayreuth e-mail: Susanne.Hoffmeister@uni-bayreuth.de

Jörg Rambau
University of Bayreuth e-mail: Joerg.Rambau@uni-bayreuth.de

specify an action depending on the state. Then the system will move to another state in the next stage according to the transition probabilities, which depend on state and action. Depending on the current state, the current action, and the resulting state, a reward is granted. The goal is to find a *policy*, such that the expected (discounted) total reward over all stages (finitely or infinitely many) is maximized.

Sometimes the decision maker wants to maximize the probability that the system reaches some desired state (in our case a winning state of the game). This can be modelled by making the desirable state an absorbing state (i.e., a state in which the system remains forever) and granting a reward of one for entering this state (rewards are zero everywhere else). Thus, this way we can model how to maximize the probability to win a sports game.

The problem with the applicability of MDP models in sports is the following. The number of different situations in sports games is, strictly speaking, inaccessibly large, infinite, or even uncountable. Since it is not explicitly defined what the "true" state space of a sports game is, one needs to seek for an approximation. We deliberately chose to look for a representation of the really substantially different situations by a finite number of states and actions. The finite-state-action approach comes with the question of how detailed such a finite MDP should be. For a model with continuous state space components, the analogous question arises for the number of state-space dimensions.

If one uses too many states and actions, one may end up with a too complicated MDP: good policies may be too detailed to be easily adopted by the players during the gameplay, and (near) optimal policies may even be very difficult to determine. Moreover, the more complicated a model is, the more modeling parameters usually need to be estimated, which may become a problem, if the solution is sensitive to parameter deviations.

If one uses only few states and actions, then the estimation of transition probabilities becomes difficult. Just think of the simplest possible MDP with only the three states "start", "win", "loss". Then the whole problem is reduced to the estimation of the transition probabilities. Thus, in order for a coarse MDP to be realistic, a lot of information has to be contained in the transition probabilities. Therefore, these probabilities inevitably depend on a combination of various aspects (e.g., the combination of players on the field). If for some such combination we have no or not enough historical data, then a straight-forward estimate is not available.

One idea in statistics is to model using a formula with (few) parameters how the data is generated from other (observable) data. Then one can estimate the (few) parameters and use the evaluated formula in unknown terrain. We use the underlying principle of this idea. However, we will not employ a parametrized formula but we use two *Markov Decision Problems (MDPs)* to answer the benchmark question. Our *strategic* MDP models the influence of principle strategic options. Optimal policies for it answer our question. Because these policies are principle strategic decisions, players can follow the corresponding recommendations. However, the strategic MDPs transition probabilities are difficult to estimate. Thus, a *gameplay* MDP models the gameplay in detail. Its transition probabilities are easier to estimate. By simulating the gameplay MDP, we estimate the transition probabilities of the strate-

gic MDP. We carry out this program for our example benchmark question. It will turn out that in our coarsest MDP model the optimal decision among *risky* play and *safe* play in a service or a field attack situation depends neither on the score nor on the right to serve. We will see that the optimal decision can be found by evaluating a certain expression in the transition probabilities. If this expression is positive, then playing *risky* throughout is optimal, otherwise playing *safe* throughout is better.

To arrive at our conclusions, we have developed an all new machinery based on Markov decision problems (MDP). Our method differs from existing MDP approaches in the manner that we combine *two* related MDPs to answer a *single* strategic question. A detailed MDP has transition probabilities that can be estimated, a coarse MDP provides conclusions that can be used in practice. By relating the two, we can provide useful conclusions on the basis of data that can be estimated.

We implemented our approach for the example beach volleyball. For the first time, we develop a strategic MDP that models whether safe or risky play yields a higher winning probability in a particular match ("s-MDP"); we analytically characterize an optimal policy for it in terms of the (very few) s-MDP transition probabilities ("optimize"); we develop a gameplay MDP that models a rally in detail depending on the players' individual skills ("g-MDP"); we analyse historical competition videos from the London 2012 Olympics with an all new video analysis tool in order to estimate the single-player-dependent transition probabilities for the g-MDP ("calibrate"); we simulate the g-MDP for each policy of the s-MDP, which provides estimations of the transition probabilities for the s-MDP ("simulate"); we derive whether safe or risky is better against a given opponent by evaluating the optimality criterion for the transition probabilities ("conclude"); finally, we visualize the sensitivity of the strategic MDP's recommendation depending on skills and opponent strength in a strategy-skill score card ("present").

For general sports games, the meta-procedure "s-MDP" – "optimize" – "g-MDP" – "calibrate" – "simulate" – "conclude" constitutes the first two-scale approach to answer principle strategic questions. The resulting skill-strategy score cards can support the choice of a strategy in an upcoming match against a particular opponent. This meta-procedure defines a research program for each principal strategic question in an individual sports game.

This paper is organized as follows. In Section 2, we briefly review the related work in both MDP and sports strategy research. Section 3 describes our new approach of combining two MDPs of different scale. The two MDPs are defined in Sections 4 and 6 for our example beach volleyball. Section 5 between shows some computational results based on the s-MDP alone based on direct counts from the 2012 Olympics. In a first round of computational results, we show that a direct estimation of s-transition probabilities yields volatile results with systematic shortcomings that motivate our two-scale procedure. Section 7 specifies the parameters that define a g-MDP strategy and the implementation of the two special strategies *risky* and *safe*. Our data collection from the 2012 Olympics is based on a new video analysis tool that appears in Section 8. We present computational results in Section 9, followed by a comparison in Section 10 and a sensitivity analysis based on

skill-strategy score cards in Section 11. Our suggestion how to extend the method to game theory can be found in Section 12. We conclude the paper in Section 13.

2 Sport Strategy Optimization and MDPs

If the focus of the analysis of a sports game is on interactions and strategic aspects of that game, then a dynamic model may be more appropriate than a statistical approach. Markov chains (MCs), Markov decision problems (MDPs) and Markov games (MGs) are possible frameworks for a dynamic model. They all incorporate the Markov property: the next state may just depend on the current state and not on the complete realization history of states.

An MC is a discrete-time stochastic process that can be characterized by a set of states and a probability distribution P which specifies *transition probabilities* between *states*. With an MC, it is possible to examine how a system will evolve under P given a certain initial distribution [17].

There are several contributions that use MCs to investigate different aspects in various types of sports. We only present a sample of recent applications of MCs that consider beach volleyball or volleyball, as this is the type of sport we will later use as an example. Miskin et al. [13] investigate skill importance in women's volleyball. The authors model play sequences as discrete absorbing MCs by using a Bayesian approach to estimate the transition probabilities from the data gathered. The data was collected during the 2006 competitive season of a single women's Division I volleyball team. The 36 states consolidated in this analysis are moves that consist of a skill and a rating combination, e.g., a set is rated according to its distance from the net. The importance score of a skill is a metric that incorporates its impact to the desired outcome and its uncertainty. It is computed by the posterior distribution associated with the skill. Ferrante and Fonseca [6] use an MC approach for volleyball to compute an explicit formula of the serving team's winning probability in a set. Beside this, the mean duration of a set is computed in terms of the expected number of rallies. The authors make the assumption that the probability of winning a single rally is independent of the other rallies and constant during the game. In this way they are able to apply an MC. The states in their model correspond to different scores that may occur in a set together with an indicator which team serves next. The winning probability is computed in terms of two parameters which represent the winning probability of a rally depending on the serving team. MC properties and combinatorial arguments are used to derive the explicit formula for the winning probability. The authors applied their formula to data from the Italian Volleyball League. The calculated winning probabilities and set durations were close to real data estimates. A similar MC approach as in [13] has been used by Heiner et al. [7] for women's soccer.

An MDP is more complex than an MC. It is a *Markov decision process* supplemented by an *optimality criterion*. The decision process incorporates a decision maker that chooses at each time step (a *decision epoch*) an *action* from a specified

set of actions. This way, the transition probabilities do not only depend on the current state (like in an MC), they also involve the chosen action of the decision maker. Depending on the current state, the chosen action, and the realized next state, a reward is generated. A *policy* is a decision rule that prescribes an action choice for each state. Once a policy has been fixed, the system generates an MC. Policies can be compared with respect to the optimality criterion. There are tools to determine an *optimal* policy for the decision maker in many settings if the problem scale is not too large. We follow the notation of Puterman's textbook [18] throughout this paper.

In sports related applications, MDPs are often used in connection with general, tactical considerations that are not team or match specific. Some examples are: Clarke and Norman [4] as well as Nadimpali and Hasenbein [15] investigate a Markov Decision Problem (MDP) for tennis games to determine when a player should challenge a line call. The latter one is the more detailed model. We describe it briefly in the following: A decision point occurs when an opportunity to challenge the umpire arises. The states include the outcome of the point, the score, the number of challenges remaining, the probability that the call is incorrect, and the result of a successful challenge. There are two possible actions in each state: *challenge* and *do not challenge*. Further parameters of the model are the relative strength of the players and the fallibility of the officials. These parameters are used to generate the transition probabilities for the model. They use the standard linear programming approach for multi-chain, average cost MDPs to obtain optimal policies under a variety of parameter settings. Hirotsu and Wright model football as a four state Markov Process and use dynamic programming to determine the optimal timing of a substitution [8], the best strategy for changing the configuration of a team [9], or to determine under which circumstances a team may benefit from a professional foul [26]. Chan and Singal [2] use an MDP to compute an optimization-based handicap system for tennis. The weaker player gets 'free points' at the start of the match, such that the match-win probability of both players is equalized. Clarke and Norman [3] formulate an MDP for cricket to determine whether the batsman should take an offered run when maximizing the probability that the better batsman is on strike at the start of the next over. The model is solved analytically by dynamic programming. Norman [16] builds a more aggregated MDP for tennis games to tackle the question when to serve fast or when to serve slow at each stage of a game. The model is solved analytically using a monotonicity property of the optimal cost function and dynamic programming.

Most of the MDPs on team or match dependent sport-strategic decisions are retrospective: Terroba et al. [21] develop an MDP-based framework for tennis matches. The information needed to build the model is semi-automatically gathered from broadcast sports videos. Machine learning algorithms are executed to identify optimal policies. They also present a novel modification to the Monte Carlo tree search algorithm and apply their model to popular tennis matches of the past. They present how the player who has lost in reality could have won the match with identical skills, just by using a different policy.

To the best of our knowledge, the only MDPs that take player skills into account and could be applied to future matches exist for baseball. Wright and Hirotsu [25]

formulate a Markov model for baseball to calculate an optimal pinch hitting strategy under the 'Designated Hitter Rule'. Their method can be applied to a specific match by using the probability of each player to achieve a single, double, triple, home run, walk, or out.

An MG is a stochastic game in an MDP-like environment. Instead of one decision maker, there exists a whole set of players. At each decision epoch, each player chooses an action from his action set. So, the transition probabilities and rewards incorporate the decisions of all players. An MG gets even more complex through the different optimality criteria of the players. Therefore, policies that are simultaneous best responses, i.e., Nash equilibria, are the focus of interest. In some cases, most notably when an MG with finite strategy sets has pure Nash equilibria, they can be found algorithmically. This requires the repeated computation of best responses to fixed strategies. Our method in this paper can be also seen as a building block for this problem. In Section 12 we show first results into this direction. More on MGs can be found in Webb's textbook [24].

To the best of our knowledge, there exist only a few applications of a Markov Game (MG) to optimize the policies a-priori for a particular sports game. Kira et al. [11] formulate an MG for baseball and computed Markov perfect equilibria for both teams. The transition probabilities of the MG are assumed to depend only on the probability parameters for the hitting skills of the players. They use a dynamic-programming algorithm for solving the Bellman equations that characterize the value function of the game for both teams. However, MG models have been applied in the context of sports strategies in a more general set-up: Walker et al. [23] use *Binary Markov Games* to model a sports game like tennis and derive that under certain monotonicity properties optimal policies to win the match are a repeated application of an optimal policy to win a rally (our results on the optimality of myopic policies in Section 4 are related but in the MDP set up). Turocy [22] uses MG models fed with massive historical data in order to clarify whether there really *has been* a "last-up" advantage in baseball *on average in the past*. Routley and Schulte [19] employ MG models to rank ice hockey players according to their skills. In an upgraded MG model also location information is included [20]. Anbarc et al. [1] have tried to decide the fairness of tie break mechanisms on the basis of MG models.

Why did we choose to utilize MDPs instead of MGs in this paper, although our policy might be influencing the policy of our opponent? The answer is twofold: first, in order to investigate MG models, the problem of characterizing best-responses to given policies is important. By investigating MDPs, we cover this step. Second, for strategical decision support, MGs would guide us to the best policy against a *strategically perfect* opponent. In most cases, this is not what we want; we consider it rather more successful to adapt to the special strengths and weaknesses of a particular opponent. In future models, we plan to incorporate dimensions like variability into the MDP setting. This will at least cure some of the short-comings of MDP models in this regard. If the sets of strategy choices are finite and small (like for the benchmark problem in this paper), our approach can be applied to solve finite constant-sum games modeling the behaviours of both opponents (see Section 12).

3 The New Two-Scale MDP Approach

We seek for elementary strategic guidelines for a sports game. The dilemma in MDP-modelling of sports strategy optimization is the following: A compact MDP directly developed for the strategic question ("Should I play safe or risky?") may allow us to analytically or numerically solve for an optimal policy based on the input data. However, this input data, in particular the transition probabilities (e.g., for an attack directly winning a point in a beachvolleyball rally) often depend on the combination of all players involved. Consequently, they are hard to estimate, since historical data for all combinations of players is needed (e.g., for the direct point probability, the setting and hitting skills of the attacking team's players and reception skills of the receiving team's players are all relevant at the same time).

For a more detailed model where transitions only depend on individual actions (e.g., the hard smash aimed at a certain spot in the field happened as intended) transition probabilities can be estimated easier, since only individual success probabilities for a single player are needed. Such detailed MDP models with billions of states could be very complicated to solve for optimal policies, but that is not the main problem. A detailed MDP will inevitably produce recommendations on the level of exact individual actions in all kinds of special situations (e.g., whenever your opponent has taken certain positions and the ball is in another position and flies in a certain direction and your team mate's position is in another certain position ..., then your next hit should be a set to a certain position.) Since strategy recommendations have to be implemented by humans eventually, such outcomes would be impractical.

However, the idea to use coarser, more principle MDP models can lead to a very difficult input-data problem: The details that we may want to leave out in the model are not irrelevant. They appear in aggregated form in the transition probabilities, which often depend on the opponent's behaviour. The consequence is that these probabilities can hardly be estimated whenever up-to-date observations for our team and the actual opponent are not available. In contrast, in a detailed model the transition probabilities may refer to very simple state transitions. This could be the probability that a certain hit is performed successfully, almost successfully, or failed completely. Such probabilities can be observed in special training sessions or in videos of historic events, independently of the skills of an opponent.

It seems that each modelling granularity has something to offer. Therefore, we use both, i.e., we employ two MDPs instead of one for the optimization of a sports strategy and relate them to each other. One is coarse and one is detailed. We obtain our new *two-scale method*.

The coarse MDP is called the *strategic MDP*, *s-MDP* for short. It represents the principle influence of the strategic decision in question on the winning probability. It uses as its basis a plausible segmentation of the gameplay into strategic pieces. This could be a phase of ball-possession or the like. This s-MDP will have moderate size and a simple structure so that finding an optimal policy is within reach analytically or numerically. Moreover, players can implement the resulting recommendation in practice. However, its transition probabilities need not be easily observable. The

detailed MDP is called the *gameplay MDP*, *g-MDP* for short. It is an ordinary MDP but with a very fine granularity. Since the g-MDP is only used in a simulation, it does not matter if it has billions of states. It represents the dynamics of the detailed gameplay in greater detail in such a way that its gameplay-decisions *(g-decisions)* and gameplay-state transitions *(g-transitions)* can be related to strategy decisions *(s-decisions)* and strategy-state transitions *(s-transitions)* in a meaningful way.

Neither the size nor the structure of the g-MDP are restricted. What we care about is that all transition probabilities in the g-MDP can be observed up to an acceptable accuracy, possibly by some additional effort like special training sessions or video analysis

Whether or not we are scoring a direct point depends on us *and* the opponent. So, observations must be classified according to pairs of opponent teams, and there are many. This leads to a very sparse data basis to estimate probabilities. One way out is to develop a model how the probabilities come about. Our idea means: instead of standard parametric statistics, we use a g-MDP to generate the transition probabilities. A suitable g-MDP has single moves and hits as actions. The g-states store the players' and the ball's positions plus some technical information like how many times in a row the ball has been touched by the same team, who touched the ball before and whether the ball was hit hard before.

In order to gain an advantage over using the s-MDP alone, we allow state transitions whose *g-transition probabilities only depend on the skills of the player hitting the ball*. For example, the action "smash targeted at a certain position in the opponent's field" leads to follow-up states only depending on to what extent the smash was carried out successfully. Whether or not the opponent can return the smash in a controlled fashion solely depends on the returning player's skills. The resulting g-MDP will be complicated. It will possibly be hard to find optimal g-policies. And: a g-policy will be complicated to implement during gameplay. But: By Monte-Carlo simulation *(g-simulation)* of the g-MDP we can in certain cases estimate the resulting s-transition probabilities.

More specifically, we have to relate the g-MDP to the s-MDP as follows: For each s-policy we have to specify what g-actions fit to this policy in the g-MDP. Call a feasible sequence of g-decision rules over the epochs of a phase of ball possession an *attack plan*. Any set of such attack plans is called an *attack type*. A probability distribution over an attack type is called an *attack style*. Now, we assign to each s-policy an attack style.[1] We call this assignment the *s-g-implementation*. For example, if at a certain score our team is in possession of the ball and wants to play the s-policy *risky*, then we can assign to it a probability distribution over the set of all attack plans ending with the most risky (i.e., close to the border of the field or a hard hit like a smash) attack-hit *available in the respective situation*.[2] The set-up of such

[1] In other words, an attack style is a mixed partial g-policy consisting only of decision rules that belong to some attack type.

[2] Note that it would not be sufficient to assign a probability distribution over a set of *actions* to an s-policy, since any hit in a sequence of actions could fail with some probability, resulting in a state without feasible actions. In contrast to this, an attack plan, which uniquely determines a sequence of decision rules, returns for all possible resulting failure states an action to cope with it.

an s-g-implementation requires the classification of all g-decision functions in each g-state by s-decisions. Thus, a vast amount of case-by-case analysis is necessary using expert knowledge of the particular game. Usually, more than one g-decision function is possible to represent a single s-decision; in that case, we choose one uniformly at random. A different viewpoint is that the s-g-implementation is the formal definition of what a coach actually means by playing *safe* or *risky*.

Using this connection, one can count *in simulation* how often one realization of the risky attack style results in a direct point, a direct failure, or a continuation of the rally, which correspond to outcomes of s-transitions in our example. Essentially, the g-MDP is utilized in the same way as a family of Markov-chains, parametrized by classes of g-decisions, each induced by the possible s-decision it implements. An MDP model is the better viewpoint here, since a formal connection between the s-MDP and the g-MDP needs the concept of g-decisions.

Since the transition probabilities of the s-MDP usually depend on the combination of all players involved, it is difficult to estimate them by historical observations. At worst, a certain combination of players may have never played in a match before. Here, our two-scale approach comes into play. The related g-MDP models each player's individual actions (in basketball, e.g., this might mean dribble, pass, shoot, from where, to where, ...). The *individual* player probabilities, called *skills*, describe the outcomes of these player's actions and constitute the g-transitions. The advantage of the g-MDP is that the g-transitions only depend on a single player's skills and such skills are easier to estimate. They do not depend on combination of players and can therefore be estimated from arbitrary matches of that single player or even from training experiments. Usually, several g-transitions in the g-MDP (i.e., again for basketball: pass around – no-look-pass into the attack area – shot – score) constitute one s-transition (i.e., we score) in the s-MDP. The s-transition probabilities of the s-MDP can then be estimated by counting the g-transitions in g-MDP-simulations on the basis of estimated skills.

Given the simulated s-transition probabilities, one can solve the s-MDP and find out an optimal s-policy, which represents a principle strategic recommendation. Note that in order to show-case the concept in this paper in a more concise way, we have chosen to restrict our s-MDP for beach-volleyball to only few possible policies. In priciple, which out of two (or few) policies is the best could be estimated by simulating the corresponding Markov chains in the g-MDP for the durations of complete games (rather than single phases of ball-possession) at the cost of longer computation times. Even in our simple case such a brute-force numerical approach would miss out some important information: The analysis of the s-MDP provides us with structural results (Theorems 1 and 2 and the winning probabilities of the tie-game in Section 4) and with useful sensitivity information (see Section 11). This information follows from the fact how exactly the simulated probabilities influence the qualities of the policies. This information would be substantially more difficult to obtain by simulation alone. Furthermore, in our approach the policies in the s-

For example, in a failed attempt to set the ball properly, the most risky smash available might be much less aggressive than the safest smash available after an excellent set – this possibility could not be covered by classifying actions.

MDP can be chosen to be much more involved as long as the s-MDP can be solved fast analytically or numerically. The evaluation of the g-MDP can even be performed on-demand inside a possible interative solution method for the s-MDP. Beyond this, our setting allows for the following extension: we can make the s-g-implementation the subject of optimization for each s-strategy. Implementing this extension, however, would be beyond the scope of this paper.

In the gameplay situation, the players have to decide about the explicit g-actions they perform next, depending on the optimal s-policy and the states they encounter. They will do this exactly by mimicking the s-g-implementation. Whenever in reality all attack plans in an attack type are carried out similarly often as in the g-simulation, the actual s-transition probabilities will be similar to the simulated s-transition probabilities.

The exact form of the s-g-implementation is defined through expert knowledge. It can be implicitly based on intuitive understanding of the players. This has the advantage that no brain power is needed for it during gameplay. Alternatively, the team may want to establish an explicit encoding what an s-policy (e.g., play *risky*) is supposed to mean in terms of an attack style (play "any attack combination with the hardest-possible smash closest-possible to the boundary of the field" or the like). For beach volleyball, we have implemented this idea for one strategic question (see Sections 4 through 9).

Although in this paper, the only worked example is from beach volleyball (other examples like tennis can be worked out in a similar way), the two-scale MDP paradigm can be used for other sports games as well. For the particular sports game, one first has to develop an s-MDP, which models the strategic question. Second, one needs a sufficiently related g-MDP with observable transition probabilities. The g-MDP serves as a device to estimate the transition probabilities in the s-MDP. For this purpose, s-transitions are counted in simulations of the g-MDP.

Consider, e.g., basketball. One interesting strategic question is whether to provoke a very fast tempo with high risk against a then less consolidated defence or to play calmly with low risk against a completely settled defence configuration. Or soccer: Should one preferably play high long passes behind the defending lines or should one play short low passes. A possible s-MDP would then consist of states corresponding to the principal situations (score, ball possession, phase of attack, shot opportunity) in which our team can choose to play "fast" or "slow" (basketball) or "high" or "low" (soccer) in order to influence the transition probabilities. Given these probabilities we could solve the s-MDP and make a recommendation: "fast" or "slow" and "high" or "low".

In the following sections, we will present an s-MDP/ g-MDP pair for beach volleyball that finds rules when to play *risky* or *safe*, depending on the skills of the individual players and the opponent's skills. Even for beach volleyball, we note that our special choices of an s-MDP and a g-MDP are by no means unique. Our choices were guided by the wish to base the answer to the team-strategic question on the skills of the individual players for common hitting techniques. All rules concerning beach volleyball can be found in official documents by the Fédération Internationale

de Volleyball [5]. We will briefly sketch the most important scoring rules when they become relevant in Sections 4 and 6.

The two-scale MDP paradigm can be transferred to other sports games, but the concrete implementation in this paper, i.e., the s-MDP, the g-MDP, and the data estimation are tied to the beach volleyball example with the benchmark question. A concrete implementation of our new paradigm for interesting strategic questions in basketball and other sports games is research in progress.

4 A Strategic MDP for Beach Volleyball

From now on we show how our two-scale model can be applied to a particular sports game, namely beach volleyball, and a particular strategic question, namely safe versus risky play. At this point, safe and risky play are just names for two strategies in the s-MDP. Eventually, it is only the s-g-implementation that will give this a concrete meaning in terms of classes of detailed g-decision rules.

In this section, we specify an s-MDP for our benchmark question. Recall that we want to find out in which situations (score, possession, serve or field attack) *risky* play will lead to a higher set-winning probability than *safe* play. Our strategy is to construct the s-MDP as simple as possible. The benchmark question requires to model the actions of one of the two teams. Moreover, we distinguish between service play and field attack play – it might be optimal to serve *safe* and to attack *risky* or vice versa.

Let Team P be the team whose strategy we want to optimize, and let Team Q be P's opponent. The control set in all states s where team P possesses the ball is given by $A_s := \{risky,\ safe\}$. The control set in the states where team Q possesses the ball contains a unique dummy control. Aiming at the benchmark question, a state has to contain the current score, which team starts the next attack plan and an indicator whether the state is a serving state or not. Thus, the simplest possible state space with respect to the benchmark question is $S^{\text{reg}} := \big\{(x,y,k,\ell) \mid x,y \in \mathbb{N},\ k \in \{P,Q\},\ \ell \in \{0,1\}\big\}$. Here, x and y denote the scores of Team P and Q, respectively. Moreover, k specifies which team possesses the ball, and ℓ encodes whether or not this is a serving state ($\ell = 1$) or a field attack state ($\ell = 0$).

Let us restrict to matches consisting of a single set to 21 points in the following. The state set *winning states* S^{win} contains all states, where team P has won the set, e.g., states where P has 21 points and Q no more than 19. Similarly, the state set *losing states* S^{lose} contains all states, where team P has lost the set, e.g., where Q has 21 points and P no more than 19. At states with a score of $20:20$, the so called "tie game" starts, where a team wins if it has a lead of 2 points. As our s-MDP should have a finite number of states, we use a different state representation for the tie-game. Instead of remembering the number of points for team P and team Q separately, we only denote the point difference of the two teams in a state. So, the states of the tie game are

$$S^{\text{tie}} = \{(z,k,\ell) \mid z \in \{-2,-1,0,1,2\},\ k \in \{P,Q\},\ \ell \in \{0,1\}\},$$

which are only finitely many states. This kind of state representation is not possible in the regular set, since the absolute number of 21 points must be reached to win a set. In the tie game, only a relative criterion must be fulfilled. Note that with this simpler representation it is not possible to make the s-transition probabilities dependent on the duration of the tie game. Incorporating dependence on the duration requires a more complicated solution procedure, generally using a countably infinite number of states. In this paper, we stick to stationary probabilities, which is not an uncommon assumption in professional sports [12]. Using the relative notation for the tie-game states, there we have finitely many states that describe a beach volleyball set. Let S^{win} and S^{lose} contain all winning and losing states respectively for team P which are modelled as absorbing states that are, once entered, never left.

A decision epoch starts when Team P gains control over the ball and starts its attack. The decision epoch ends when Team P makes a fault or a point, or when the attack is successful but Team Q has gained control over the ball and starts its own attack plan. The actions of Team Q are modelled as part of the transition probabilities in the s-MDP. These decision epochs in general allow for infinitely many stages. Let $p^+(P)_a$ [$p^-(P)_a$] be the probability that Team P playing action a directly wins [loses] the rally. The corresponding probabilities for Team Q are denoted by $p^+(Q)$ and $p^-(Q)$, respectively. As abbreviations, we denote the probabilities that none of this happens by $p^0(P)_a := 1 - p^+(P)_a - p^-(P)_a$ and $p^0(Q) := 1 - p^+(Q) - p^-(Q)$, respectively. Since a serving attack has transition probabilities clearly different from a field attack, we distinguish between them. This is denoted by a superscript *field* or *serve* on the transition probabilities. In the following, we are considering the strategic options *risky* and *safe* either for the service or for the field attack. Thus, the evolution of the system is governed by twelve probabilities $p^+(P)_a^{att}$, $p^-(P)_a^{att}$, $p^+(Q)^{att}$, $p^-(Q)^{att}$, where $a \in \{risky,\ safe\}$, $att \in \{serve, field\}$.

These probabilities induce all transition probabilities by incrementing points and changing the right to serve in the obvious way. Entering a winning state yields a reward of one; all other transitions have reward zero. Table 1 summarizes our s-MDP. Note how the whole construction of our s-MDP was guided only by the benchmark question, not by finding the most compact representation or a being able to solve the model. In Figure 1, we illustrate the resulting transition diagram in the case that P services first in the set for a simplified beach volleyball set requiring only two (instead of 21) points for a win. At the states $(1,1,P,1)$ and $(1,1,Q,1)$, the tie game starts.

Our s-MDP was constructed with a symmetric view on teams P and Q: The only difference is that team P can choose a strategy whenever in possession of the ball whereas team Q's strategy is fixed. In practice, we aim at optimizing the strategy for team P using a strategy for team Q that has been estimated from earlier games in the same tournament or the like. In Sections 11 and 12 we will discuss sensitivity issues and extend this best-response approach to a finite constant-sum game setting, which allows to prepare for more than one opponent strategy.

Table 1 Strategic MDP (s-MDP)

Strategic MDP	Beach Volleyball Set between Team P and Team Q
Decision Epochs:	$T = \{1,2,3,\ldots\}$
State Sets:	$S = S^{\text{reg}} \cup S^{\text{tie}}$ $S^{\text{reg}} = \{(x,y,k,\ell) \mid x,y \in \{0,\ldots,21\}$ with $x \le 19 \vee y \le 19,$ $k \in \{P,Q\},\ \ell \in \{0,1\}\}$ $S^{\text{tie}} = \{(z,k,\ell) \mid z \in \{-2,\ldots,2\},\ k \in \{P,Q\},\ \ell \in \{0,1\}\}$ $S^{win} = \{(21,y,k,\ell) \in S^{\text{reg}}\} \cup \{(2,k,\ell) \in S^{\text{tie}}\}$ $S^{lose} = \{(x,21,k,\ell) \in S^{\text{reg}}\} \cup \{(-2,k,\ell) \in S^{\text{tie}}\}$
Action Set:	$A_s = \begin{cases} \{risky, safe\} & \forall s = (x,y,P,1) \in S^{\text{reg}},\ s = (z,P,1) \in S^{\text{tie}} \\ \{risky, safe\} & \forall s = (x,y,P,0) \in S^{\text{reg}},\ s = (z,P,0) \in S^{\text{tie}} \\ \emptyset & \text{else.} \end{cases}$ There exist an artificial action at the absorbing states $S^{win} \cup S^{lose}$.
Transitions:	regular game and transition to tie-game Let $s = (x,y,P,1) \in S^{\text{reg}} \setminus \{S^{win} \cup S^{lose}\}$, $a \in A_s$. $p((x+1,y,P,1) \mid s,a) = p^+(P)_a^{serve}$ if $(x,y) \neq (19,20)$, $\quad p((0,P,1) \mid s,a) = p^+(P)_a^{serve}$ if $(x,y) = (19,20)$. $p((x,y+1,Q,1) \mid s,a) = p^-(P)_a^{serve}$ if $(x,y) \neq (20,19)$, $\quad p((0,Q,1) \mid s,a) = p^-(P)_a^{serve}$ if $(x,y) = (20,19)$. $p((x,y,Q,0) \mid s,a) = p^0(P)_a^{serve}$ Let $s = (x,y,P,0) \in S^{\text{reg}} \setminus \{S^{win} \cup S^{lose}\}$, $a \in A_s$. $p((x+1,y,P,1) \mid s,a) = p^+(P)_a^{field}$ if $(x,y) \neq (19,20)$, $\quad p((0,P,1) \mid s,a) = p^+(P)_a^{field}$ if $(x,y) = (19,20)$. $p((x,y+1,Q,1) \mid s,a) = p^-(P)_a^{field}$ if $(x,y) \neq (20,19)$, $\quad p((0,Q,1) \mid s,a) = p^-(P)_a^{field}$ if $(x,y) = (20,19)$. $p((x,y,Q,0) \mid s,a) = p^0(P)_a^{field}$
Transitions:	tie-game Let $s = (z,P,1) \in S^{\text{tie}} \setminus \{S^{win} \cup S^{lose}\}, a \in A_s$. Let $s = (z,P,0) \in S^{\text{tie}} \setminus \{S^{win} \cup S^{lose}\}, a \in A_s$. $p((z+1,P,1) \mid s) = p^+(P)_a^{serve}$ $\quad p((z+1,P,1) \mid s) = p^+(P)_a^{field}$ $p((z-1,Q,1) \mid s) = p^-(P)_a^{serve}$ $\quad p((z-1,Q,1) \mid s) = p^-(P)_a^{field}$ $p((z,Q,0) \mid s) = p^0(P)_a^{serve}$ $\quad p((z,Q,0) \mid s) = p^0(P)_a^{field}$ The transitions of team Q are modelled analogously. $S^{win} \cup S^{lose}$ are modelled as absorbing states and all other transitions have zero probability.
Rewards:	$r(s,a,s') = \begin{cases} 1 & \text{if } s \notin S^{win},\ s' \in S^{win} \\ 0 & \text{else.} \end{cases}$
Objective:	maximize the total expected reward

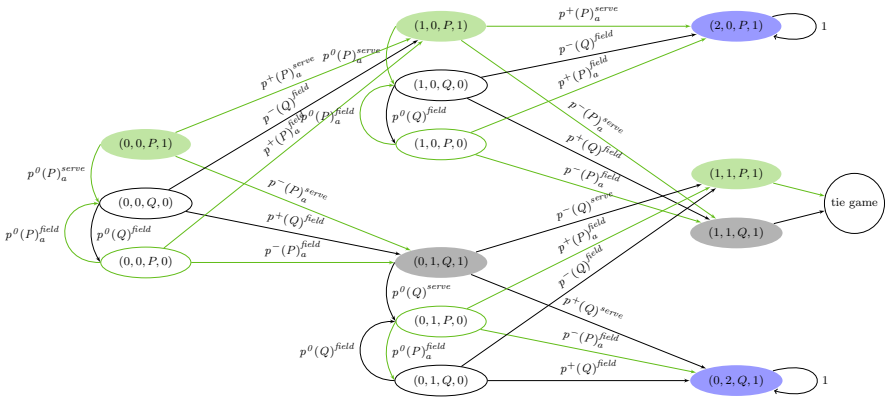

Fig. 1 General s-MDP

It will turn out that the problem to find an optimal policy can be partitioned into two special cases of the s-MDP because the optimal policies for them are myopic (see Appendix 13) and, thus, do not interfere with each other. One case is the serving s-MDP, where the attack types differ only in the serving technique, i.e., $A_s = \{risky, safe\}$ for all serving states $s = (x, y, P, 1)$ of P. The other case is the field attack s-MDP, where the attack types differ in the attack plans used during a field attack, i.e., $A_s = \{risky, safe\}$ for all non-serving states $s = (x, y, P, 0)$ of P.

One may observe that there are many states in the two subproblem s-MDPs in which no action of P is required. In an MDP these states are unnecessary. By concatenating all paths between states in which P has to make a decision, we can transform the s-MDPs into versions where decision states of P, absorbing states and only some additional states – for better readability – occur.

Moreover, since there exists a stationary optimal policy and the choice of an action in A for P depends in both subproblems only on the score, P plays the same action in all decisions states with identical scores. Therefore, all transitions that are neither changing the score nor involve actions of P can be merged. This requires the evaluation of some geometric series in a straight-forward fashion. For simpler notation of the result, we defined the following probability terms for score changes in the serving s-MDP (based on an arbitrary fixed field attack strategy for team P) and in the field attack s-MDP (the events are denoted in parentheses):

$$\alpha_Q^{serve} = \frac{p^-(Q)^{field} + p^0(Q)^{field} p^+(P)^{field}}{1 - p^0(Q)^{field} p^0(P)^{field}}, \qquad \text{(reception } Q\text{, point for } P\text{)}$$

$$\beta_Q^{serve} = \frac{p^+(Q)^{field} + p^0(Q)^{field} p^-(P)^{field}}{1 - p^0(Q)^{field} p^0(P)^{field}}, \qquad \text{(reception } Q\text{, point for } Q\text{)}$$

$$\alpha_P^{serve} = \frac{p^+(P)^{field} + p^0(P)^{field} p^-(Q)^{field}}{1 - p^0(P)^{field} p^0(Q)^{field}}, \qquad \text{(reception } P\text{, point for } P\text{)}$$

$$\beta_P^{serve} = \frac{p^-(P)^{field} + p^0(P)^{field} p^+(Q)^{field}}{1 - p^0(P)^{field} p^0(Q)^{field}}, \qquad \text{(reception } P\text{, point for } Q\text{)}$$

$$\alpha_a^{field} = \frac{p^+(P)_a^{field} + p^0(P)_a^{field} p^-(Q)^{field}}{1 - p^0(P)_a^{field} p^0(Q)^{field}}, \qquad \text{(ball possession } P\text{, point for } P\text{)}$$

$$\beta_a^{field} = \frac{p^-(P)_a^{field} + p^0(P)_a^{field} p^+(Q)^{field}}{1 - p^0(P)_a^{field} p^0(Q)^{field}} \qquad \text{(ball possession } P\text{, point for } Q\text{)}$$

$$\gamma_a^{serve} = p^+(P)_a^{serve} + p^0(P)_a^{serve} \alpha_Q^{serve}. \qquad \text{(service } P\text{, point for } P\text{)}$$

Note that $\alpha_a^{field} + \beta_a^{field} = 1$. Figure 2 and Figure 3 illustrate the outcome of this transformation for the two-point field attack s-MDP. The same transformation is also possible for the tie-game since the structure of the transition probabilities in the tie-game is identical to the regular game.

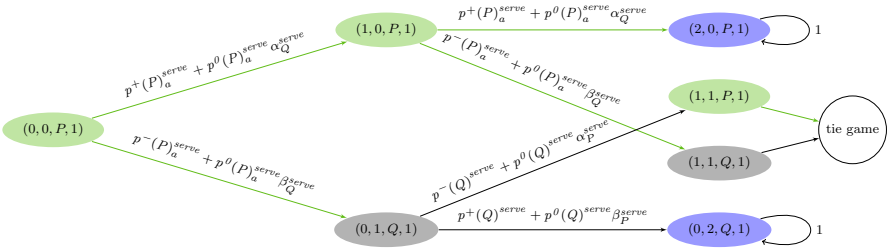

Fig. 2 Serving s-MDP

The s-MDP can be solved analytically for an optimal policy. The total expected reward is monotonically increasing in the number of points of P and monotonically decreasing in the number of points of Q (Appendix 13 provides a formal proof for this plausible proposition). Due to this property, a myopic policy that maximizes the probability to win the next point is optimal. This result was found independently of a result by Walker et al. [23], who proved that given a monotonicity property a myopic policy is optimal for binary Markov games. Since the transition probabilities are identical in every stage of the game, the optimal myopic policy stays the same throughout the game. Our theoretical main result is the following:

Theorem 1 (Optimal Policy – Field Attack s-MDP). *There exists a stationary optimal policy that chooses in each state the action a^* with $\alpha_{a*}^{field} \geq \alpha_a^{field}$ for all $a \in A$.*

Theorem 2 (Optimal Policy – Serving s-MDP). *There exists a stationary optimal policy that chooses in each state the action a^* with $\gamma_{a^*}^{serve} \geq \gamma_a^{serve}$ for all $a \in A$.*

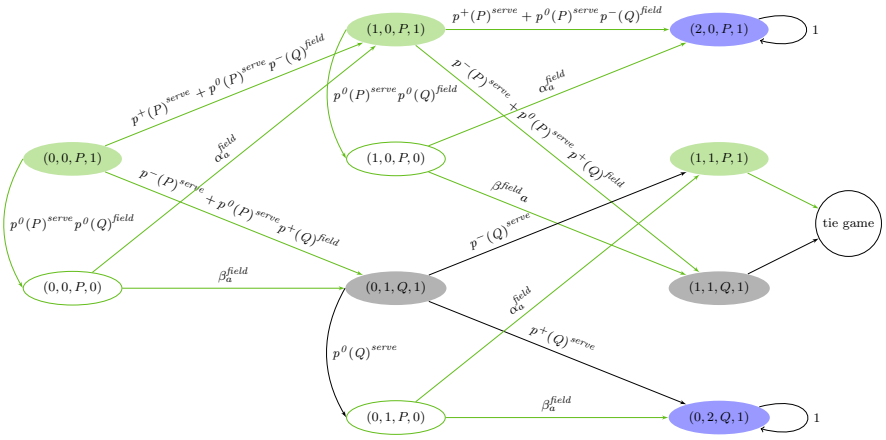

Fig. 3 Field Attack s-MDP

Note that it is important for this result that a rally cannot result in a draw (like in the famous MPD-text-book example on a multiple-game chess match, where it is optimal to play safe when in the lead). Also games in which time marks the end (like in soccer) need a different analysis.

Since we have only two actions in A, *risky* play ($a^* = risky$) is optimal throughout whenever $\alpha^{field}_{risky} - \alpha^{field}_{safe}$ is positive in the field attack s-MDP or when $\gamma^{serve}_{risky} - \gamma^{serve}_{safe}$ is positive in the serving s-MDP, respectively. The optimal policy is unique if these expressions are strictly positive. Observe that α^{serve}_Q depends on the played field attack strategy. For determining the best combination of a field attack and a serving strategy one first has to compute best field attack strategy according to Theorem 1, calculating α^{serve}_Q for this field attack strategy and then apply Theorem 2 to determine the optimal serving strategy based on the optimal field attack strategy.

As an optimal decision rule in the s-MDP does only depend on the situation type, and furthermore only selects between two actions, we define a decision in the s-MDP as a mixture between *risky* and *safe* that only depends on the situation type:

Definition 1 (s-MDP decision rule). A s-MDP decision rule is a mixture between the two actions *risky* and *safe* where

$$mix_{sit}$$

is the fraction by which the action *risky* is chosen in situation *sit*.

Even if Theorem 1 and Theorem 2 are sufficient to characterize an optimal strategy in the whole s-MDP, we want to give an analytic formula for the winning probability of the tie-game. As before, stationary data guarantees the existence of an optimal stationary policy and we can aggregate the transitions and states such that we get a simplified representation that consists only of serving states. The summarized transition probabilities are computed from the original s-MDP transition

probabilities by using the geometric series. The cumulated probability of gaining or losing a point if serving variant a and field attack variant b is played is:

$$\alpha_P^{a,b} := p^+(P)_a^{serve} + p^0(P)_a^{serve} \cdot p^-(Q)^{field} + p^0(P)_a^{serve} \cdot p^0(Q)^{field} \cdot \alpha_{b,P}^{field}$$
$$\beta_P^{a,b} := p^-(P)_a^{serve} + p^0(P)_a^{serve} \cdot p^+(Q) + p^0(P)_a^{serve} \cdot p^0(Q)^{field} \cdot \beta_{b,P}^{field}.$$

So, we compute the terms for every combination of a serving variant a with a field attack variant b. In a serving state of team Q, team P has only to choose a field attack variant b. The cumulated probability of gaining, or losing, a point if field attack variant b is played is:

$$\alpha_Q^b := p^-(Q)^{serve} + p^0(Q)^{serve} \cdot \alpha_{b,P}^{field} \qquad \beta_Q^b := p^+(Q)^{serve} + p^0(Q)^{serve} \cdot \beta_{b,P}^{field}.$$

Figure 4 visualizes the aggregated tie-game.

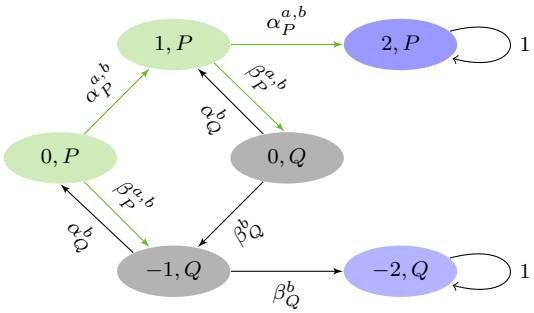

Fig. 4 Tie Game s-MDP

For the winning probabilities $v_{0,P}^{a,b}$ and $v_{0,Q}^{a,b}$ for team P in the tie states $(0,P)$ and $(0,Q)$, we can set up a system of equations

$$\begin{aligned}
v_{0,P}^{a,b} &= \alpha_P^{a,b} \cdot v_{1,P}^{a,b} + \beta_P^{a,b} \cdot v_{-1,Q}^{a,b} \\
v_{1,P}^{a,b} &= \alpha_P^{a,b} \cdot 1 + \beta_P^{a,b} \cdot v_{0,Q}^{a,b} \\
v_{0,Q}^{a,b} &= \alpha_Q^b \cdot v_{1,P}^{a,b} + \beta_Q^b \cdot v_{-1,Q}^{a,b} \\
v_{-1,Q}^{a,b} &= \alpha_Q^b \cdot v_{0,P}^{a,b} + \beta_Q^b \cdot 0.
\end{aligned}$$

Solving this system of equations yields the following formula for the winning probability of P depending on the service strategy a and the field attack strategy b:

$$v_P^{a,b} = \frac{(\alpha_P^{a,b})^2}{(1-\alpha_Q^b\beta_P^{a,b})^2 - \alpha_P^{a,b}\alpha_Q^b\beta_P^{a,b}\beta_Q^b}, \quad v_Q^{a,b} = \frac{\alpha_P^{a,b}\alpha_Q^b(\alpha_P^{a,b}\beta_Q^b - \alpha_Q^b\beta_P^{a,b} + 1)}{(1-\alpha_Q^b\beta_P^{a,b})^2 - \alpha_P^{a,b}\alpha_Q^b\beta_P^{a,b}\beta_Q^b}.$$

We could now answer the benchmark question if we knew the twelve governing probabilities for Teams P and Q. However, whether or not P scores directly depends also on the skills of Q, and vice versa. Thus, a direct estimation of these probabilities would require to make historical observations of Team P for each opponent team separately. In the following section, we ignore the dependence of these probabilities on the opponent and try to estimate them from all matches in the same tournament.

5 Computational Results I

After having found an analytic solution of the s-MDP, we can try to answer the benchmark question raised in the introduction of this paper. First, the analytical solution of the s-MDP showed us that the optimal strategy is myopic. So, independent of the current score, it is always optimal to choose the strategy that maximizes the probability to win the next point. We now turn our attention to the estimation of the s-transition probabilities from historic observations directly.

All our computational results in this paper are based on a video analysis of the Olympics 2012 in London. We took the point of view of strategic consultants for the German team Brink-Reckermann (Team P) for the final against the Brazilian team Alison-Emanuel (Team Q). In order to estimate the direct point and fault probabilities directly from the matches prior to the final, each service and field attack of each team was classified according to whether

- it was played *risky* or *safe* or can not be classified
- whether it led to an immediate point, an immediate fault, or a continued rally.

The absolute counts were used for a maximum-likelihood estimation of all the s-transition probabilities. Their dependence on the opponent team was ignored. For the purpose of comparison we also estimated these probabilities a-posteriori from the final match alone. Note that – although we take this a-posteriori measurement as a yard-stick – the observations in the final match contain only very few realizations of a distribution over possible outcomes.

Any comparison of a-priori estimates with these a-posteriori observations is subject to two types of errors: the a-priori estimates have a certain prediction error that comes from both systematic errors (like ignoring the dependence on the opponent) and sampling errors (like small observation counts), and the a-posteriori estimate has a particularly large sampling error (only the points of a single match are used). Therefore, a discrepancy between the predictions of our model and the actual outcome in the one final should not be solely attributed to modeling/estimation errors. It might as well be the case that the *actual* outcome of the final did not coincide with the *expected* outcome of the final.

Table 2 and Table 3 show the resulting maximum-likelihood estimations of the s-transition probabilities based on the match data. Table 2 evaluated the rallies of the prefinal matches while Table 3 considered only rallies of the final match. Since the s-MDP strategies *risky* and *safe* are characterized by more than one property

(hitting technique and target field) there exist observed attacks that neither belong to the *risky* nor to the *safe* strategy. Therefore, the number of observations stated in the #-column is, e.g., for risky serves in the final, quite small.

The *prefinal-mix* respective *final-mix* is the actual mixture of risky and safe services and field attacks played in the prefinal matches respective in the final match. According to Definition 1, we have

$$\textit{prefinal-mix}_{serve} = \frac{32}{134} \qquad \textit{prefinal-mix}_{field} = \frac{58}{91}$$
$$\textit{final-mix}_{serve} = \frac{1}{39} \qquad \textit{final-mix}_{field} = \frac{12}{23}.$$

Table 2 Direct estimation of s-MDP probabilities – prefinal setting

strategy a	#	$p^+(P)_a^{serve}$	$p^-(P)_a^{serve}$	#	$p^+(P)_a^{field}$	$p^-(P)_a^{field}$	γ_a^{serve}	α_a^{field}	win. Prob.
risky-risky	32	16%	22%	58	66%	17%	37%	34%	77%
risky-safe	32	16%	22%	33	48%	0%	36%	33%	60%
safe-risky	102	4%	9%	58	66%	17%	34%	34%	69%
safe-safe	102	4%	9%	33	48%	0%	32%	33%	50%
prefinal-mix	134	7%	12%	91	59%	11%	34%	34%	65%
	#	$p^+(Q)^{serve}$	$p^-(Q)^{serve}$	#	$p^+(Q)^{field}$	$p^-(Q)^{field}$			
prefinal-mix	165	2%	10%	105	58%	15%			

Table 3 Direct estimation of s-MDP probabilities – postfinal setting

strategy a	#	$p^+(P)_a^{serve}$	$p^-(P)_a^{serve}$	#	$p^+(P)_a^{field}$	$p^-(P)_a^{field}$	γ_a^{serve}	α_a^{field}	win. Prob.
risky-risky	1	0%	0%	12	42%	25%	29%	29%	14%
risky-safe	1	0%	0%	11	64%	9%	34%	34%	73%
safe-risky	38	3%	0%	12	42%	25%	30%	29%	18%
safe-safe	38	3%	0%	11	64%	9%	36%	34%	77%
final-mix	39	3%	0%	23	52%	17%	33%	31%	44%
	#	$p^+(Q)^{serve}$	$p^-(Q)^{serve}$	#	$p^+(Q)^{field}$	$p^-(Q)^{field}$			
final-mix	28	4%	14%	39	59%	15%			

Using the optimality criteria of Theorems 1 and 2, we can compute that a risky service and a risky field attack (short: *risky-risky*) is the optimal policy under the considered policies in the prefinal setting against the Brazilian team playing their prefinal strategy, compare column 8 and 9 in Table 2. By dynamic programming, we computed the winning probabilities in the last column and found that the prefinal

recommended *risky-risky* strategy would have led to the highest winning probability of 77% by a seemingly large margin.

We wish to evaluate how our prefinal recommendation would have proven in the final match. However, an evaluation of the directly estimated s-transition probabilities from the final match can not be given large weight due to the small number of observations. If the compared strategies were more specialized (e.g. requiring certain positions of players on the court), the number of observations would have been even smaller. Also, if a team likes to evaluate a strategy they have not played before in the tournament, a direct estimation of the s-transition probabilities is not possible for that new strategy.

In order to see how well the estimations for the s-transition probabilities and the corresponding s-MDP-based winning probabilities match what actually happened in the final, one can compare the values from Table 3. The resulting differences in winning probabilities are substantial, and it is not clear whether these are due to the prefinal estimations and their systematic short-comings, to the postfinal estimations due to their small number of observations, or to an actual deviation of the teams' skills exposed in the final compared to the matches before.

Since in a direct estimation of the s-transition probabilities from the matches are independent of the opponent, the estimations in the prefinal setting are always identical, no matter which team Germany faces in the final match. Since the s-transition probabilities describe events that depend on both teams (e.g., the probability of an ace depends on the serving skills of the serving team as well as the reception skills of the opponent team) the s-transition probabilities estimates should vary for different opponents. The more the opponent in the final match varies from the prefinal opponents, the more probable it is that the optimal strategy against the opponent in the final is different. It is not possible to directly estimate the s-transition probabilities dependent on the opponent team for two reasons: First, like in our example of the beachvolleyball tournament of the Olympic games 2012, the team may not have faced the opponent of the final match prior to the final. Second, estimating from at maximum one match, the resulting number of observations would be as small as in Table 3. A clustering of opponent teams may be tried to cure the first problem. In this application example, however, Germany did not even play any team of similar strength as Brazil prior to the final.

For those cases in which the results based solely on the s-MDP are not satisfying (as is the case for our benchmark question), we suggest our two-scale approach. The use of an adequate g-MDP can help to overcome the issues resulting from the direct estimation of the s-transition probabilities. Moreover and more importantly, with the new method it will become possible to define and analyse many more strategic options on the basis of the same individual player-skill estimates. This will be demonstrated in Section 12.

6 A Gameplay MDP for Beach Volleyball

This section sketches the infinite-horizon, stationary g-MDP for a beach volleyball rally. The formal details can be found in Appendix 1. Its purpose is to provide an approximation architecture for estimating opponent-dependent s-transition probabilities from simulation runs based on the individual players' skills, which govern the g-transition probabilities.

In this section, it becomes relevant how a point is won in beach volleyball. We summarize briefly the scoring rules as far as they are relevant to the model. A point is won whenever after a hit the ball touches the ground inside the opponent's field. As soon as the opponent touches the ball, this is counted as the hit of the opponent. A point is lost if the ball is hit outside the court, grounds on the own court, or a hit is not executed properly. Moreover, a point is lost when a team touches the ball four times in a row or a player hits the ball twice in succession. A block is a passive contact close to the net in order to rebound an attack-hit of the opponent. The first hit after a block may be executed by both players. However, a blocking contact counts for the rules concerning the number of contacts of a team. Furthermore, a point is lost when a player touches the net. Each rally starts with a service from behind the ground line that must not be blocked.

In order to maintain the symmetry of the gameplay, team Q's action sets will be analogously modelled to the action sets of team P. However, team Q plays a fixed probability distribution over a defined set of attack plans. The transition probabilities encompass both the randomized choices of Q's actions and the realizations of random variables in the system dynamics.

We split up the team actions of a team into two individual player actions. Any player's action will be defined as a combination of a hit and a move. The decomposition of a complex team action into actions of a single player makes the definition of the action sets more manageable. Furthermore, this decomposition allows us to build the g-MDP solely on individual player probabilities, which is the big advantage of the g-MDP.

The g-MDP is an infinite horizon, discrete time MDP with stationary data. Stationary data can be reasonably assumed for a professional beach volleyball rally. In [12], Koch found that the temporal position within a rally did neither effect the type nor the quality of the attack-hit. Our decision epochs are the points in time at which one of the players is about to hit the ball. If a player has decided about his next action, which may be, e.g., an attack-hit, he has to stick to it. This is because an interruption of the current action to revise the decision would lead to a delay. In a professional match this delay will usually outweigh any potential improvement of the new decision. Each time a player contacts the ball, the state of the system is observed and a decision about the next hit or movements must be made by team P. However, there is one exception from that rule: if the ball is touched in a blocking action, then no state will be observed until the ball is rebounded and hits the ground or is touched by the next player. Additionally, a new state is observed when the ball hits the ground or one team has made a fault. In that case, the rally is completed, and

one of the teams has scored a point. We number the points in time in a rally where the state of the system is observed by natural numbers.

The crucial property of our g-MDP is that the g-transition probabilities only depend on the skills of the individual players. To this end, any possible action (which can be a move or a hit) can result in a *success*, a *failure*, or a *deviation*. In our model, positions are classified by a grid on the field. Moves are represented by changes of a player's position in the grid with an upper bound on the range. In our model, moves are always successful. If an attack hit or a service is successful, then the ball lands in the intended grid field on the opponent's court side with the intended hardness. If a defensive hit is successful, then the ball is received and passed on to the intended grid field so that the team mate can continue the counter-attack. A fault means that the hit is in the net or was carried out with an execution fault. A deviation means that the ball passes the net, but not as intended, e.g., it lands in a neighboring grid field (which may also be out). The skill level for a special hit is defined as a probability distribution over these three possible outcomes. Depending on the outcomes of an attack hit, the opponent faces various situations, depending on which his hits lead to success, failure, or deviation. Similarly, depending on the outcomes of a reception, setting and smashing have to be performed in different situations influencing the their probabilities for success, failure, and deviation. An analogous mechanism works for blocking, though slightly more complicated (see Appendix 1 for details). This way, the skills of the two teams and of the two players in a team are decoupled completely in the g-MDP. Therefore, the number of necessary probability estimates is linear in the number of players as opposed to quadratic in the number of teams if the team formations are fixed or even of degree four in the number of players in general. Depending on the g-state, there are various actions possible. Note that this setup also accounts for how the quality of receptions and sets influence the possible follow-up actions: A deviated reception makes subsequent setting impossible, and an additional reception with move has to be performed; a deviated setting results in the impossibility of a smash making a more difficult shot necessary, etc. By a plausibility analysis we excluded from all rule-compliant actions the non-professional choices. The remaining choices are made based on a g-strategy, which is linked to the s-strategies. This will be the topic of the next section.

7 Gameplay MDP strategy

Each team in the g-MDP plays a team specific g-MDP-strategy. A g-MDP-strategy is a variation of some *basic strategy* used as a default together with an modification of the blocking, serving and attack-hit strategies according to parameters that characterize a team strategy, e.g., *risky* or *safe*. Besides these configurable decisions, all strategies in the g-MDP use the same default decision rules of the basic strategy.

In the context of MDPs, a g-MDP-strategy is a stationary policy that consists of a decision rule that prescribes, depending on the state, which action should be chosen.

However, since we are in a sports environment, we will speak of strategies instead of stationary policies.

The basic strategy is implemented to guarantee a reasonable match flow. It excludes unrealistic and moreover obviously non-optimal combinations of player actions or sequences of team actions, and chooses uniformly at random one of the plausible options in each situation.

Some parts of the basic strategy are parametrized such that extreme strategies, like *risky* and *safe*, can be derived from it. We chose in this example to configure the blocking, serving, and attack-hit parts of a strategy. In general, other or more complex parametrizations of the basic strategy are possible. With an implemented basic strategy, it is not necessary to implement an individual decision rule for each possible state in the g-MDP – all straight-forward actions are inherited from the basic strategy.

The decision rules of the basic strategy split up into one decision rule for each state category. The ten different state categories are serving, reception, setting, attacking and defending states from the perspective of both teams. Each category is determined by values of the state variable *counter* and $side(pos(ball))$, where $side(pos(ball))$ states on which court side the ball is. We refrain from a complete definition in print of the straight-forward decision rules of the basic strategy that are used also in all other strategies under consideration – it is just a very long list of very plausible rules.

Instead, we restrict ourselves to those decision rules in which the strategies of interest differ. We represent all strategies as randomized policies over identical sets of plausible deterministic policies representing extremal ways to play. The investigated strategies only differ in the selection probabilities. This way we obtain a parametrized set of randomized strategies. More specifically, all strategy-specific decision rules are encoded by a vector π whose components determine the probability for choosing true in a Boolean decision. It represents the probabilities with which the strategy chooses one out of two extremal ways to play in various dimensions. In the basic strategy, e.g., all components of π are set to 0.5 which means that in each dimension both decision possibilities are equally probable.

For example, the blocking strategy is specified by π_b, which states with which probability player 1 of a team is the designated blocking player in the next rally. It follows that with probability $(1-\pi_b)$ player 2 is the blocking player. The parameter π_s determines the serving strategy of a team. With probability π_s, a serve on player 1 of the opponent team is made, i.e., the target field of the serve belongs to the opposing court half that is covered by player 1.

Further, a technique and target field decision of the serve and attack-hit are included in π_h. The two parts π_h^{serve} and π_h^{field} of π_h comprise the strategies corresponding to service and field attack, respectively. Each component further splits up into a technique and target field decision that can be different for both players ρ, i.e., $\pi_h^{sit} = (\pi_{h,tech}^{sit}(\rho),\ \pi_{h,target}^{sit}(\rho))^T$ with $sit \in \{serve, field\}$. The subscript term indicates if the decision is related to the technique (*tech*) or target field (*target*) decision. Now, we can summarize all parameters that are necessary for defining a g-MDP strategy of team P:

Definition 2 (g-MDP strategy). A strategy of the g-MDP is a parametrization of the basic strategy and characterized by the parameters:

$$\pi = \begin{pmatrix} \pi_h \\ \pi_b \\ \pi_s \end{pmatrix}, \quad \pi_h = \begin{pmatrix} \pi_h^{serve} \\ \pi_h^{field} \end{pmatrix}, \quad \pi_h^{sit} = \begin{pmatrix} \pi_{h,tech}^{sit}(\rho) \\ \pi_{h,target}^{sit}(\rho) \end{pmatrix}, \; \begin{array}{l} sit \in \{serve, field\}, \\ \rho \in \{P_1, P_2\}. \end{array}$$

For a higher memorability, we defined the values of the components of π_h always as the probability for the more risky opportunity. In our example, we have two serving techniques available in the g-MDP, namely the float serve S_F and the jump serve S_J. The float serve is considered as a safe hit and the jump serve as a risky hit. (All classifications of this type have been determined by personal communication with high-level amateur beach volleyball players.) So $\pi_{h,tech}^{serve}(\rho)$ is defined as the probability that ρ chooses a S_J. For the attack-hit, we have three techniques available the smash F_{SM}, a planned shot F_P and an emergency shot F_E. The emergency shot is normally only played if none of the other attack-hits is possible, and in such a case it is chosen with certainty by each strategy. The smash is considered as a risky hit and the planned shot as a safe hit. So $\pi_{h,tech}^{field}(\rho)$ is defined as the probability that ρ chooses a F_{SM}. Furthermore, we define all fields that are near the touch of the court as border fields. For example, on court side of team Q the border fields are $\partial F := \{Q11 - Q31, Q14 - Q34\}$. These are more risky target fields than non-border fields. So $\pi_{h,target}^{serve}(\rho)$ and $\pi_{h,target}^{field}(\rho)$ are the probabilities with which a border field is chosen as a target field. Should there be several possible *risky* or *safe* options the *risky* or *safe*, respectively, strategy chooses one of them uniformly at random. Using this randomization is a means to prevent complete predictability, although this advantage cannot be measured in the reward function of the current MDP-setup. It can be seen as injecting some general key learnings from game theory into the system.

After having introduced the general concept of a g-MDP strategy, we want to specify two hitting strategies that implement the s-MDP strategies *risky* [*safe*] as g-MDP strategies. For answering our benchmark question, we will compare them later in the computational results section, see Section 9. We call the two special strategies the risky hitting strategy π_h^{risky} and the safe hitting strategy π_h^{safe}. They are the most extreme hitting strategies. The strategy π_h^{risky} takes always a risky technique and chooses always a border field as target field. The π_h^{safe} strategy chooses always a safe hit with a non border field as target field. Table 4 summarizes the techniques and target fields chosen by the two extreme strategies. The assignment of $risky \mapsto \pi_h^{risky}$ and $safe \mapsto \pi_h^{safe}$ is the s-g-implementation of our benchmark question. In the following, we will refer to π_h^{risky} [π_h^{safe}] when we write about the *risky* [*safe*] strategy in the g-MDP-setting.

Table 4 Overview risky hitting strategy π_h^{risky} versus safe hitting strategy π_h^{safe}

Strategy	π_h^{risky}	$\pi_h(\rho)$	π_h^{safe}	Strategy	π_h^{risky}	$\pi_h(\rho)$	π_h^{safe}
Serve				Attack-Hit			
serving technique S_J	1	$\pi_{h,tech}^{serve}(\rho)$	0	attack technique F_{SM}	1	$\pi_{h,tech}^{field}(\rho)$	0
border field	1	$\pi_{h,target}^{serve}(\rho)$	0	border field	1	$\pi_{h,target}^{field}(\rho)$	0

8 Gameplay MDP validation

For calibrating the g-MDP for the German team Brink-Reckermann against the Brazilian team Alison-Emanuel in the final match of the Olympic 2012 games, we need estimations for the skills of all players as input parameters. To estimate the skills, we evaluated all matches they played in the tournament except the final match. Details of the data collection process and the complete presentation of all data tables can be found in [10]. For illustration purposes, the skill estimates for Julius Brink based on the pre-final matches are included in Table 5 and Table 6.

Table 5 Input data from all matches except final: Julius Brink – Serves and Attack-Hits

	target fields	Q11-Q14			Q21-Q24			Q31-Q34		
	performance	#	*succ*	*fault*	#	*succ*	*fault*	#	*succ*	*fault*
Serve										
S_F	P01 - P04	34	0.88 (0.88)	0.00 (0.00)	43	0.88 (0.88)	0.12 (0.12)	-	-	-
S_J	P01 - P04	34	0.94 (0.94)	0.00 (0.00)	16	0.75 (0.75)	0.19 (0.19)	-	-	-
Attack-Hit										
F_{SM}	out	0	0.86 (–)	0.02 (–)	0	0.86 (–)	0.02 (–)	-	-	-
	P11-P14	0	0.86 (–)	0.02 (–)	0	0.86 (–)	0.02 (–)	-	-	-
	P21-P24	55	0.85 (0.85)	0.04 (0.04)	17	0.94 (0.94)	0.00 (0.00)	-	-	-
	P31-P34	7	0.77 (0.71)	0.01 (0.00)	2	0.89 (1.00)	0.02 (0.00)	-	-	-
F_E	out	0	0.76 (–)	0.06 (–)	0	0.76 (–)	0.06 (–)	-	-	-
	P11-P14	0	0.76 (–)	0.06 (–)	1	0.79 (1.00)	0.05 (0.00)	-	-	-
	P21-P24	7	0.73 (0.71)	0.11 (0.14)	7	0.82 (0.86)	0.02 (0.00)	-	-	-
	P31-P34	1	0.70 (0.00)	0.05 (0.00)	1	0.79 (1.00)	0.05 (0.00)	-	-	-
F_P	out	0	0.95 (–)	0.05 (–)	0	0.95 (–)	0.05 (–)	0	0.95 (–)	0.05 (–)
	P11-P14	0	0.95 (–)	0.05 (–)	0	0.95 (–)	0.05 (–)	0	0.95 (–)	0.05 (–)
	P21-P24	8	0.99 (1.00)	0.01 (0.00)	30	0.97 (0.97)	0.03 (0.03)	0	0.95 (–)	0.05 (–)
	P31-P34	2	0.96 (1.00)	0.04 (0.00)	3	0.88 (0.67)	0.12 (0.33)	0	0.95 (–)	0.05 (–)

Table 6 Input data from all matches except final: Julius Brink – Defence, Reception, Set, Block

	attack strength		*normal*			*hard*	
	performance	#	*succ*	*fault*	#	*succ*	*fault*
Defence	d	20	0.85 (0.85)	0.05 (0.05)	14	0.71 (0.71)	0.21 (0.21)
	d_m	29	0.93 (0.93)	0.00 (0.00)	13	0.46 (0.46)	0.38 (0.38)
Reception	r	34	1.00 (1.00)	0.00 (0.00)	9	0.90 (0.89)	0.10 (0.11)
	r_m	42	0.95 (0.95)	0.02 (0.02)	3	0.97 (1.00)	0.02 (0.00)
Set	s	117	0.99 (0.99)	0.00 (0.00)	-	-	-
	performance	#	*direct point*	*over net but no point*	*fault*	*misses ball*	
Block	b	5	0.20	0.20	0.20	0.40	

In Table 5, the maximum-likelihood estimates of the individual player probabilities of Julius Brink for all types of serves and attack-hits are presented. We aggregated different player positions and target fields together to get a larger number of observations. The number of observations for a certain combination of player position and target field is stated in the #-column. The probabilities shown in brackets are the maximum-likelihood estimates for the specified hit whereas the other probabilities are the maximum a-posteriori probability estimations which include a prior assumption [14]. For categories with more than eleven observations both probabilities are equal. More details on the a-posteriori skill estimation can be found in [10]. The column *succ* states for each combination the probability that the hit lands in the target field and the column *fault* contains the probability of a technical error. The remaining probability is the probability that the hit was successful but the ball deviated into a neighbour-field of the target field.

Table 6 specifies the estimated probabilities of Julius Brink for defence, receptions, settings, and blocks. The estimated probabilities fit our intentions that we had when we defined the hits, e.g., receptions have a higher success rates than defence actions and hard balls are harder to defend or receive than normal balls. For the blocking skills, the first three columns after the number of observations describe the possible results of a block that catches the ball, while the last column is the probability that the block misses the ball. Since Jonas Reckermann is the designated blocking player in the German team, Julius Brink has done nearly no blocks in all these matches [10]. We have done the same estimations of the individual probabilities of the other players. The respective tables are presented in [10].

Before going on with strategic recommendations in the next section, we want to check how well the g-MDP model fits a real beach volleyball match. We use the final match of the Olympic games as a benchmark for our model and the estimated input skills. It is the only match of the tournament where we have estimations of the skills of both teams.

Table 7 Estimated final and prefinal strategies of Brink-Reckermann and Alison-Emanuel of the Olympic 2012 games

	final π^{final}					prefinal $\pi^{prefinal}$				
	GER		BRA			GER		BRA		
	Brink	Reckermann	Alison	Emanuel		Brink	Reckermann	Alison	Emanuel	
$\pi^{serve}_{h,tech}$	19.23%	24.14%	68.00%	36.67%	$\pi^{serve}_{h,tech}$	39.37%	50.41%	37.24%	35.34%	
$\pi^{serve}_{h,target}$	3.85%	17.24%	16.00%	26.67%	$\pi^{serve}_{h,target}$	18.90%	36.36%	17.93%	17.29%	
$\pi^{field}_{h,tech}$	73.91%	69.23%	92.86%	81.48%	$\pi^{field}_{h,tech}$	65.32%	71.67%	82.00%	86.19%	
$\pi^{field}_{h,target}$	26.09%	33.33%	42.86%	53.70%	$\pi^{field}_{h,target}$	38.71%	45.00%	36.00%	30.94%	
π_b	1.56%		96.08%		π_b	2.44%		95.85%		
π_s	27.27%		34.55%		π_s	50.00%		50.00%		

Table 7 shows the strategy estimations for both teams in terms of the strategy definition of Section 7. It contains the strategy estimated from observations of the final match and the estimation from the prefinal matches. The estimated strategy of the final is used for validating the g-MDP model definition. In the prefinal strategy, we used 50% as the estimate for the team's serving strategy. Since the teams faced different opponents in the prefinal matches, we could not derive any meaningful value from the observations. Recall, that the classification of a hit as *safe* or *risky* depends on the state of the system, e.g., whether or not the setting resulted in a deviation or not.

We collected the realized s-MDP transition probabilities from the Olympic final by counting the number of serves and field attacks as well as the direct points and faults. These values are presented in the first line of Table 8.

For validating our approach, we simulated 1000 batches of 100 beach volleyball rallies each where both teams played the estimated strategy of the final match. We implemented a special-purpose simulation code in Java. The g-MDP simulation has been slightly tweaked from the ideal descriptions of the g-MDP in order to match as closely as possible our interpretation of the video data for the data collection.

By counting the number of serves and field attacks as well as their outcome (direct point, fault or a subsequent attack), we calculated the s-MDP transition probabilities from the g-MDP simulation by a maximum-likelihood estimation. That is, we counted the number of times the action was performed and the number of times it resulted in a success, a failure, or a deviation. The quotients were taken as estimates for the probabilities. Since the case of a deviation of target field for an attack hit is difficult to judge upon (because we cannot tell the intention from the video), we only classified outs as a deviation. Whenever there were fewer than eleven (a number determined in many experiments) observations for an action, we used actions from an extended category to add additional observations. The results for different skill estimations are shown in the last three lines of Table 8.

Table 8 Validation of simulated s-MDP transition probabilities based on different skill estimates

estimation method	$p^+(P)^{serve}_{\pi^{final}}$	$p^-(P)^{serve}_{\pi^{final}}$	$p^+(Q)^{serve}_{\pi^{final}}$	$p^-(Q)^{serve}_{\pi^{final}}$
realized probabilities final	2%	4%	4%	14%
simulating the g-MDP with				
skills of all matches except final	2%	15%	3%	15%
skills of all matches	2%	12%	4%	14%
skills of final only	1%	2%	6%	10%
estimation method	$p^+(P)^{field}_{\pi^{final}}$	$p^-(P)^{field}_{\pi^{final}}$	$p^+(Q)^{field}_{\pi^{final}}$	$p^-(Q)^{field}_{\pi^{final}}$
realized probabilities final	49%	17%	55%	16%
simulating the g-MDP with				
skills of all matches except final	32%	15%	26%	19%
skills of all matches	36%	15%	36%	19%
skills of final only	46%	12%	50%	15%

The deviations of the predictions in the prefinal row from the direct observations in the final in the very first row are partly small and encouraging and partly quite large. However, as discussed before, one should not consider the first row as containing the "true" probabilities because the values from the final are an estimation for the probabilities, too, and one with a large sampling error, given the small number of observations. Still, it can be seen that the pattern of which probability is large and which probability is small looks quite similar. We have to keep in mind, though, that any interpretation of an outcome of our analysis must be accompanied by a thorough sensitivity analysis. We will show a possibility to implement a concept for this in Section 11.

9 Computational Results II

After having found an appropriate g-MDP model, we can estimate s-MDP transition probabilities from simulating the g-MDP and answer the benchmark question raised in the introduction of this paper. Using the g-implementation of *risky* and *safe* as described in Table 4 and the a-priori skill estimations of all player, we can estimate the s-MDP transition probabilities, see Table 9. Assuming Brazil plays a similar strategy as their prefinal strategy $\pi^{prefinal}$, their s-MDP transition probabilities can be estimated from the g-MDP simulation as well, see last line in Table 9.

Using the optimality criteria of Theorem 1 and 2, we can compute that a risky service and a risky field attack (short: *risky-risky*) is the optimal policy under the considered policies in the prefinal setting against the Brazilian team playing their

Table 9 Estimation of s-MDP probabilities from g-MDP simulation – prefinal setting

strategy a	$p^+(P)_a^{serve}$	$p^-(P)_a^{serve}$	$p^+(P)_a^{field}$	$p^-(P)_a^{field}$	γ_a^{serve}	α_a^{field}	winning Prob
risky-risky	5%	20%	40%	16%	46%	55%	80%
risky-safe	5%	20%	16%	13%	39%	46%	31%
safe-risky	1%	13%	40%	16%	48%	55%	84%
safe-safe	1%	13%	16%	13%	40%	46%	34%
$\pi^{prefinal}$	2%	16%	32%	14%	45%	53%	73%

	$p^+(Q)^{serve}$	$p^-(Q)^{serve}$	$p^+(Q)^{field}$	$p^-(Q)^{field}$
$\pi^{prefinal}$	2%	12%	24%	18%

Table 10 Estimation of s-MDP probabilities from g-MDP simulation – postfinal setting

strategy a	$p^+(P)_a^{serve}$	$p^-(P)_a^{serve}$	$p^+(P)_a^{field}$	$p^-(P)_a^{field}$	γ_a^{serve}	α_a^{field}	winning Prob
risky-risky	2%	8%	52%	14%	36%	37%	49%
risky-safe	2%	8%	40%	12%	33%	34%	25%
safe-risky	1%	2%	52%	14%	37%	37%	52%
safe-safe	1%	2%	40%	12%	34%	34%	28%
π^{final}	1%	2%	47%	12%	36%	36%	41%

	$p^+(Q)^{serve}$	$p^-(Q)^{serve}$	$p^+(Q)^{field}$	$p^-(Q)^{field}$
π^{final}	6%	10%	51%	14%

prefinal strategy. This result coincides with the prefinal recommendation based on the direct estimated transition probabilities presented in Section 5.

The reader may have noticed that there are differences in the performances and the played strategy when comparing the prefinal matches with the final match: The skill estimates based on the prefinal matches differ from the estimates based on the final match, compare e.g. Tables for the skills of Reckermann in [10]. Furthermore, the strategy of Brazil in the final match deviates slightly from their prefinal strategy, see Table 7. Because of these differences, we wish to evaluate how our prefinal recommendation would have proven in the final match. Table 10 shows the estimated s-MDP transition probabilities estimated from the g-MDP simulation provided with the postfinal setting. Applying again the optimality criteria of Theorems 1 and 2, we derive that in the postfinal setting a *safe* service strategy and a *risky* field-attack strategy would be the best response to the Brazilian final strategy. However, the a-priori recommendation proves to be quite good. By dynamic programming, we computed the winning probabilities and found that the a-priori recommended *risky-risky* strategy would have led in the postfinal setting to a winning probability of 49%, which is better that the actual played strategy of Germany in the final (41% winning probability) and only slightly worse than the optimal *safe-risky* strategy (52% winning probability).

It should be said that we experienced deviations between our predictions and the outcomes of the final in this case as well, as in the pure s-MDP case in Section 5. The deviations are attributed to changes in the individual skills. The skill estimates from the prefinal matches differ from skill estimates based on the final match only. However, the number of observation of one match is not satisfactory to estimate the individual skills: The whole point of basing strategic recommendation on skill estimates rather than direct point probabilities is that the data of all matches prior to the final can be used independently of the opponent.

In order to keep this paper focused on the as-simple-as-possible benchmark question, we sticked to a comparison of two possible strategies only. Given the level of detail in the g-MDP, we could easily compare more strategy combinations involving the blocking player or the positioning of the players in the field *on the basis of the same skill estimations*. Note, moreover, that the skill estimations used in this paper are based on very few data. In practice, we suggest to evaluate the skills of our team also in training sessions and other matches prior to the tournament.

10 Comparison of Methods

We have carried out the whole concept for the German beach volleyball team Brink Reckermann at the the Olympic games 2012. We tried to give a recommendation for the German prior to the final match against Brazil. Thereby we answered the strategic question "does *risky* or *safe* play lead to a higher winning probability?" twice: first, by using direct estimates of the s-MDP transition probabilities and a second time by using estimates from from the g-MDP simulation. The prefinal recommendations of both methods coincided. Since both estimation methods for the s-transition probabilities have independent weaknesses, in this case the prefinal recommendation can be trusted even more. It seems that the two-scale approach did not gain anything new. However, that can only be said for the recommendation alone. In Sections 11 and 12 we will see how the two-scale approach yields sensitivity and game theoretic insights about strategies that have *never been played before*, which is even more important in the presence of only loosely validated data estimation. This would be impossible with the direct estimation of s-transition probabilities from historic data only.

In Table 11 and Table 12, we present the estimation results of the s-MDP transition probabilities for the final strategy to compare both estimation methods. The Table splits up into two Tables: Table 11 presents the estimates for the serving situation and Table 12 for the field attack situation. Both methods, the direct estimation of the s-MDP transition probabilities, see Section 5, and the estimation from the g-MDP simulation, see Section 9, are compared to the realized transition probabilities in the final match. In the g-MDP simulation, we used as an estimate for the final strategy π^{final} the values of Table 7. The results of the g-MDP simulation, i.e., line 2 and 4 in each Table, are a compilation of the results of Table 8 where we validated the g-MDP simulation. The results of the direct estimation based on the final

match is also a compilation of the results presented in Table 3. However, the direct estimation for the final strategy based on the prefinal matches contains new values. We used the prefinal estimates for *risky* and *safe* of Table 2 and mixed them by the *final-mix*. For example, for the serving probabilities of team P, we get,

$$p^+(P)^{serve}_{\textit{final-mix}} = \textit{final-mix}_{serve} \cdot p^+(P)^{serve}_{risky} + (1 - \textit{final-mix}_{serve}) \cdot p^+(P)^{serve}_{safe}$$
$$p^-(P)^{serve}_{\textit{final-mix}} = \textit{final-mix}_{serve} \cdot p^-(P)^{serve}_{risky} + (1 - \textit{final-mix}_{serve}) \cdot p^-(P)^{serve}_{safe}$$

by the direct estimation method based on prefinal matches. As the proportion team Q used a *risky* or a *safe* strategy in the final and the estimates for the transition probabilities of those strategies has not been presented yet, we put them in Table 15 in the Appendix.

Table 11 Comparison between two scale and direct approach: The realized s-MDP transition probabilities of the final are the benchmark – strategicMDP transition probabilities for serving situation.

estimation method	data	dec. rule	$p^+(P)^{serve}$	$p^-(P)^{serve}$	$p^+(Q)^{serve}$	$p^-(Q)^{serve}$	Avg. L^1-Error
observations final			2%	4%	4%	14%	-
g-MDP simulation	prefinal skills	π^{final}	2%	15%	3%	15%	3.27%
s-MDP direct	prefinal matches	*final-mix*	4%	9%	2%	12%	2.86%
g-MDP simulation	final skills	π^{final}	1%	2%	6%	10%	2.23%
s-MDP direct	final match	*final-mix*	3%	0%	4%	14%	1.10%

Although, the average absolute difference of the realized transition probabilities of the final match and the direct estimates of the field attack transition probabilities is in all cases smaller than the deviation of the estimates from the g-MDP simulation, the approximation errors are of a similar order of magnitude, given that the "benchmark" count in the final represents only the counts of one sample match, which is a random experiment, too. Only the direct point probability for Brazil after a field attack has been underestimated by a large margin (26% versus 55%) by the g-MDP: This is subject to further investigation.

Note that the direct estimation of the s-transition probabilities relevant for the final could only be done for those strategies that have been played before. If one wants to optimize over strategy sets containing strategies that have not been played before (in the tournament against similar opponents), then the g-MDP simulation is the only method that yields estimates at all.

Table 12 Comparison between two scale and direct approach: The realized s-MDP transition probabilities of the final are the benchmark – s-MDP transition probabilities for field attack situation.

estimation method	data	dec. rule	$p^+(P)^{field}$	$p^-(P)^{field}$	$p^+(Q)^{field}$	$p^-(Q)^{field}$	Avg. L^1-Error
observations final			49%	17%	55%	16%	-
g-MDP simulation	prefinal skills	π^{final}	32%	15%	26%	19%	12.87%
s-MDP direct	prefinal matches	*final-mix*	57%	9%	60%	17%	5.71%
g-MDP simulation	final skills	π^{final}	46%	12%	50%	15%	3.40%
s-MDP direct	final match	*final-mix*	52%	17%	59%	15%	2.27%

11 Sensitivity and Skill Strategy Score Cards

In the following, we discuss what we call *Skill-Strategy Score Cards*. Skill-Strategy Score Cards are a visualization of the sensitivity of strategy recommendations on probability estimates. Given the substantial uncertainty in the probability estimates, this is paramount to the correct assessment of overly detailed computational results in practice. They indicate for various individual skill levels and for various opponent types the differences in the winning probabilities of two strategies. The skill probabilities $p_{succ,\rho}(pos(\rho), h)$ and $p_{fault,\rho}(pos(\rho), h)$ for one hit h are varied in each small plot from zero to one. The $p_{dev,\rho}(pos(\rho), h)$ is implicitly determined through the two varied probabilities. The colour of each square-shaped data point in the plot reflects the difference between the winning probabilities of the safe hitting strategy π_h^{safe} and the risky hitting strategy π_h^{risky}, that were both introduced in Section 7. A green colour means that it is better to play *safe*, the red colour suggests *risky* play, and the yellow colour indicates that there is no difference in the winning probabilities of both strategies.

As a reminder, the smash is played in a field attack in *risky* whereas the planned shot is played in *safe*. Lines in each small plot indicate the real skill level of the varied hit. For example, in Figure 5 the smashing skills are varied, and the three lines indicate the average skill level of the smash as presented in the input data.

One can use a Strategy-Skill Score Card as follows: Imagine you want to choose between *safe* and *risky* field attack play against a particular opponent. You can estimate all necessary s-transition probabilities, either as described in this paper via the direct estimation or via the simulation of the g-MDP (or by some other method). You are interested in the sensitivity of the recommendation w.r.t. the opponent's strength and your players' skills on complementary hits representing the *safe* and *risky* field attack styles. For example, the skill level for a planned shot influences the winning probability of the *safe* attack style, and the skill level of smash (the complementary

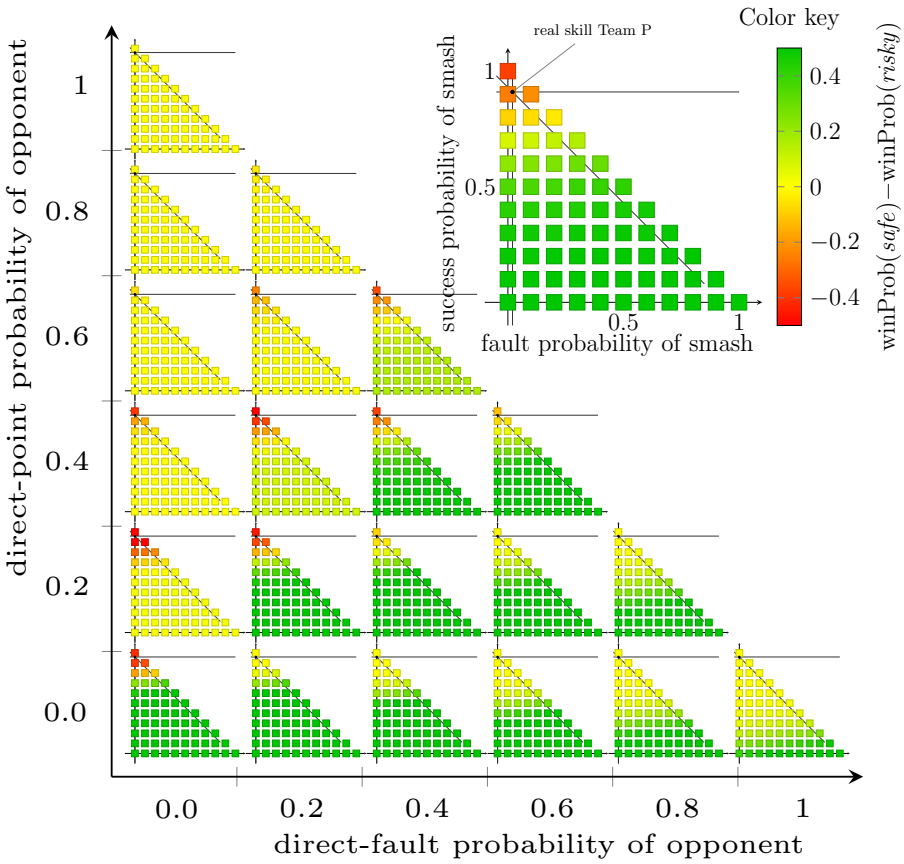

Fig. 5 Skill-Strategy Score Card: difference between the winning probabilities of *safe* and *risky* play for different opponents and varying smash skills $p_{succ,\rho}\,(pos(\rho),\, F_{SM})$ and $p_{fault,\rho}\,(pos(\rho),\, F_{SM})$

hit) influences the winning probability for the *risky* attack style; you want to know how the strategic recommendation changes as you vary the skill levels of these hits. Then you do the following:

1. Produce a Strategy-Skill Score Card for smash with variable skills for smash and the estimated skills for shot. This yields a chart like the one in Figure 5. Do the same with variable shot skills and the estimated smash skills. This yields a chart like the one in Figure 6.

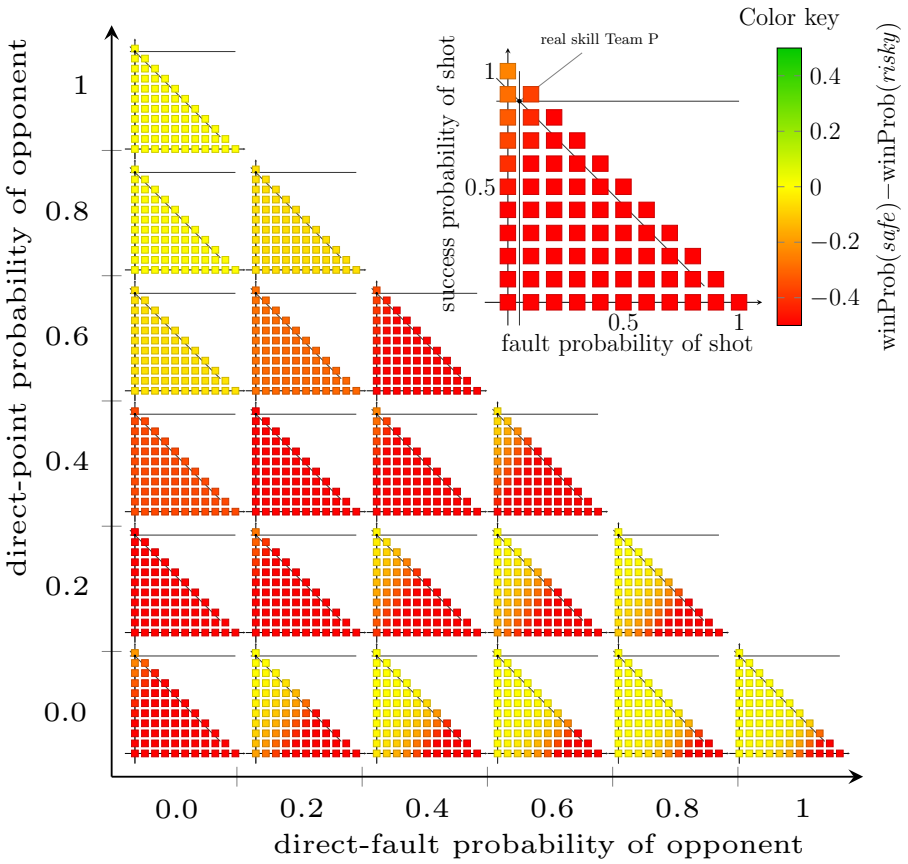

Fig. 6 Skill-Strategy Score Card: difference between the winning probabilities of *safe* and *risky* play for different opponents and varying shot skills $p_{succ,\rho}\,(pos(\rho),\, F_P)$ and $p_{fault,\rho}\,(pos(\rho),\, F_P)$

2. Focus on the little chart at the respective $(p^+(Q)^{field}, p^-(Q)^{field})$-coordinate, where $p^+(Q)^{field}$ denotes the estimated opponent's direct-point probability and $p^-(Q)^{field}$ the estimated opponent's error-probability.
3. Focus on the square at the $(p_{succ,\rho}\,(pos(\rho),\, F_{SM}), p_{fault,\rho}\,(pos(\rho),\, F_{SM}))$-coordinate in the little chart, where $p_{succ,\rho}\,(pos(\rho),\, F_{SM})$ denotes your players' estimated success probability for smash and $p_{fault,\rho}\,(pos(\rho),\, F_{SM})$ the corresponding error-probability.

4. The result for the estimated probabilities is: the greener the square, the more does *safe* outperform *risky*.
5. Since all probabilities are only estimates, the graphics show some sensitivities: The neighbouring little squares show how the superiority of *safe* over *risky*, or vice versa, changes as your players' smash skills (or planned shot skills, respectively) vary: the squares above are for larger success probabilities, the squares to the right are for larger error probabilities, and the squares to the top-right are for larger deviation probabilities (no error but with deviation into a neighbouring field).
6. The neighbouring little charts show how the superiority of *safe* over *risky*, or vice versa, changes as your opponent's strength varies: the little charts above are for a larger direct point probability of your opponent in the field, the little charts to the right for a larger error probability of your opponent in the field. This way, our team can base a decision on a larger area of plausible probabilities.
7. Moreover, it is possible to assess the critical probability values where the superiority of *safe* over *risky* flips fast (narrow yellow areas between green and red).

Let us draw some conclusions from the example cards in Figures 5 and 6.

If the opponent is strong (many direct points), then we have to look at the top little chart, where the difference between the two strategies is very small (yellow all over): Against such a strong opponent, the choice of a strategy does not matter.

We see in Figure 5 that for the field strategies a weak opponent (opponent with many direct errors) leads to the yellow-green little chart to the right that has yellow only if the skills for a successful-and-on-target smash are large enough. That is, the *risky* field strategy against such an opponent is quite robustly never better than *safe*, and both strategies are equally good if our smash skills are good enough. The little chart at the origin, however, (opponent with few direct points and few errors) shows a sharp dependence of the superiority of *safe* over *risky* on the smash skills of our players. This is plausible because against such an opponent we get many more chances for a smash during a rally, and its quality will influence the winning probability of a *risky* attack style substantially.

12 Extension: Two Person Constant Sum Game

Using the two-scale approach and the skill data from the previous section, we were able to generate the two person constant sum game presented in Figures 7 and 8. We evaluated the s-MDP with transition probabilities resulting from a simulation of the g-MDP based on the estimated individual skills. Tables 7 uses skill estimates from the prefinal matches and Table 8 skill estimates from the final match. Each Tables presents the winning probability of Germany for 32×32 strategy combinations. These Tables are an extension of the benchmark question, which compares only two strategies against a static opponent, to game theory, where also the opponent team can vary between strategies. Note that a generation of such a table from direct estimated s-transition probabilities would require a massive amount of data to

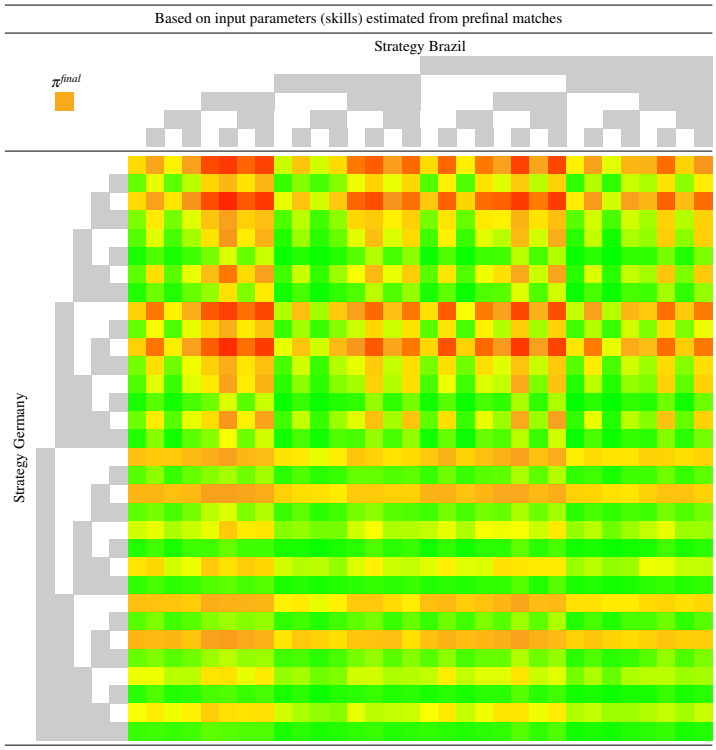

Fig. 7 Winning probabilities for Germany for different strategy combinations of both teams in the pre final setting.

get enough estimates that belong to a strategy combination. Considering how often certain teams meet within a season, this project would seem to have no chance of success.

The strategies compared in the tables are all extreme strategies that are generated by the g-MDP strategy parameters

$$\pi_b,\ \pi_{h,*}^{serve}(\rho 1),\ \pi_{h,*}^{field}(\rho 1),\ \pi_{h,*}^{serve}(\rho 2),\ \pi_{h,*}^{field}(\rho 2).$$

Observe that the used technique and the target field for one situation are combined into one value. The serving strategy π_s is in the prefinal setting fixed to 0.5 and in the postfinal setting to the observed value of π_s in π^{final}. By an extreme strategy, we mean that each parameter is alternating between 0 and 1. So, we consider 5 parameters, each of which can be 0 or 1, resulting in 32 different strategies. For presenting the strategies in the table in a clear manner, we use a pattern to indicate the used strategy. A parameter that takes the value 0 is represented by a white coloured field, a 1 by a grey coloured field. Furthermore, we used the ordering of the parameters as presented above. So, for example, the first line in both tables corresponds to Ger-

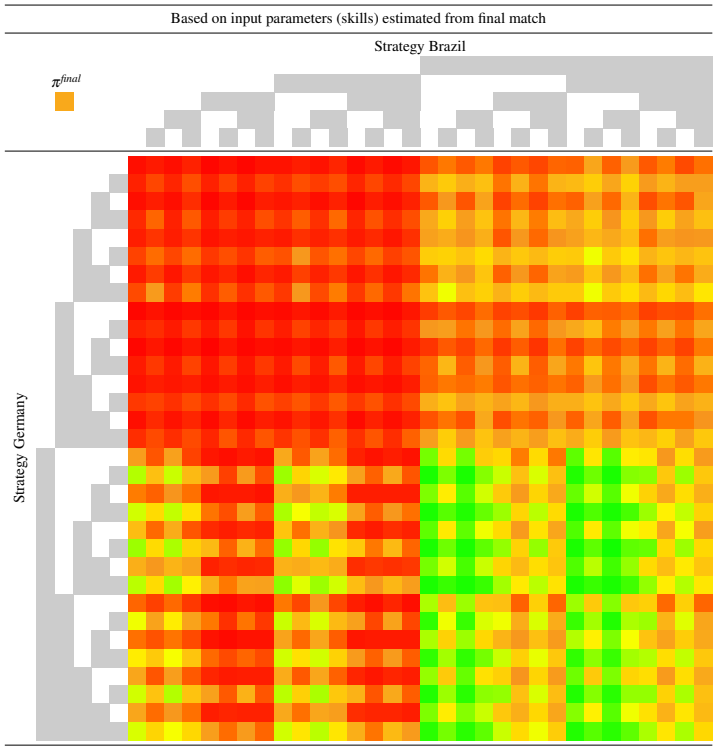

Fig. 8 Winning probabilities for Germany for different strategy combinations of both teams in the postfinal setting.

many playing the strategy $0,0,0,0,0$, which means that player 2 is always blocking and both player play a safe serve and a save field attack.

Pure green means Germany wins; yellow depicts a 50-50-chance; and red means Germany loses. The intermediate colours indicate the intermediate values. We comment only on obvious patterns. Prior to the match we would have recommended the following for Germany:

- Since the lower half of the table is greener, player 1 (Brink) should be the blocking player for Germany. This need not necessarily come from blocking skills but also from the order in which an attack is performed: the blocking player is most often also the setting player. This result surprises us, since actually player 2 (Reckermann) is the usual blocking player for Germany. And most probably for some reason. We still have to investigate whether this is an artefact or a reasonable option.
- Every other row is greener, thus, player 2 should play *risky* field attacks; this is also the case for player 1, but very less so (only a slight visible change in every other group of four rows).

- The most evenly green rows are the ones with pattern $1, *, 1, *, 1$. All strategies matching this pattern achieve a high winning probability against all possible Brazilian strategies.

After the final, we see that things have changed quite substantially:

- The most prominent impression is that there is far more red: The winning probabilities for the skills estimated from the final only were much reduced for Germany. One possible reason for this is that the performance for the important actions of the German players was not as good as before or the performance of the Brazilian team improved over the prefinal estimates.
- In spite of this big change, the strategic a-posterior recommendation is still to have player 1 be the blocking player who plays *risky* in the field. It is much less important in hindsight, however, what player 2 does in the field.

13 Conclusion

We presented a new concept to answer principle, match-dependent strategic questions in sports games. The question itself is modelled by a strategic MDP (s-MDP) containing only information relevant to the question. If the direct estimation of the s-MDP from the available data is not satisfying, an adequate gameplay MDP (g-MDP) can help to derive valuable s-MDP transition probabilities. The important property of the g-MDP is, that its transition probabilities depend only on the separate skills of the players. With the probabilities derived from the g-MDP, the analytic solution of the s-MDP can be evaluated to answer the strategic question.

We have extensively analysed the Olympic final 2012 by this new method. Some results are encouraging, and some surprising outcomes have yet to be investigated further. Since all estimations of probabilities are quite fragile, a sensitivity analysis of the results is a must. We have presented skill-strategy score cards as a means to graphically present sensitivity information showing the hot spots in parameter space where decisions switch fast.

We think that other strategic questions with a similar structure like the benchmark question in this paper can be treated by following our concept of multi-scale modelling with MDPs. Future research will deal with situations in which skills are dependent on the current score or with the role of variability for success. Moreover, one could try to optimize the detailed meanings of only roughly described strategies like *risky* and *safe* by find the best possible s-g-implementation.

Appendix 1

In this section, we give the details of the proofs for Theorem 2 and Theorem 1. One important observation is the following plausible lemma that holds for both special

cases of s-MDPs. It says that it is no disadvantage for us when we have more points or the opponent has fewer points.

Lemma 1. *The optimal expected reward-to-go* $v^*(x,y,k,\ell)$ *satisfies* $v^*(x,y,k,\ell) \leq v^*(x+1,y,k,\ell)$ *and* $v^*(x,y,k,\ell) \geq v^*(x,y+1,k,\ell)$.

Proof. We prove this by comparing all possible realizations of the game separately. First of all, the outcome of future rallies does not depend on the score. Each winning scenario starting from state (x,y,k,ℓ) corresponds to a winning scenario with identical transitions starting from state $(x+1,y,k,\ell)$ with one stage less that has at least the same probability. Thus, the total winning probability starting from $(x+1,y,k,\ell)$ is no smaller than the one starting in (x,y,k,ℓ). Moreover, each losing scenario starting from state (x,y,k,ℓ) corresponds to a losing scenario with identical transitions starting from state $(x,y+1,k,\ell)$ with one stage less that has at least the same probability. Thus, the total losing probability starting from $(x,y+1,k,\ell)$ is no smaller than the one starting in (x,y,k,ℓ). The claim expresses exactly this in terms of the optimal reward-to-go in the respective states.

In the previous lemma we compared the winning probabilities in states with identical service components. We now explain why the winning probability increases when we win the next point.

Lemma 2. *The optimal expected reward-to-go satisfies* $v^*(x+1,y,P,1) \geq v^*(x,y+1,Q,1)$.

Proof. Team P, in order to win starting at state $(x,y+1,Q,1)$, first has to reach a score of $x+1$ at some point in time. Thus, the main observation, denoted by $(*)$, is that all winning scenarios starting from state $(x,y+1,Q,1)$ pass through exactly one of the states $(x+1,y+z,P,1)$, $z=1,\ldots,21-y$. Let W be the event that P wins, let E be the event that state $(x,y+1,Q,1)$ is passed, and for $z=1,\ldots,21-y$ let E_z be the event that state $(x+1,y+z,P,1)$ is passed. Then we compute:

$$\begin{aligned}
v^*(x,y+1,Q,1) &= \text{Prob}(W|E) \\
&= \sum_{z=1}^{21-y} \text{Prob}(E_z|E)\,\text{Prob}(W|E_z) && \text{(Markov-Property and } *) \\
&= \sum_{z=1}^{21-y} \text{Prob}(E_z|E)\,v^*(x+1,y+z,P,1) \\
&\leq \sum_{z=1}^{21-y} \text{Prob}(E_z|E)\,v^*(x+1,y,P,1) && \text{(Lemma 1 and induction)} \\
&\leq v^*(x+1,y,P,1). && \text{(by } *)
\end{aligned}$$

Thus, an optimal policy is myopic:

Corollary 1. *The policy that always maximizes the probability to win the next point is optimal for the s-MDP for beach volleyball.* □

Appendix 2

This appended section defines the details for the infinite-horizon, stationary g-MDP for a beach volleyball rally that was sketched in Section 6.

Let P and Q be the teams participating in the game. P_1 and P_2 are the players of P; Q_1 and Q_2 are the players of Q. Team P is the team for which we want to choose an optimal playing strategy, whereas team Q is the uncontrolled opposing team. That means, as in the s-MDP, team P is the decision making team, and the behaviour of team Q is part of the system disturbance in the transition probabilities. We have decision epochs $T = \{1,2,3,\ldots\}$, and $t \in T$ is the total number of ball contacts minus the blocking contacts in the rally so far.

A state in the g-MDP is a tuple that contains the players' positions, the ball's position, a counter of the number of contacts, the information which player last contacted the ball, a Boolean variable that indicates the hardness of the last hit, and the designated blocking player of the defending team for the next attack. A general formulation for a state is

$$(pos(P_1), pos(P_2), pos(Q_1), pos(Q_2), pos(ball), counter, lastContact, hard, blocker).$$

The function $pos(\cdot)$ returns the position of a player or the ball. A position on the court is defined on basis of the grid presented in Figure 9.

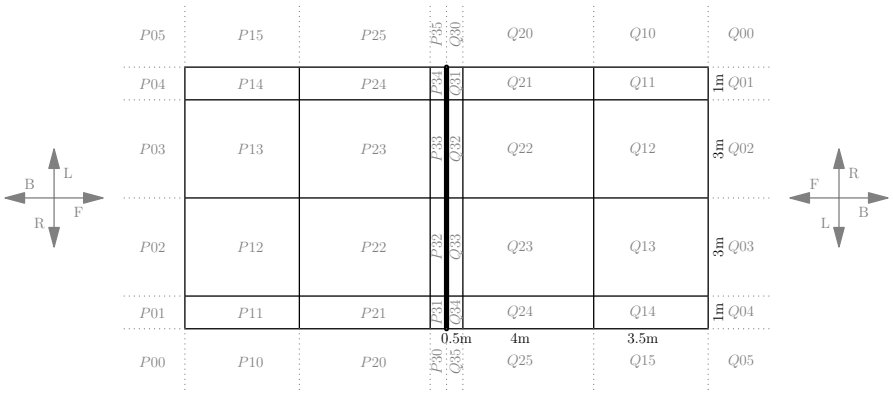

Fig. 9 Court grid

The components *counter* and *lastContact* are needed to implement the three-hits and the double contact rule respectively. The state variable *counter* can take values from the set $\{-1,0,1,2,3\}$. The case "-1' marks a service state. This way it is possible to forbid a blocking action on services. The counter stays -1 if the ball crosses the net after a serve. This helps to distinguish between a reception or defence action. Consequently, if the counter is 0, the ball crossed the net via an attack-hit performed in a field attack. The information which player last contacted the ball is needed to

implement the double-contact fault into the model. The state variable *lastContact* takes values in $\{P_1, P_2, Q_1, Q_2, \emptyset\}$. If the ball has just crossed the net or the state is a serving state, a $\emptyset$-symbol shows that both players are allowed to execute the next hit. The Boolean state variable *hard* indicates the power of the last hit. If $hard = 1$, then the ball has a high speed when reaching the field, else the ball has normal speed. Finally, the state variable *blocker* takes values in $\{P_1, P_2, Q_1, Q_2\}$ and indicates the designated blocking player of the currently defending team. It is necessary to save it in the state since the decision who blocks is made once at the beginning of the opponents attack plan and followed more than one time step. Besides these generic states, the g-MDP contains the absorbing states *point* and *fault*, where *point* and *fault* is denoted from the perspective of team P. The resulting g-MDP has around one billion different states. As an example $(P02, P33, Q12, Q13, P02, -1, \emptyset, 0, -)$ is a typical serving state for team P.

Of course, some of the states occur more often in practice than others. Depending on the current state, there are different actions available to each player. The individual player actions of a player ρ consist of a hit h and a move μ. We distinguish between a one-field and a two-field movement. Also, the direction (f:= forward, fr:= forward-right, ...) of the movement matters. A blocking action belongs to the group of movements since ball possession is not required to perform a block. A blocking action can only be performed if the player is in a field at the net. All possible moves for team P are listed in Table 13. The moves of the players that belong to team Q are defined analogously.

Table 13 Move specification for ρ belonging to team P

Symbol	Specification	Description	Requirements
$\emptyset$	-	Stay	none
m	$f, fr, r, rb, b, bl, l, lf$	Move one field	none
M	f, r, b, l	Move two fields	none
b	-	Block	$pos(\rho) \in \{P31, \ldots, P34\}$, $counter \neq -1$

Depending on the position of a player and on the position of the ball relative to the player, each player has a set of available hits. Sometimes, this set can consist solely of the hit *no hit*. A hit h^{tech}_{field} is defined by a hitting technique *tech* and a target field *field*. Depending on the hit's degree of complexity, there are different requirements such that the hit is allowed in the model. The function $neighbour(field)$ returns a set of all neighbouring fields of *field* according to the grid presented in Figure 9 and the field itself. All hitting techniques with their possible target fields and requirements are listed in Table 14. The hitting techniques for a player of team Q are defined analogously.

There are rules in the model that restrict the possible combinations of a hit with a move to a player action as well as rules that restrict the possible combinations

of two player actions to a team action. Reasons for these restrictions are practical considerations. There are three rules on combining a hit with a movement to a player action. The first one is: If a player makes a real hit, i.e., a hit that is not *no hit*, due to timing reasons only a one-field movement is allowed. The second one is: If a player makes a hit that is performed with a jump, like, e.g., a jump serve, only a one-field movement in forward direction (i.e., towards the net) can follow. The third one is: If the hit requires a movement before executing the hit, no additional movement afterwards is allowed. This is, e.g., the case for a reception that takes place in a neighbouring field of the hitting player. We incorporate one restriction to the combination of player actions: If two player actions are combined to a team action, only one player may make a real hit. Team actions that themselves or whose player actions do not follow these rules are not available in the model – for both teams. Further conceivable restrictions could be easily implemented in the model whenever they only depend on the current state.

Transition probabilities determine the evolution of the system if a certain action in a certain state is chosen. Assume, we know for each player ρ and each hitting technique h^{tech}_{target} the probability

$$p_{succ,\rho}\left(pos(\rho),\, h^{tech}_{target}\right) := \mathbb{P}\big(pos^{t+1}(ball) = target \mid pos^{t}(\rho), h^{tech}_{target}\big),$$

i.e., the probability that the specified target field *target* from ρ's position at time t is met. In the notation used above, the terms $pos(\rho)$ and h^{tech}_{target} show the dependence on the position of the hitting player and the hit he uses. The probability is time-independent. The t on the right-hand side of the last equation is only used to indicate that $pos^{t}(\rho)$ is the position of ρ at time t while $pos^{t+1}(ball)$ is the position of the ball in the subsequent state. Similarly, assume, we know the probability of an execution fault

$$p_{fault,\rho}\left(pos(\rho),\, h^{tech}_{target}\right) := \mathbb{P}\big(s^{t+1} = fault \mid pos^{t}(\rho), h^{tech}_{target}\big)$$

for player ρ using hit h^{tech}_{target} from position $pos(\rho)$. An execution fault includes hits where the ball is not correctly hit such that the referee terminates the rally. For serves and attack-hits an execution fault also includes that the ball is hit into the net.

Furthermore, assume that we know the blocking skills of each player. The parameter $p_{block,\rho}$ denotes the probability that player ρ touches the ball when performing the block b against an adequate attack-hit from the opponent's side of the court. The probability $p_{block,\rho}$ is independent of the skills of the attacking player. There are three possible outcomes of that block. The block can be so strong that it is impossible for the opponent team to get the returned ball, and the blocking team wins the rally. This probability is denoted by $p_{block,\rho,point}$. Also, the block can result in a fault with probability $p_{block,\rho,fault}$. That happens if the ball is blocked into the net and can not be regained or the blocking player touches the net, which is an execution fault. None of the above happens with probability $p_{block,\rho,ok} := p_{block,\rho} - p_{block,\rho,point} - p_{block,\rho,fault}$. This is called an *"ok"-block*, and the ball lands in one random field on the opponents or own court side. We define $p_{no\ block,\rho} := 1 - p_{block,\rho}$ as the probability that the blocking player fails to get his

Table 14 Hit specification for player ρ of team P and ball *ball*; requires always $\rho \neq lastContact$ except the action *no hit* (* if $pos(ball) \neq pos(\rho)$ then no movement afterwards allowed)

tech	*target*	description	Requirements *counter*	position
$\emptyset$	-	no hit	*	none
Serve				
S_F	$Q11-Q24$	float serve	$=-1$	$pos(\rho)=pos(ball)\in P01-P04$
S_J	$Q11-Q24$	jump serve (hard)	$=-1$	$pos(\rho)=pos(ball)\in P01-P04$
Reception				
r	$P11-P34$	receive	$=-1$	$pos(ball)=pos(\rho)$
r_m	$P11-P34$	receive with move	$=-1$	$pos(\rho)\in neighbour(pos(ball))^\star$
Setting				
s	$neighbour(pos(\rho))\setminus(Q,\cdot)$	set	>0	$pos(\rho)=pos(ball)$
Attack-Hit				
F_{SM}	$Q11-Q24$	smash (hard)	>1	$pos(\rho)=pos(ball)$ or $pos(\rho)+m_f=pos(ball)$
F_E	$Q11-Q24$	emergency shot	>1	$pos(\rho)\in neighbour(pos(ball))^\star$
F_P	$Q11-Q34$	planned shot	>0	$pos(\rho)=pos(ball)$
Defence				
d	$P11-P34$	defence	$\neq -1$	$pos(ball)=pos(\rho)$
d_m	$P11-P34$	defence with move	$\neq -1$	$pos(\rho)\in neighbour(pos(ball))^\star$

hands at the ball. In this case, the landing field of the ball is not affected by the block. In total, the blocking probabilities are

$$p_{no\ block,\rho}+\underbrace{p_{block,\rho,point}+p_{block,\rho,fault}+p_{block,\rho,ok}}_{p_{block,\rho}}=1.$$

From all these input probabilities, we generate all transition probabilities in the g-MDP. We explain how the next state evolves from the current state and the played team actions: The next player's position depends only on the current position and the movement the player makes. An allowed movement will always be successful. The crucial component is the next position of the ball. Here, the individual skills of the hitting player enter the model. Assume first, no player of the opposing team is blocking. Then with probability $p_{succ,\rho}\left(pos(\rho),\ h^{tech}_{target}\right)$ the ball's next position will be the desired target field, and with probability $p_{fault,\rho}\left(pos(\rho),\ h^{tech}_{target}\right)$ the hitting player makes an execution fault. The remaining probability

$$1-p_{succ,\rho}\left(pos(\rho),\ h^{tech}_{target}\right)-p_{fault,\rho}\left(pos(\rho),\ h^{tech}_{target}\right)=:p_{dev,\rho}\left(pos(\rho),\ h^{tech}_{target}\right)$$

will be the probability that the ball lands in a neighbouring field of the target field. We assume each neighbouring field is equally probable.

If the hit is an attacking hit to the opponent's court side, then the ball may be blocked. The blocking action must be made from an adequate position[3] such that the block can have an impact. If all preconditions are fulfilled, we first evaluate whether the hit is successful. A hit is *successful* if no execution fault occurs, the ball crosses the net, and approaches the target field or one of its neighbours with the respective probabilities. Given a successful attack, we evaluate in the next step the result of the block. If the blocking player does not touch the ball, then the next position of the ball will not be affected by the block. Otherwise, the outcome of the block is evaluated according to the blocking skill of that player and may be a point, fault or a different position of the ball. This need not automatically mean a point for the attacking team, since the defending team may perform a successful defence action in the next time step. Finally, in case of an execution fault or if the ball is not hit by any player, then the next state will be *point* or *fault*, respectively, from the perspective of team P.

Appendix 3

Table 15 Direct estimation of s-MDP probabilities for team Q

Based on prefinal matches						
strategy	#	$p^+(Q)^{serve}$	$p^-(Q)^{serve}$	#	$p^+(Q)^{field}$	$p^-(Q)^{field}$
risky-risky	19	5%	32%	78	65%	21%
risky-safe	19	5%	32%	27	37%	0%
safe-risky	146	1%	7%	78	65%	21%
safe-safe	146	1%	7%	27	37%	0%

Based on final match						
strategy	#	$p^+(Q)^{serve}$	$p^-(Q)^{serve}$	#	$p^+(Q)^{field}$	$p^-(Q)^{field}$
risky-risky	6	17%	33%	32	69%	19%
risky-safe	6	17%	33%	7	14%	0%
safe-risky	22	0%	9%	32	69%	19%
safe-safe	22	0%	9%	7	14%	0%

References

1. Anbarc, N., Sun, C., Ünver, M.: Designing Fair Tiebreak Mechanisms: The Case of Penalty Shootouts. Tech. rep., Boston College Department of Economics (2015)

[3] The attacking player must be in front of the blocking player. An attack-hit from the last row of the court ($P11-P14$ or $Q11-Q14$) can not be blocked.

2. Chan, T.C.Y., Singal, R.: A Markov Decision Process-based handicap system for tennis. Journal of Quantitative Analysis in Sports **12**(4), 179–189 (2016). DOI 10.1515/jqas-2016-0057
3. Clarke, S.R., Norman, J.M.: Dynamic programming in cricket: Protecting the weaker batsman. Asia Pacific Journal of Operational Research **15**(1) (1998)
4. Clarke, S.R., Norman, J.M.: Optimal challenges in tennis. Journal of the Operational Research Society **63**(12), 1765–1772 (2012)
5. Fédération Internationale de Volleyball: Official Beach Volleyball Rules 2009-2012 . Tech. rep. (2008)
6. Ferrante, M., Fonseca, G.: On the winning probabilities and mean durations of volleyball. Journal of Quantitative Analysis in Sports **10**(2), 91–98 (2014)
7. Heiner, M., Fellingham, G.W., Thomas, C.: Skill importance in women's soccer. Journal of Quantitative Analysis in Sports **0**(0), 287–302 (2014). DOI 10.1515/jqas-2013-0119
8. Hirotsu, N., Wright, M.: Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. Journal of the Operational Research Society **53**(1), 88–96 (2002). DOI 10.1057/palgrave/jors/2601254
9. Hirotsu, N., Wright, M.: Determining the best strategy for changing the configuration of a football team. Journal of the Operational Research Society **54**(8), 878–887 (2003). DOI 10.1057/palgrave.jors.2601591
10. Hoffmeister, S., Rambau, J.: Skill estimates - olympic beach volleyball tournament 2012 (2019). URL `https://epub.uni-bayreuth.de/4150/`
11. Kira, A., Inakawa, K., Fujita, T., Ohori, K.: A dynamic programming algorithm for optimizing baseball strategies **10** (2015)
12. Koch, C., Tilp, M.: Analysis of beach volleyball action sequences of female top athletes **4**(3), 272–283 (2009)
13. Miskin, M.A., Fellingham, G.W., Florence, L.W.: Skill importance in women's volleyball. Journal of Quantitative Analysis in Sports **6**(2) (2010)
14. Mitchell, T.M.: Estimating Probabilities. In: Machine Learning, chap. 2, pp. 1–11 (2017)
15. Nadimpalli, V.K., Hasenbein, J.J.: When to challenge a call in tennis: A Markov decision process approach. Journal of Quantitative Analysis in Sports **9**(3), 229–238 (2013)
16. Norman, J.M.: Dynamic programming in tennis – when to use a fast serve. The Journal of the Operational Research Society **36**(1), pp. 75–77 (1985)
17. Norris, J.: Markov Chains. Cambridge University Press 1997 (1997)
18. Puterman, M.L.: Markov Decision Processes – Discrete Stochastic Dynamic Programming. John Wiley & Sons, New York (2005)
19. Routley, K., Schulte, O.: A Markov Game Model for Valuing Player Actions in Ice Hockey. Uncertainty in Artificial Intelligence (UAI) pp. 782–791 (2015)
20. Schulte, O., Khademi, M., Gholami, S., Zhao, Z., Javan, M., Desaulniers, P.: A Markov Game model for valuing actions, locations, and team performance in ice hockey. Data Mining and Knowledge Discovery (2017). DOI 10.1007/s10618-017-0496-z
21. Terroba, A., Kosters, W., Varona, J., Manresa-Yee, C.S.: Finding optimal strategies in tennis from video sequences. International Journal of Pattern Recognition and Artificial Intelligence **27**(06), 31 pages (2013). DOI 10.1142/S0218001413550100
22. Turocy, T.L.: In Search of the "Last-Ups" Advantage in Baseball: A Game-Theoretic Approach. Journal of Quantitative Analysis in Sports **4**(2) (2008). DOI 10.2202/1559-0410.1104
23. Walker, M., Wooders, J., Amir, R.: Equilibrium play in matches: Binary markov games. Games and Economic Behavior **71**(2), 487 – 502 (2011). DOI http://dx.doi.org/10.1016/j.geb.2010.04.011
24. Webb, J.N.: Game Theory – Decisions, Interaction and Evolution. Springer, London (2007)
25. Wright, M., Hirotsu, N.: A markov chain approach to optimal pinch hitting strategies in a designated hitter rule baseball game. Journal of the Operations Research Society of Japan **46**(3), 353–371 (2003)
26. Wright, M., Hirotsu, N.: The professional foul in football: Tactics and deterrents. Journal of the Operational Research Society **54**(3), 213–221 (2003). DOI 10.1057/palgrave.jors.2601506