DRAFTING AGENT-BASED MODELING INTO BASKETBALL ANALYTICS

Matthew Oldham
Andrew T. Crooks

Department of Computational and Data Sciences
George Mason University
4400 University Drive
Fairfax, VA 22030, USA
moldham@gmu.edu

ABSTRACT

The growth of sports analytics (SA) has raised numerous research topics across a variety of sports, including basketball. Agent-based modeling (ABM) has great potential to assist and inform SA, but to date it has not been utilized. To support the use of ABM in SA, a model of a basketball game, which considers most fundamentals of play, is presented. Additionally, player behavior is partially predicated on assessing the length of a player's shooting streak (testing the "hot-hand" effect) and the consideration a team gives to a streak and their franchise player. The model's output is used to calibrate and validate it against statistics from the National Basketball Association (NBA). Via a set of experiments, the model indicates that an increased belief in the franchise player leads to increased scoring action, but a belief in the hot-hand a minor effect. Thereby, demonstrating the utility of ABM to SA, thus opening a new research field.

Keywords: agent-based modeling, sports analytics, hot-hand effect

1   INTRODUCTION

The burgeoning field of sports analytics (SA) has presented researchers from varying disciplines a rich stream of topics to investigate, as detailed in Section 2. The increased availability of high-fidelity data has fed the growth of SA, with the data utilized in unison with behavioral theories to provide insights into understanding and improving performance. SA, initial success came in baseball, before spreading to other sports, including football, golf, motor racing, and basketball. In basketball, numerous areas of research have arisen including: at the macro level the distribution of scoring activity is a mixture of random walk processes and power-law behavior, and at the individual level players often exhibit cognitive biases. An example of a bias is whether players consider a hot-hand effect and their reaction to such a phenomena. As detailed in Section 2.3, the hot-hand effect refers to the behavior of people who erroneously consider the length of successive success in the past when forming expectations regarding the probability of future success.

As the field of SA expands, it has become apparent that the application of innovative approaches is required. This point is made by Merritt and Clauset (2014) when they indicate that while game theory can help explain dynamics in some circumstances, it is of limited value in assessing complex games. One method that has great potential to assist and inform those engaged in SA, but which to date has not been utilized, is agent-based modeling (ABM). The advantage of creating agent-based models is that it allows researchers to assess, in silico, the micro-level interactions that give rise to verifiable macro outcomes. This outcome is achieved by employing a bottom-up modeling approach, where heterogeneous interactive agents make decisions, adapt and evolve to meet the requirements of their environment.

A model of a basketball game is presented to support the use of ABM in SA, where the fundamentals of play including player and court positions, a shot clock, and shooting performance are included. Also, the model allows for one player in each team to be a franchise player – think Michael Jordan – that the team recognizes as the "main" player. Section 3.1 details the implementation of this player. The players are continuously assessing whether to shoot, pass, or dribble, with the decision-making process involving the player in possession of the ball assessing allocation cues, including shooting performance, and the consideration the team gives to hot-streaks (included to test the hot-hand effect), and the team's franchise player. The probabilistic nature of the model allows for insights into the dynamics of scoring actions to gained and assess whether it matches a random walk. The model captures extensive data which was used to calibrate and validate it against comparable statistics from the National Basketball Association (NBA).

The motivation for implementing an agent-based model to simulate basketball comes from the growing literature (see Section 2.2) that proposes the hypothesis that the game dynamics mimic a complex adaptive system (CAS). The foundation of the approach comes from the fact that an analysis of micro inter-play data shows that the scoring dynamics follow a random walk for a majority of the game, with the exception being the final minute of a quarter, at which time the scoring dynamics produce a power-law distribution (Martín-González et al. 2016). The hot-hand debate encapsulates the question of whether a player experiences an extended period of above-average performance. The hot-hand refers to the fact that people form (erroneously) the belief that if a player is involved in a successful outcome concerning a random event – for example making a basket – the probability of future success in that same activity increases.

An agent-based model is utilized to assess the implications of a team's belief in their franchise player and the hot-hand. Crucial insights emerge from a set of experiments where what the players consider a streak (the threshold), the consideration they gave the streak (the hot-hand effect), and their bias towards favoring their team's franchise player are varied. The main implications are that an increased belief in the franchise player leads to increased scoring action, with the franchise player taking a higher proportion of the shots, but a belief in the hot-hand has at best a minor effect. Results across other indicators such as lead changes and maximum leads showed that as a team's belief in their franchise player increased, the outcome or the course of the game was not affected. The general results support the hypothesis that in general play-scoring activity indeed is a random walk, with only the distribution of scoring among players affected. In the remainder of the paper, Section 2 provides a literature review; in Section 3, the specifications of the model are detailed; and Section 4 presents the results of both the simulations for the baseline model and the various experimental settings. Finally, Section 5 concludes the paper and offers a map for future research.

2 BACKGROUND

2.1 Sports Analytics

SA is experiencing a meteoric rise, with the trend forecast to continue. According to Modor Intelligence (2017), the SA market was valued at USD 83.6 million in 2015 and is forecast to grow to USD 447.2 million by 2020. However, this growth was a long time in the making, with the genesis of SA being traced back over 50 years (Wright 2009). The delayed recognition of the field was partly due to its reputation of being a frivolous pursuit and a hobby industry (Wright 2014) and, as told in *Moneyball* (Lewis 2003), a reluctance of sporting organizations to include SA in their decision-making. The eventual success of baseball teams that implemented SA eventually saw the floodgates open and the sports world is now teeming with opportunities to utilize SA (Fry and Ohlmann 2012). The ability of users to collect data at an ever-increasing rate has undoubtedly aided the rise of SA.

The existing SA work has typically employed Operations Research (OR) methods to research a growing list of research questions (Wright 2014). As such, the focus has been on providing insight into the decision-making process through analytical techniques such as optimization, Markov processes, and regression. However, with the advancements in data collection, and the rise of new analytical methods, the utilization of alternative approaches such as ABM is now feasible. Despite the growing use of SA, according to

Coleman (2012), the field remains fragmented in an academic sense. This situation has occurred because initially the innovative work in the discipline was being carried out within the sports organizations, with the work hidden from the public realm. However, this is changing given the massive amount of publicly available sporting data; for example, Hutchins (2016) discusses the systematic rise of data availability across numerous sports. This paper intends to add to the growing literature of SA by implementing an agent-based model to highlight the potential of calibrating and verifying a model with the easily accessible game data. Section 3.1 explains why the use of ABM is appropriate by highlighting the advantages of the framework, which include the ability to consider individual decision-making, heterogeneity, adaptation, and various temporal and spatial aspects that are absent in the traditional SA techniques. The opportunity exists to extend the ABM process to many other sports as they share many of these characteristics.

2.2 Sports and Complex Adaptive Systems

With the increased availability of data, the spectrum of questions that SA can address ranges from the binary outcome of a win or a loss expectation to micro-level intra-play events. The availability of more detailed data has allowed researchers to consider whether various sports, including basketball, have the characteristics of a CAS. The consideration of sports being a CAS is a newly established field of research (see for example: McGarry et al. (2002), Andriani and McKelvey (2009), Gabel and Redner (2012), Merritt and Clauset (2014), Clauset et al. (2015) and Martín-González et al. (2016)).

The critical ramification for this paper comes from Gabel and Redner (2012), who report that the scoring activity in basketball, with consequences for the number of lead changes, the magnitude of the game's margin, and the persistence of the lead, follows a random walk. A random walk does not imply that skill has no role; instead, the relatively even skill levels at the highest level tend to negate any systematic advantage, leaving a random walk process as the primary determinant (Clauset, Kogan, and Redner 2015). Additional rationale for the random walk process, according to Merritt and Clauset (2014), is that players are short-term focused, as they attempt to maximize their scoring while minimizing the immediate opportunities of their opponents. This dynamic means that there is little strategic play and players are reactive to events as they occur. The random walk process also implies that teams cannot get "hot" – that is, record an extended streak of "making a basket." Section 2.3 discusses this issue in greater detail.

A notable exception to the random walk mechanism is when a quarter is nearing a close. At this point, the distribution of the scoring activity becomes consistent with that predicted by a power-law distribution (Martín-González et al. 2016). Martín-González et al. (2016) provide further support for this paper when they suggest that "basketball is a collaboration-opposition sport where non-linear local interactions among players are reflected in the evolution of the score." The point is that if an environment is found to operate as a CAS, it implies that a small number of rules applied locally are responsible for generating complex, emergent global phenomena. The connection to ABM is that this approach is the ideal method for exploring and discovering the dynamics responsible for the emergent outcomes in a CAS. Section 3.1 details how the implemented agent-based model achieves these goals.

2.3 The Hot-hand Effect

SA played a pivotal role in the 1980s in identifying the cognitive bias coming from an unjustified belief in the law of small numbers (Tversky and Kahneman 1971). The implication related to sports was that players and supporters formed a belief, based on a limited sample, that successive attempts are positively related. Hence, in basketball if a player has made(missed) 2 shots in a row, the belief is that they are more likely to make(miss) their next shot; that is, they have a "hot(cold)-hand" (Gilovich, Vallone, and Tversky 1985). This research kicked off a 20-year debate regarding the existence of the hot-hand effect. The evidence for the vast majority of sports is that the hot-hand is non-existent, and the random process of the game can partially explain any period of enhanced performance (Bar-Eli, Avugos, and Raab 2006).

The importance of the hot-hand research to the implemented model is not so much whether streaks do or do not exist but, if people believe they do, what are the behavioral ramifications? Especially, having formed

the belief that a player is capable of a having a hot-hand, how might a team adapt their behavior if they feel a teammate is on a hot-streak. An example of this comes from the research of Gilovich, Vallone, and Tversky (1985) who suggested that if players believed in the hot-hand, this would lead them to being more likely to pass to a player that has just made a shot and for that player to shoot. Burns (2004) assessed, through a Markov model and computer simulations, the possible behavioral outcomes stemming from a belief in the hot-hand and a franchise player. The basis of the model was to assess how players could best use the allocation cues stemming from a game. That is, if a player makes a successful shot how should other players utilize that information in deciding whether to take the next shot themselves or pass it to the player that just made the basket. A motivation for this paper comes from the desire to agentize the model of Burns (2004), with the benefits arising from players having different skills and responsibilities and considering temporal, environmental, and cognitive factors.

3   MODEL DESIGN AND RATIONALE

3.1 Why Utilize Agent-based Modeling and the Model's Background

This section justifies introducing ABM to SA, highlights how the implemented model leverages the benefits of ABM, and how the model was designed to address the research questions. The defining feature of ABM is that the system is assessed from the perspective of individual agents, thereby employing a bottom-up design approach (Macal 2016). The methodology of ABM involves agents being placed in a dynamic environment and provided with a set of decision-making rules, which they utilize to decide their best course of action after considering their environment and interaction with other agents. Also, the agents are capable of adapting and evolving to meet the requirements of the environment and improve their outcome, thereby mimicking the characteristics of a CAS. The need to undertake such an approach is articulated by Axtell (2000), when he suggests that the key benefits of agent-based models over other simulation techniques are that they may be the only way to uncover the solution structure, test the dynamics of the system, and examine the sensitivity of a model's output against its parameters and assumptions. Agent-based models are also ideally suited to consider temporal and spatial aspects, both of which are vital aspects in a basketball game. The ability to utilize all these aspects (decision-making, spatial, and temporal) is what distinguishes ABM from the more traditional techniques applied in SA.

Martín-González et al. (2016) provide support for pursuing the implementation of an agent-based model basketball game by highlighting the critical characteristics of basketball. These characteristics are that scoring is continuous, to the point of being almost instantaneous, and information related to scoring allows an assessment of the level of self-organization of a team and the actions undertaken by the team (and players). Further, regarding spatial aspects, the game is contained within the court, the probability of scoring is dependent on court position, and player movement patterns are influenced by their allocated position. Concerning the temporal aspects, a game, using NBA standards, is 48 minutes long but, more importantly a team has 24 seconds in which to take a legitimate shot once they gain possession of the ball.

For the interested reader the model (implemented in the 3D version of NetLogo) and a detailed Overview, Design concepts, and Details (ODD) (Grimm et al. 2010) document can be found at https://tinyurl.com/ABMBasketBall. The rationale for utilizing this approach is that it allows for a broader dissemination of the model, and its methodology, into the ABM community, while providing sufficient information for the reader of this paper to appreciate the intent of the model. The document also provides the details of how the model addresses the various aspects of a CAS.

Figure 1 provides a screenshot of the model's graphical user interface (GUI), with the court design, the metrics captured by the model, and the various user-defined parameters visible. The primary feature of the court design is the varying colors. The colors are important because they define whether the player scores 2 or 3 points and, more important, they provide the probability of a player scoring from each position on the court. Readily available NBA data informed the probabilities for scoring from each location. For example, Mala (2015) provides the details of generating a scoring map for a specific player. The court

layout also restricts the movements of the players. The ability to consider spatial aspects and divide the court into separate segments and allocate a specific scoring probability is a demonstration of benefits of implementing an agent-based model and the potential of using real-world data to inform a model.
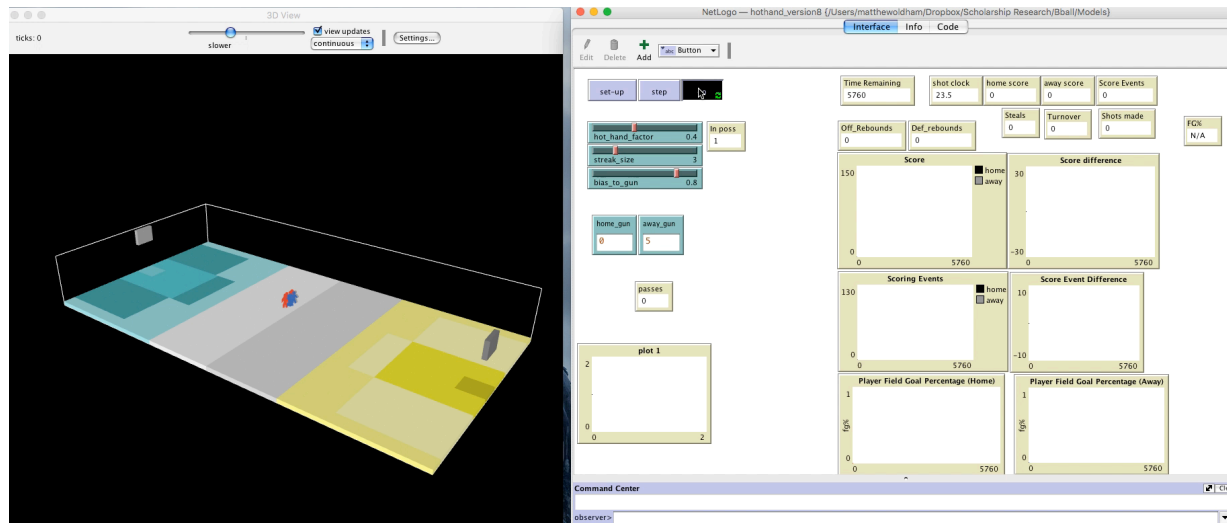

Figure 1: Provided on the left is a screenshot of the 3D NetLogo model. Visible are the various sections of the court which determine the probability of scoring from a given spot. On the right are the different game metrics and the user-defined variables.

To reflect a game of basketball, ten agents are initiated, with five allocated to a home team and five to an away team. With a vital component of the model being to allow for the assessment of the effect of players considering the hot-hand and a franchise player, a player from each team is chosen by the user as the franchise player at initiation, with the status remaining intact for the entirety of the game. This is achieved by the user selecting the appropriate ID number of the chosen players via the model's GUI. To allow for this each player is assigned an identification variable (ID number) which is used to allocate playing positions in a manner consistent with a typical basketball team; that is, there are two guards, two forwards and a center. The relevance of the positions is that they dictate the areas where the offensive players move when they are in possession. Regarding allowable movements, guards have the broadest range, including moving around the 3-point zone, while the centers only move in and around the key. Direct opponents whom the players guard when their team is not in possession are also allocated to each player.

Regarding the relevance of the franchise player, as mentioned in Section 1 and utilized in Burns (2004), a team may for whatever reason tend to favor a superstar player. Within this implementation of the model, the user can adjust the belief a team has in their franchise player. The presence of and belief in a franchise player will affect who the players tend to pass to and the likelihood of the player in possession taking a shot. Section 3.2 (with direct reference to (1) and (2)) details how this occurs. In addition, along with an explanation of the various player parameters, Table 1 explains the franchise player functionality.

3.2 Model Dynamics

The model proceeds such that every half second (the timeframe of a step) players decide their actions. Therefore, to be consistent with an NBA game of 48 minutes a simulation consists of 5760 steps, and an uninterrupted shot clock cycle is 48 steps (24 seconds). The decision to implement half second time steps ensured that the model's scoring action was equivalent, thereby calibrated, to the NBA data. However, future iterations may change this to 0.2 of a second based on the work of Foschi and Leone (2009).

With two teams involved, separate procedures exist. For the defensive team – the team not in possession – the players execute the *defensive* procedure, which means that the defensive players locate their opponents, with these players moving around, and set their heading to match the player before moving to a location on

the defensive side of their opponent – that is, a position between the basket and their opponent. Following a shot, the defensive team, via the *rebound* procedure, turns and moves toward their defensive backboard. The *rebound* procedure, as explained later, determines which team secures the rebound.

Table 1: The Vital Player Variables.

Variable name	Code	Role
Hot-hand factor	HF_i	The variable sets the inclination players (see (1) and (2)) have for considering the hot-hand factor. The range of the variable is 0 (no belief) to 1 (total belief) and is incremented by 0.01.
Streak size threshold	SS_i	This variable sets the benchmark of consecutive shots a player must make for the offensive team, to conclude that the player is on a hot-streak. (1) highlights the role of this variable in that it helps determine whether the player in possession shoots, dribbles, or passes.
Hot-streak tally	HS_{it}	The first part of tracking the performance of each player is this variable. The variable records the number of consecutive shots a player has made. Via (1) the variable helps determine the action of the players.
Franchise player status	F_i	The user selects one player from each team to be the franchise player. For the selected players this variable is set to 1, with it being set to 0 for all others. The variable, via (1) and (2), influences the actions of the player in possession of the ball and the attractiveness of having the ball passed to them.
Belief in the franchise player	BF_i	The variable sets the inclination players (see (1) and (2)) have for favoring/believing in their franchise player. The range of the variable is 0 (no belief) to 1 (total belief) and is incremented by .01.
Field goal percentage (FG%)	$fg\%_{it}$	The second part of tracking player performance is to update the percentage of successful shots made. This is a popular metric used in sports analytics and helps determine whether the player is a viable passing option.
Attractiveness to pass to	$A2P_{it}$	Offensive players update this variable at each step, per (2). The variable is utilized in determining the player's attractiveness as a passing option.

Figure 2 provides the cycle implemented at each step for the offensive team. The first step is to call the *strategy* procedure, which takes half a second off the shot clock and begins the process of determining the appropriate action for the offensive team. The first step is whether the defensive team steals the ball. This step sees a random-float drawn between 0 and 1. If the number is less than .03, then the *turnover* procedure is called, and the possession of the ball is switched to the other team and play resets. The value of .03 was chosen to calibrate this metric to the equivalent level of steals from 17 seasons of NBA game data. A similar approach was utilized across multiple game metrics and outcomes, thus highlighting how a real-world data set can be utilized to inform and calibrate an agent-based model.

If a steal does not occur, the player in possession of the ball will check the shot clock. If there less than 2 seconds on the clock or the player decided to shoot on the previous step the *shoot* procedure is called. The player taking the shot determines the probability of making the shot (SP_{it}), which is based on court location. To determine whether the shot is successful a random-float between 0 and 1 is compared to the pre-determined probability of the player successfully making the shot from that player's current position on the court. If the shot is successful the score is updated (reflecting either a 2 or 3-point shot), along with a host of game statistics, including the hot-streak tally (HS_{it}), and shots-made percentage ($fg\%_{it}$) – which are all vital to determine the future actions of the players. Next, the model calls the *inbound* procedure, which has teams switching possession, with the play process resetting. If the shot misses, then the *rebound*

procedure is called. Here a random-float between 0 and 1 is called, and if the figure is less than 0.24, the current offensive team rebounds the ball. The ratio of offensive rebounds to defensive rebounds comes from NBA game data. In the event of an offensive rebound, the play cycle and the shot clock; alternatively the defensive team rebounds, the teams switch responsibilities, and with play cycle and shot clock resetting.
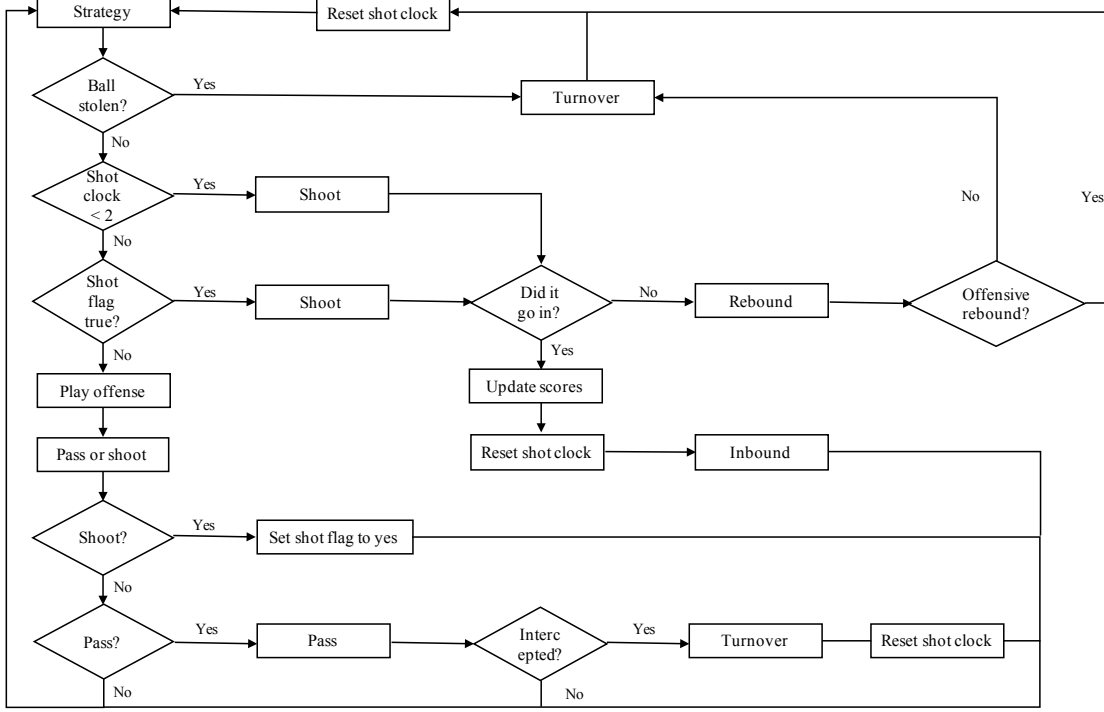

Figure 2: The play cycle of the model.

On the condition that a player has not shot, the team in possession of the ball will enact the *offense* procedure, which moves the offensive players around in their offensive half. To move each player, the players add a random value to the x and y coordinates of their designated offensive positions and move towards their newly determined offensive positions. Next, the player in possession decides to shoot, pass, or dribble. The process starts with the player evaluating the color of the patch upon which the player stands. From the color of the patch, the player gains an insight into the probability of making the shot as per the court design described in Section 3.1. Based on the location the player's SP_{it} value is updated. Next, the player checks the time left on the shot clock and assesses whether to shoot based on

$$SorP_{it} = \log((^1/_{TL_{it} \, * \, .2}) + (SP_{it} * 2))^{(1 + (HS_{it}/SS_i) \, * \, HF_i) + (BF_i * F_i))}. \tag{1}$$

Table 1 provides the definitions for the variables in (1) except for, TL_{it}, which is the time left on the shot clock, and SP_{it}, which is the probability of the player making the shot from their current position. The intention of the formula is that without the influence of the hot-hand or franchise player there would be an increased probability of the player shooting as the shot clock wound down (that is, TL_{it} decreases) and the higher probability of making the shot (given by SP_{it}). The effect of the hot-hand (HF_i) and the extra confidence from being the franchise player (F_i) are compounding factors that increase the probability of the player taking the shot. Noting that the F_i variable is binary (1 for the franchise player or 0), meaning that the BF_i variable is only relevant in the franchise player's decision-making. The influence of the hot-hand effect manifests itself in two ways. The first is the weight that players give the effect through the HF_i (hot-hand factor) variable. The second is through the length of the streak the current player (given through HS_{it}) is on and the threshold that players consider as a streak (SS_i). Therefore, as a player's streak (HS_{it}) grows, the influence of the hot-hand effect grows, with a lower streak threshold amplifying the effect.

After calculating the $SorP_{it}$ value, the player compares it to a shooting threshold, which is drawn from a random-float between 0 and (2 / (1 / TL_{it}). The justification for the approach is that the threshold is a decreasing function concerning the time left on the shot clock. The threshold was tuned to calibrate the model, so the number of field goal attempts was consistent with a typical NBA game. If the $SorP_{it}$ value is greater than the threshold, the player will set their *shoot?* flag to true, with the player taking the shot at the next step. Alternatively, the player will continue to dribble or pass, as described next.

If the player has decided not to take the shot, the *pass* procedure is called. This procedure has the player in possession assess the position of the other players in the team. Having established the legitimate passing options, the player will pass to the player with the highest attractiveness. The attractiveness of a player as a passing option is given by (2), with this value updated at the start of the play cycle. The formula's intention is that the attractiveness of a player as a passing option is dependent on each player's performance, and on the team's trust in their franchise player and belief in the hot-hand effect. By way of additional definitions, $team\%_{it}$ is the percentage of points the player being assessed has scored for the team, thereby providing a measure of performance. The influence of the hot-hand effect ($HS_{it} * HF_i$) and the belief in the franchise player ($BF_i * F_i$) is linear, with the effect of the franchise player belief being binary, as in

$$A2P_{it} = fg\%_{it} + team\%_{it} + (HS_{it} * HF_i) + (BF_i * F_i). \qquad (2)$$

Before effecting the pass, the model updates the *intercept_options* list. This list includes the opponent of the current dribbler and the player who is about to receive the pass, with one of these players chosen as the possible interceptor. A pass interception occurs if a random-float between 0 and 1 is less than .05 (again calibrated to the average ratio of passes to turnovers in the NBA). A successful pass has the receiver's status updated to reflect the player has possession of the ball and the play cycle will continue; alternatively the turnover procedure will be called.

3.3 Calibration and Validation

To assist in the calibration and validation processes for the model, 17 seasons (2000 to 2016) of individual game data was scraped from NBA.com. The data collected included; game scores, winning margins, field goal attempts, the percentage of field goals made, rebounds, steals, and turnovers. From the NBA game data, density functions were calculated to aid calibrating certain aspects of the model, as discussed in Section 3.2. Per Figure 3, the density functions are also compared to the density functions derived from the data generated from the parameter sweep of the model, a step undertaken to determine a level of validation for the model. The a priori expectation is that if NBA basketball follows a random walk process, the results from the model should provide a good first approximation given its stochastic nature. The parameter sweep involved running each parameter setting, as outlined in Table 2, 100 times, resulting in the equivalent of 6,400 games. The main comment regarding Figure 3 is that the model was able to be specified and calibrated such that it was able to produce outcomes in line with real-world data, noting no formal statistical tests were undertaken to support this statement. Therefore, Figure 3 provides a level of validation supporting the proposition that the model does provide a good approximation of the random walk in NBA basketball. The model's best match is the game ending margin (Figure 3b), with the relevance being the connection to Clauset et al. (2015) research on safe margins and lead changes. The one critique may be that the model has a fatter tail, which is explainable by NBA teams resting players once they feel a game is safe. Hence, the final scores are not genuinely reflective of skill, a factor the model does not consider.

Regarding the actual scores of the games, Figure 3a illustrates that the NBA scores, in general, tend to be higher. A feasible explanation for the differences is that the model does not accommodate foul shots or bonus shots. Also, the prevalence of 3-point shots needs investigating. Figure 3c illustrates the field-goal-made percentage (FG%) are comparable for the model and NBA data, noting the peaks of both distributions are in the late 40% range, with the NBA data showing a wider variation. Possible explanations for the variation, which will form the basis of future work, include: there is no defensive pressure applied in the model; there are no substitutions or fatigue, so the skill level of the game does not diminish when bench players enter the game; and the probability of hitting a shot is the same for all players. From comparing the

rebound densities, in Figure 3d it is seen that the model produces, on average, more rebounds, and the distribution is more peaked. This result comes from the fact that despite the FG% being comparable, the model sees more shots being taken in a game (a density plot of shots taken has similar characteristics to Figure 3d). One explanation is that with no fouls or out-of-bounds the game clock does not stop or reset; therefore, more scoring attempts are made. Another explanation is that the game maintains a consistent tempo as players do not "milk" the clock at various stages of the game.
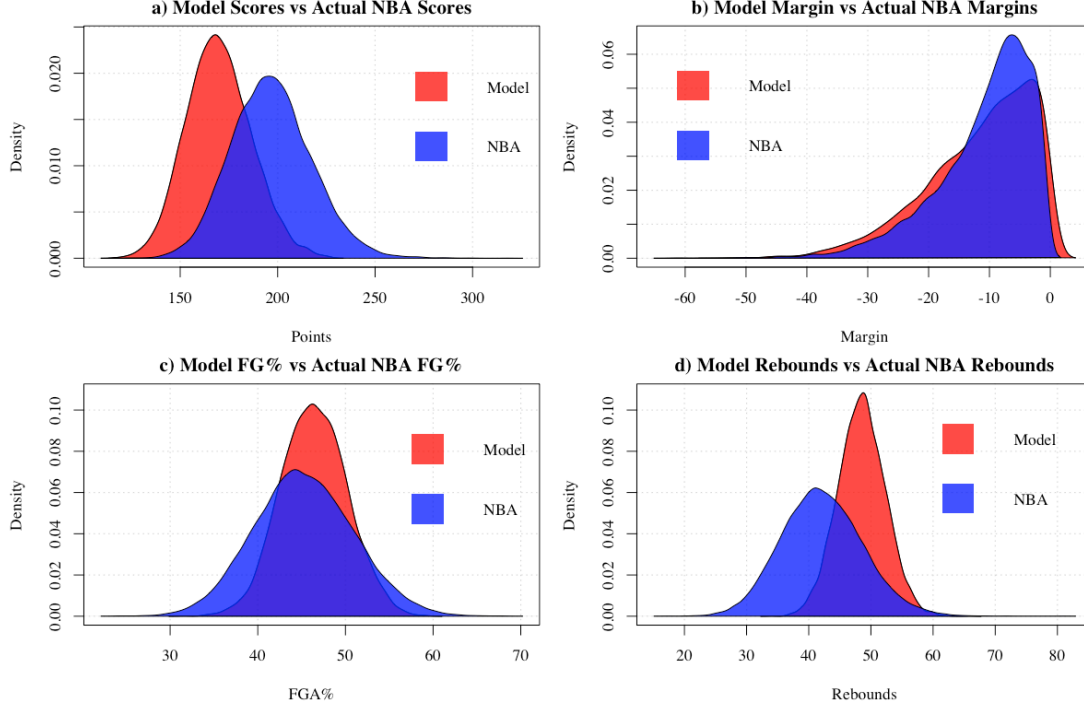

Figure 3: Distribution of the varying NBA game metrics compared to the model.

4   RESULTS

This section summarizes the results of the various experiments designed to test the influence of the players' beliefs regarding the importance of the franchise player and the hot-hand effect. The main output used to assess the effects is the total game score. Table 2 provides the range of settings used in the parameter sweep.

Table 2: Experimental Settings.

Variable	Settings
IDs for the franchise players	0 and 5
Belief in the franchise player (BF_i)	0, 0.33, 0.66, 1
Hot-hand factor (HF_i)	0, 0.33, 0.66, 1
Streak size thresholds (SS_i)	2, 3, 4 ,5

Figure 4a provides boxplots for the total game scores generated by the various interactions of the two belief variables. For ease of interpretation the boxplots are color-coded, with color distinguishing players' bias towards favoring their franchise player (BG_i), with the subplots within each color block distinguishing the various beliefs in the hot-hand effect (HF_i). The interpretation of the boxplots is that the x-axis marks the

interaction between the players' belief in the hot-hand and their franchise player; that is, a label of $0.33 \cdot 0.00$ is the result of combining a 0.33 belief in the hot-hand and no belief (0.00) in the franchise player. Remembering, the franchise player is the one player perceived by their teammates to have superior skills (see Table 1). The interpretation of the boxplots is that an increase in the belief in the franchise player has a more consistent effect in terms of increasing the scoring activity in the game. Alternatively, the hot-hand effect appears to be limited in its influence, with its effect only material with a belief level of 0.33.
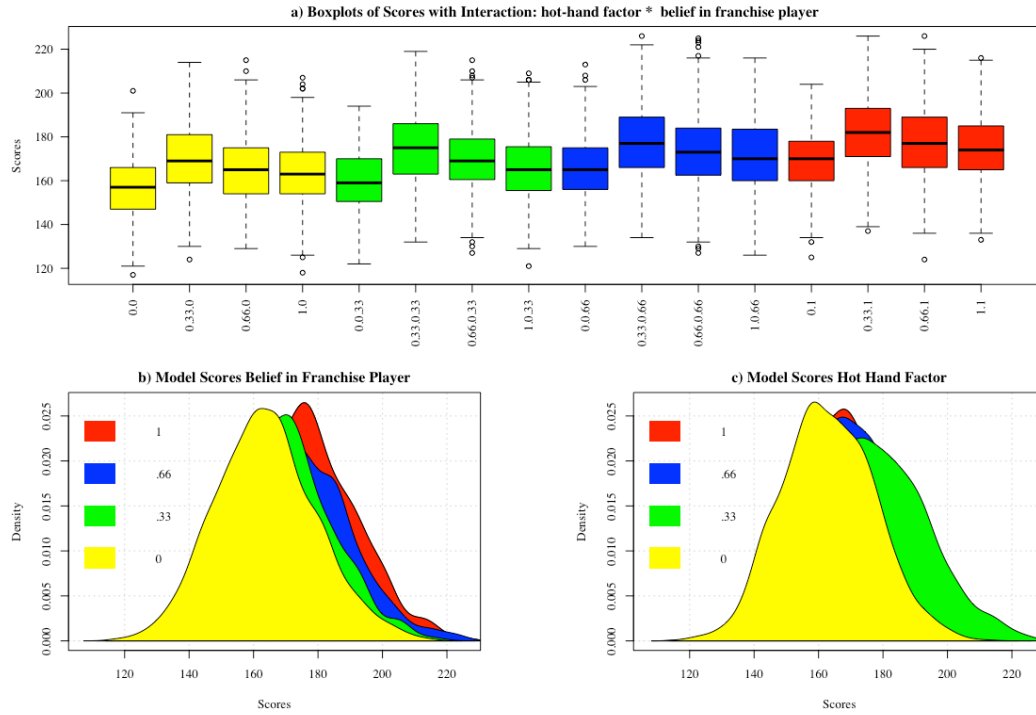

Figure 1: Results of the varying the belief in the hot-hand (HF_i) and the franchise player (BF_i).

Figure 4b and 4c are density plots showing the distribution of scores resulting from varying the various beliefs, thereby isolating the effects of each variable. The plots show that an increasing belief in the franchise player has a more consistent effect on scoring because the peaks of the scoring distributions move consistently to the right as the belief increases. Alternatively, by increasing the hot-hand belief, the results are less consistent. A vital point, per the previous analysis, is that a belief value in the hot-hand of 0.33 produces the most significant increase in scoring, a matter requiring further investigation.

The interpretation from the initial results is that if a team has a belief in their franchise player, this is more likely to lead to higher scoring than a belief in the hot-hand effect. This result is intuitive, as the characteristic of being a franchise player is binary, as a player is either "the" player or not. Therefore, if the team believes in their franchise player, the effect will manifest itself in the franchise player receiving more passes and that player having greater confidence to take more shots. This result is in line with what is seen in the NBA, as it is not difficult to find examples of key franchise players taking a bulk of the scoring responsibility. For the hot-hand factor, its effect is evident but hampered by the inability of a player to consistently generate a streak of material length, as discussed next.

Figure 5 provides boxplots for the total game scores divided by the interaction of the length of what players consider a streak (SS_i) and the weight they give to the streak (HF_i). The x-axis marks the interaction between the streak size and the belief in the hot-hand; that is, a label of $5.0 \cdot 0.33$ is the result of combining a streak threshold of 5 and a belief in the hot-hand of 0.33. The implication is that a lower threshold has a positive effect on scoring. This finding is understandable because the probability of seeing a streak of 2 or 3 is higher than the other settings; therefore, the hot-hand belief becomes influential. The outcome that a

hot-hand belief of 0.33 has the most significant influence is again evident in Figure 5, but only when players have a lower threshold for what they consider a streak.
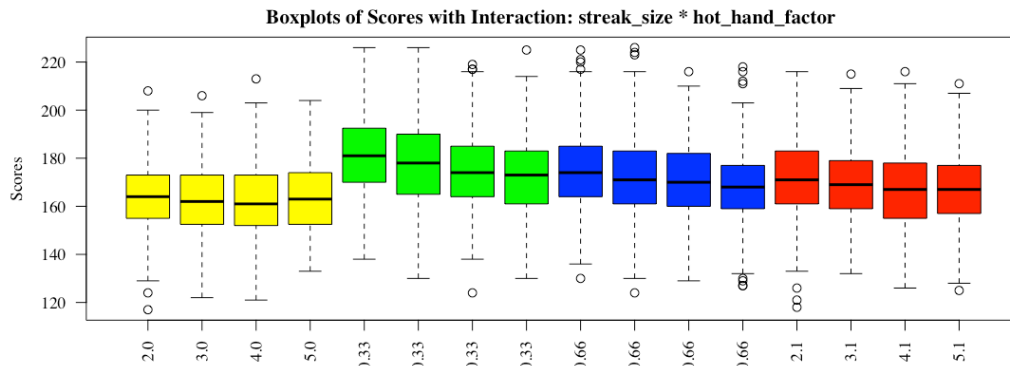

Figure 5: Results of the varying the influence of streak size (SS_i) and the belief in the hot-hand (HF_i).

The results presented in this section validate the expanded use of ABM in SA. The justification comes from adequately matching high-level data and giving meaningful insights into the effects of various cognitive biases. With the utility of utilizing ABMs for SA established, the challenge exists to integrate even more actual game data into the model design and validation process.

5   SUMMARY AND CONCLUSIONS

This paper delivers a proof-of-concept model that ambitiously combined three fields of research: SA, ABM, and cognitive biases. With an encouraging level of validation against actual NBA game data, it suggests that further research is warranted. The obvious candidates for extending the research include: calibrate player movements with actual data; make the players' beliefs dynamic; instigate consequential defensive pressure; include fouls; substitute inferior players into the game; and introduce the effects of player fatigue. These extensions could extend the application of the model to roster optimization and even game strategy. Finally, the model could allow players to consider the time left in the quarters/game and the game score in an attempt to replicate the varying scoring activity dynamics as mentioned in Section 2.2. Concerning the paper's contribution to the hot-hand debate, the model provided evidence that through a simple stochastic process a player is capable of recording an extended period of success, thereby indicating that people should not believe a streak is unbelievable or unusual. The more significant contribution from the model is how the beliefs of players affect the system. Here it was seen that confidence in the ability of a franchise player is a vital component in the scoring output of the game. In closing, it is apparent that the application of ABM to understanding sporting outcomes is a worthwhile endeavor. This result comes from the ability of agent-based models to identify the micro-interaction of agents responsible for generating system level outcomes.

REFERENCES

Andriani, P., and B. McKelvey. 2009. "Perspective - From Gaussian to Paretian Thinking: Causes and Implications of Power Laws in Organizations." *Organization Science* 20 (6): 1053–71.

Axtell, R. 2000. "Why Agents?: On the Varied Motivations for Agent Computing in the Social Sciences." *Center on Social and Economic Dynamics Brookings Institution*.

Bar-Eli, M, S. Avugos, and M. Raab. 2006. "Twenty Years of 'Hot Hand' Research: Review and Critique." *Psychology of Sport and Exercise* 7 (6): 525–53.

Burns, B. D. 2004. "Heuristics as Beliefs and as Behaviors: The Adaptiveness of the 'Hot Hand.'" *Cognitive Psychology* 48 (3): 295–331.

Clauset, A., M. Kogan, and S. Redner. 2015. "Safe Leads and Lead Changes in Competitive Team Sports." *Physical Review E* 91 (6).

Coleman, B., 2012. "Identifying the 'Players' in Sports Analytics Research." *Interfaces* 42 (2): 109–18.

Foschi, Renato, and Matteo Leone. 2009. "Galileo, Measurement of the Velocity of Light, and the Reaction Times." *Perception* 38 (8): 1251–59.

Fry, M.J., and J.W. Ohlmann. 2012. "Introduction to the Special Issue on Analytics in Sports, Part I: General Sports Applications." *Interfaces* 42 (2): 105–8.

Gabel, A., and S. Redner. 2012. "Random Walk Picture of Basketball Scoring." *Journal of Quantitative Analysis in Sports* 8 (1).

Gilovich, T., R. Vallone, and A. Tversky. 1985. "The Hot Hand in Basketball: On the Misperception of Random Sequences." *Cognitive Psychology* 17 (3): 295–314.

Grimm, Volker, Uta Berger, Donald L. DeAngelis, J. Gary Polhill, Jarl Giske, and Steven F. Railsback. 2010. "The ODD Protocol: A Review and First Update." *Ecological Modelling* 221 (23): 2760–68.

Hutchins, Brett. 2016. "Tales of the Digital Sublime: Tracing the Relationship between Big Data and Professional Sport." *Convergence: The International Journal of Research into New Media Technologies* 22 (5): 494–509.

Lewis, M. 2003. *Moneyball: The Art of Winning an Unfair Game*. 1st ed. New York: W. W. Norton.

Macal, C.M. 2016. "Everything You Need to Know about Agent-Based Modelling and Simulation." *Journal of Simulation* 10 (2): 144–56.

Mala, E. 2015. "How to Create NBA Shot Charts in R." 2015. https://thedatagame.com.au/2015/09/27/how-to-create-nba-shot-charts-in-r/.

Martín-González, J.M., Y. Guerra, J.M. García-Manso, E. Arriaza, and T. Valverde-Estévez. 2016. "The Poisson Model Limits in NBA Basketball: Complexity in Team Sports." *Physica A: Statistical Mechanics and Its Applications* 464 (December): 182–90.

McGarry, T, D.I. Anderson, S.A. Wallace, M.D. Hughes, and I.M. Franks. 2002. "Sport Competition as a Dynamical Self-Organizing System." *Journal of Sports Sciences* 20 (10): 771–81.

Merritt, S., and A. Clauset. 2014. "Scoring Dynamics across Professional Team Sports: Tempo, Balance and Predictability." *EPJ Data Science* 3 (1).

Modor Intelligence. 2017. "Global Sports Analytics Market - Growth, Industry Analysis, Trends, Forecast (2017-2022)."

Tversky, A., and D. Kahneman. 1971. "Belief in the Law of Small Numbers." *Psychological Bulletin* 76 (2): 105–10.

Wright, M. 2009. "50 Years of OR in Sport." *Journal of the Operational Research Society* 60 (1): 161–68.

Wright, M. 2014. "OR Analysis of Sporting Rules – A Survey." *European Journal of Operational Research* 232 (1): 1–8.

AUTHOR BIOGRAPHIES

MATTHEW OLDHAM is a Ph.D. candidate in Computational Social Science (CSS) program at George Mason (GMU). He holds a MAIS from GMU and a Bachelor of Economics from the University of Tasmania. His research interests are the application of ABM, network analysis, and machine learning to financial markets. His email address and website is moldham@gmu.edu and www.aussiecas.com.

ANDREW CROOKS is an Associate Professor of Computational Social Science within the Department of Computational and Data Sciences at GMU. His research interests relate to exploring and understanding of the natural and socio economic environments using geographical information systems (GIS) and ABM methodologies. His email address and website are acrooks2@gmu.edu and http://gisagents.org.