

# A Multiresolution Stochastic Process Model for Predicting Basketball Possession Outcomes

Daniel Cervone<sup>1</sup>, Alex D’Amour<sup>1</sup>, Luke Bornn<sup>1</sup>, and Kirk Goldsberry<sup>2</sup>

<sup>1</sup>Department of Statistics, Harvard University, Cambridge, MA 02138

<sup>2</sup>Institute for Quantitative Social Science, Harvard University, Cambridge, MA 02138

October 3, 2018

## Abstract

**Abstract** Basketball games evolve continuously in space and time as players constantly interact with their teammates, the opposing team, and the ball. However, current analyses of basketball outcomes rely on discretized summaries of the game that reduce such interactions to tallies of points, assists, and similar events. In this paper, we propose a framework for using optical player tracking data to estimate, in real time, the expected number of points obtained by the end of a possession. This quantity, called *expected possession value* (EPV), derives from a stochastic process model for the evolution of a basketball possession; we model this process at multiple levels of resolution, differentiating between continuous, infinitesimal movements of players, and discrete events such as shot attempts and turnovers. Transition kernels are estimated using hierarchical spatiotemporal models that share information across players while remaining computationally tractable on very large data sets. In addition to estimating EPV, these models reveal novel insights on players’ decision-making tendencies as a function of their spatial strategy.

**KEYWORDS:** Optical tracking data, spatiotemporal model, competing risks, Gaussian process.

## 1. INTRODUCTION

Basketball is a fast-paced sport, free-flowing in both space and time, in which players’ actions and decisions continuously impact their teams’ prospective game outcomes. Team owners, general managers, coaches, and fans all seek to quantify and evaluate players’ contributions to their team’s success. However, current statistical models for player evaluation such as “Player Efficiency Rating” (Hollinger 2005) and “Adjusted Plus/Minus” (Omidiran 2011) rely on highly reductive summary statistics of basketball games such as points scored, rebounds, steals, assists—the so-called “box score” summary of the game. Such models reflect the fact that up until very recently, data on basketball games were only available in this low level

of resolution. Thus previous statistical analyses of basketball performance have overlooked many of the high-resolution motifs—events not measurable by such aggregate statistics—that characterize basketball strategy. For instance, traditional analyses cannot estimate the value of a clever move that fools the defender, or the regret of skipping an open shot in favor of passing to a heavily defended teammate. The advent of player tracking data in the NBA, coupled with the appropriate inferential framework, has provided an opportunity to fill this gap.

In 2013 the National Basketball Association (NBA), in partnership with data provider STATS LLC, installed optical tracking systems in the arenas of all 30 teams in the league. The systems track the exact two-dimensional locations of every player on the court (as well as the three-dimensional location of the ball) at a resolution of 25Hz, yielding over 1 billion space-time observations over the course of a full season.

In this paper, we present a framework for modeling NBA tracking data that targets inferences at the resolution of this exciting new data. Specifically, we estimate the expected number of points the offense will score on a particular possession conditional on that possession’s evolution up to time  $t$ . We term this quantity *expected possession value* (EPV). EPV acts like the stock ticker of an NBA possession in providing an instantaneous summary of the possession’s value given all available information. Ideally, EPV mirrors the intuition of coaches and basketball strategists by attaching value to specific spatial positionings of players and personnel configurations. For instance, EPV may be high when a good shooter has an open look at the basket, but low when the ballcarrier is heavily defended far from the basket without any clear passing options. We may also, for example, see EPV as roughly constant while players are passing the ball around the perimeter, far from the basket, but then spike upwards as a player drives through an opening in the defense towards the basket. By monitoring how the EPV curves for various possessions respond to player decisions, analysts can evaluate players and strategies in real time, continuously throughout the entire course of any possession.

Our paper focuses specifically on modeling and calculating EPV curves by viewing basketball possessions as realizations of endpoint-valued stochastic processes; that is, we assume possessions evolve probabilistically over space and time until reaching some endpoint (e.g., a made basket, or a turnover) with an observable point value. Such frameworks have been used for in-game points/run expectancy models in other sports—for instance, baseball (Bukiet, Harold & Palacios 1997) and football (Burke 2010)—yet ours is the first treatment of a process that is continuous in both space and time. To handle this additional complexity, we introduce the idea of multiresolution transitions, which distinguish between the continuous evolution of all players’ positions (microtransitions) and motifs or events that unfold over longer time scales (macrotransitions), such as passes and shot attempts. We show that multiresolution transitions crucially allow conditioning that is both interpretable and computationally tractable.

While our methodology is motivated by basketball, we believe that this research can serve as an informative case study for analysts working in other application areas where continuous monitoring data are becoming widespread, including traffic monitoring (Ihler, Hutchins & Smyth 2006), surveillance, and digital marketing (Shao & Li 2011), as well as other sports such as soccer and hockey (Thomas, Ventura, Jensen & Ma 2013). As such,

we treat our particular approach to the player tracking problem in basketball with as much generality as possible, and hope that aspects of our methodology may also find use in these other contexts.

Section 2 formally defines EPV within the context of a stochastic process for basketball. Section 3 introduces multiresolution transitions and discusses the assumptions and conditioning statements that make EPV calculations tractable as averages over future paths of a stochastic process. The models for macro- and microtransitions are discussed in Sections 4 and 5, respectively, and represent the inferential component of our model, as transition probabilities rely on players’ decision-making tendencies in various spatial and situational circumstances. Section 6 discusses Monte Carlo computation for EPV curves, given parameters for the multiresolution transition models, with some results from actual NBA possessions highlighted in Section 7. Directions for further work are discussed in Section 8.

The core of our paper is a high-level overview of the EPV estimation problem and our model, focusing on the stochastic process, multiresolution approach to EPV, and highlighting the types of inferences EPV and the associated multiresolution transition models provide basketball analysts. A substantial appendix follows the discussion section, in which we discuss specific details of our implementation for EPV estimation, including choices of prior distributions for model parameters, and computational methods for obtaining approximate inferences from such a large, high dimensional data set. More than just a guide to reproducing our results, the appendix highlights novel contributions to estimation problems in spatiotemporal data. For instance, our macrotransition model employs a family of prior distributions that shares information across space as well as across players, representing hierarchical Gaussian processes in a computationally tractable parameterization.

## 2. EXPECTED POSSESSION VALUE

Let  $\Omega$  represent the space of all possible basketball possessions in full detail, with  $\omega \in \Omega$  describing the full path of a particular possession. Every basketball possession that we consider here results in 0, 2, or 3 points scored for the offense, denoted  $X(\omega) \in \{0, 2, 3\}$ . It is possible for the offense to score exactly 1 or 4 points as well if a foul occurs and free throws are made, but we exclude fouls and free throws from  $\Omega$  due to limitations in our data. For any possession path  $\omega$ , we denote by  $Z(\omega)$  the optical tracking timeseries generated by this possession so that  $Z_t(\omega) \in \mathcal{Z}$ ,  $t > 0$ , is a “snapshot” of the tracking data exactly  $t$  seconds from the start of the possession ( $t = 0$ ).  $\mathcal{Z}$  is a high dimensional space that includes  $(x, y)$  coordinates for all 10 players on the court,  $(x, y, z)$  coordinates for the ball, summary information such as which players are on the court and what the game situation is (game location, score, time remaining, etc.), and event annotations that are observable in real time, such as a turnover occurring, a pass, or a shot being attempted and the result of that attempt.

This intuitive view of  $\Omega$  as a sample space of possession paths provides the formalism for defining EPV in probabilistic terms. We define  $Z(\omega)$  to be a stochastic process, and likewise, define  $Z_t(\omega)$  for each  $t > 0$  as a random variable in  $\mathcal{Z}$ .  $Z(\omega)$  provides the natural filtration  $\mathcal{F}_t^{(Z)} = \sigma(\{Z_s^{-1} : 0 \leq s \leq t\})$ , which intuitively represents all information available from the optical tracking data for the first  $t$  seconds of a possession. Because the point outcome of a possession ( $X$ ) is apparent from observing  $Z(\omega)$  for a sufficiently long time,  $X$  is  $\mathcal{F}_\infty^{(Z)}$ -

measurable, and we can define EPV as the expected value of the number of points scored for the possession ( $X$ ) given all available data up to time  $t$  ( $\mathcal{F}_t^{(Z)}$ ):

**Definition** The *expected possession value*, or EPV, at time  $t \geq 0$  during a possession is  $\nu_t = \mathbb{E}[X | \mathcal{F}_t^{(Z)}]$ .

**Remark** Except when introducing new summaries of the possession sample space  $\Omega$ , we will omit the dependence on  $\omega$  when writing function- or scalar-valued random variables, e.g.,  $Z$  and  $Z_t$  instead of  $Z(\omega)$  and  $Z_t(\omega)$ .

## 2.1 Possession Case Study

To illustrate the behavior of EPV, we consider the estimated EPV curve from a specific Miami Heat possession against the Brooklyn Nets from the second quarter of a game on November 1, 2013. This possession was chosen arbitrarily among those during which LeBron James (widely considered the best NBA player as of 2014) handles the ball. This is presented here in order to describe and motivate the object of our estimation; the methodology itself will be discussed in the sections that follow.

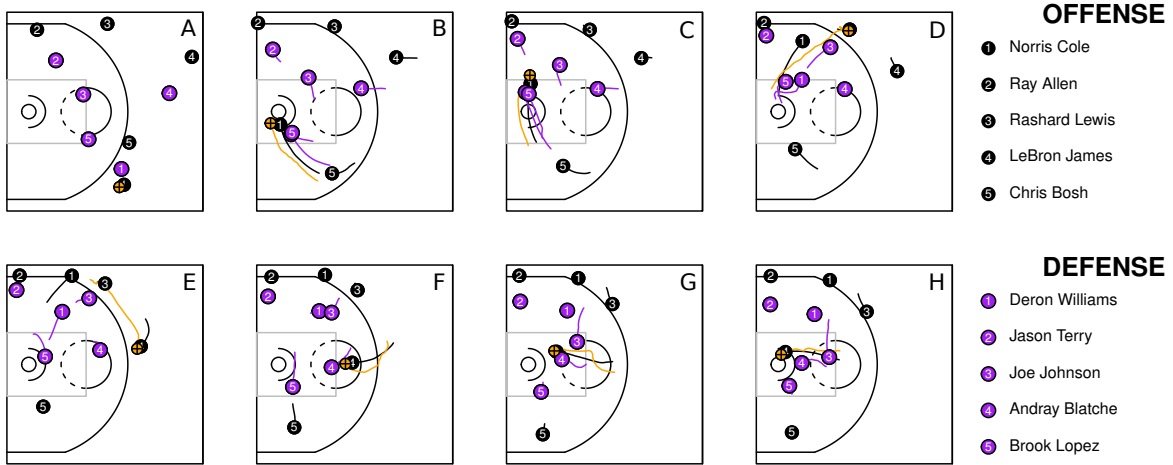


Figure 1: Miami Heat possession against Brooklyn Nets. Norris Cole wanders into the perimeter (A) before driving toward the basket (B). Instead of taking the shot, he runs underneath the basket (C) and eventually passes to Rashard Lewis (D), who promptly passes to LeBron James (E). After entering the perimeter (F), James slips behind the defense (G) and scores an easy layup (H).

In this particular possession, diagrammed in Figure 1, point guard Norris Cole begins with possession of the ball crossing the halfcourt line (panel A). After waiting for his teammates to arrive in the offensive half of the court, Cole wanders gradually into the perimeter (inside the three point line), before attacking the basket through the left post. He draws two defenders, and while he appears to beat them to the basket (B), instead of attempting a layup he runs underneath the basket through to the right post (C). He is still being double teamed and at this point passes to Rashard Lewis (D), who is standing in the right wing three position and being defended by Joe Johnson. As Johnson closes, Lewis passes to LeBron James, who

is standing about 6 feet beyond the three point line and drawing the attention of Andray Blatche (E). James wanders slowly into the perimeter (F), until just behind the free throw line, at which point he breaks towards the basket. His rapid acceleration (G) splits the defense—Joe Johnson had also begun defending James as he entered the perimeter—and gains him a clear lane to the basket. He successfully finishes with a layup (H), providing the Heat two points.

Plotting the EPV curve for this possession (Figure 2), we see several moments when the expected point yield of the possession, given its history, changes dramatically. Beginning around 0.99, the EPV first rises as Cole drives toward the basket, starting around 5 seconds into the possession. It continues rising until peaking at around 1.34 when Cole is right in front of the basket. As Cole dribbles past the basket (and his defenders continue pursuit), however, EPV falls rapidly, bottoming out at 0.77 before “resetting” to 1.00 with the pass to Rashard Lewis. The EPV increases slightly to 1.03 when the ball is then passed to James. As EPV is sensitive to small changes in players’ exact locations, we see EPV rise slightly as James approaches the three point line and then dip slightly as he crosses it. Shortly afterwards, EPV rises suddenly as James breaks towards the basket, eluding the defense, and continues rising until he is beneath the basket, when an attempted layup boosts the EPV from 1.52 to 1.62.

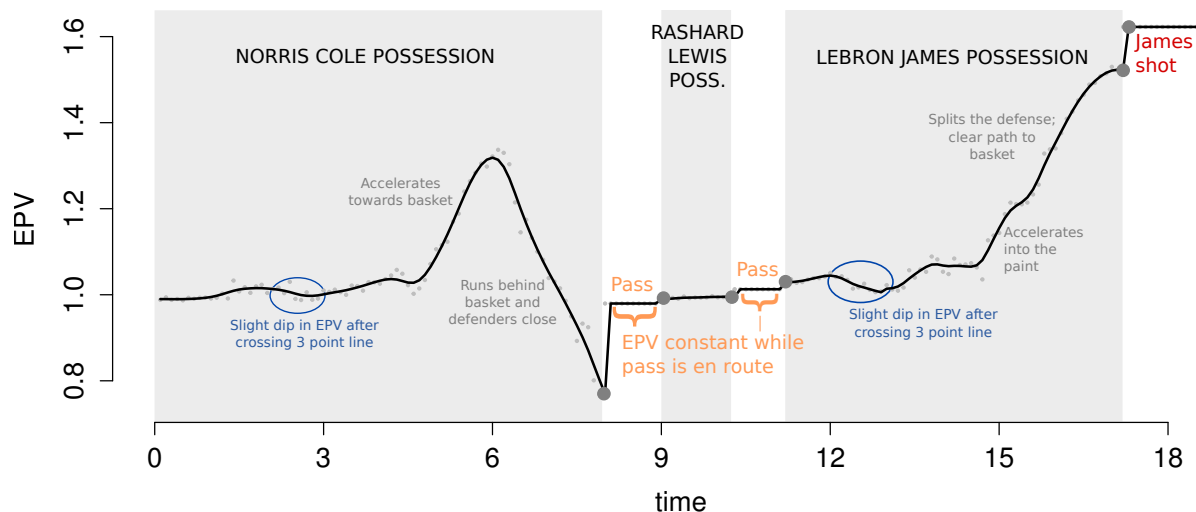


Figure 2: Estimated EPV over time for the possession shown in Figure 1. Changes in EPV are induced by changes in players’ locations and dynamics of motion; macrotransitions such as passes and shot attempts produce immediate, sometimes rapid changes in EPV. The black line slightly smooths EPV evaluations at each time point (gray dots), which are subject to Monte Carlo error.

We will revisit this example possession in more detail in Section 7. For now, it serves to highlight the behavior and potential applications of EPV. Any point on the curve in Figure 2 provides an unbiased estimate of the possession’s eventual point total as a function of its entire history, including but not limited to the identity and precise location of the ballcarrier, and those of his teammates and the defense. We see EPV rise and fall—at times dramatically—as players move the ball through the court, pass, attempt shots, and gain or lose separation from the defense.

## 2.2 Stochastic Consistency

EPV is a theoretical quantity associated with the true distribution of possession paths  $Z$ . As with all statistical estimands, the definition of EPV does not restrict the method that an investigator can use to estimate it. Indeed, because EPV is simply a conditional expectation, a number of marginal regression or classification methods that map features from  $\mathcal{F}_t^{(Z)}$  to the outcome space (either  $[0, 3]$  or  $\{0, 2, 3\}$ ) can be tempting options. While our data do not constitute the independent input/output pairs characteristic of regression (each possession outcome  $X$  has a series of inputs  $Z_t$ ), a properly specified regression model would nevertheless consistently estimate  $\nu_t$  as a function of features of  $\mathcal{F}_t^{(Z)}$ .

Regression estimates, however, lack the stochastic consistency inherent to the probabilistic formulation of EPV.  $\nu$  is an instance of the “Doob martingale” with respect to the filtration  $\mathcal{F}^{(Z)}$ —that is, a sequence of conditional expectations of the same end quantity  $X$  taken with respect to increasing elements of the filtration  $\mathcal{F}^{(Z)}$ . Thus, EPV evaluated at a particular time  $t$  can be represented as the expected EPV evaluated at a later time  $s > t$ :  $\mathbb{E}[\nu_s | \mathcal{F}_t^{(Z)}] = \nu_t$ . An important consequence of this property is that no situation can systematically yield downstream events with consistently higher or lower EPV. This is not trivial, as regression and other marginal methods that estimate each point of the EPV curve in isolation do not guarantee this coherence. For example, under a marginal estimation scheme yielding a sequence of estimates  $\hat{\nu}_t$ , it is possible for a Simpson’s paradox to arise where for some  $t$  and  $\Delta$ ,  $\hat{\nu}_{t+\Delta} < \hat{\nu}_t$  no matter what occurs at time  $t$ . On the other hand, our methodology, which explicitly computes the integral in (1) with respect to a model for the whole process  $Z$ , maintains stochastic consistency.

A simple model that does provide stochastic consistency is discretizing  $Z_t$  and modeling it as a homogeneous Markov chain. Markov chains have been commonly used for modeling final outcomes conditional on observed progress in other sports, such as in-game win probability in baseball (Bukiet et al. 1997; Yang & Swartz 2004) and in-possession point totals in football (Goldner 2012). In both these examples, the data is naturally discrete in space and time; however, our data  $Z_t$  is essentially continuous in space and time (we do observe data only at regular intervals of 1/25 second). Discretizing this process forces the investigator to trade off between the smoothness and level of detail captured in the process  $\nu_t$  (having a larger state space), and the ease of estimation and computation. While a state space with a huge number of states may in theory provide smooth, stochastically consistent estimated EPV curves, the associated transition probability matrix would be very difficult to estimate since a large number of states induces sparsity in the observed transition matrix. Moreover, computing expected values of a homogeneous Markov chain requires solving a linear system of the same dimension as the number of states, which is a cubic-time operation; this would be computationally infeasible for a huge state space.

Our methodology for estimating EPV, leveraging the idea of multiresolution transitions, largely avoids this tradeoff and offers precise, stochastically consistent EPV curve estimates. Generating such estimates is computationally demanding, but feasible given modern computing infrastructure and inference techniques.

### 3. MULTIREOLUTION MODELING

The stochastic process approach to estimating EPV requires that we integrate over the distribution of future paths the current possession can take. Letting  $T(\omega)$  denote the time at which a possession following path  $\omega$  ends<sup>1</sup>, the possession’s point total is a deterministic function of the full resolution data at this time,  $X(\omega) = h(Z_{T(\omega)}(\omega))$ . Thus, evaluating EPV amounts to integrating over the joint distribution of  $(T, Z_T)$ :

$$\begin{aligned}\nu_t &= \mathbb{E}[X|\mathcal{F}_t^{(Z)}] = \int_{\Omega} X(\omega)\mathbb{P}(d\omega|\mathcal{F}_t^{(Z)}) \\ &= \int_t^\infty \int_{\mathcal{Z}} h(z)\mathbb{P}(Z_s = z|T = s, \mathcal{F}_t^{(Z)})\mathbb{P}(T = s|\mathcal{F}_t^{(Z)})dzds.\end{aligned}\quad (1)$$

Note that we use probability notation  $\mathbb{P}(\cdot)$  somewhat heuristically, as  $\mathbb{P}(T = s|\mathcal{F}_t^{(Z)})$  is a density with respect to Lebesgue measure, while  $Z_s$  mixes both discrete (annotations) and continuous (locations) components. The best way to integrate (1) is by simulating future paths of the full resolution data with a transition kernel  $\mathbb{P}(Z_{t+\epsilon}|\mathcal{F}_t^{(Z)})$  until the simulated possession ends by reaching an observed point outcome. Such simulations provide a Monte Carlo estimate of (1). A model for this transition kernel requires a novel mixture of components for both the continuous spatial evolution of players, as well as their discrete decisions and ball movements.

Our approach is to model the possession process  $Z$  at two separate levels of resolution. In addition to modeling the short-term evolution of  $Z$  at full resolution, we simultaneously model a coarsened view of the process  $Z$  that is discrete in space and continuous in time. We combine these models in a multiresolution conditioning scheme that yields EPV calculations that are both computationally tractable (using Monte Carlo) and interpretable in terms of relevant basketball motifs.

#### 3.1 A Coarsened Process

A key component of our approach is a coarsening of the full-resolution data  $Z$  that preserves the characteristic dynamics of basketball play while shedding fine-resolution detail. For all time  $0 < t \leq T$  during a possession, assume  $C_t$  summarizes the “state” of the possession, such that  $C_t \in \mathcal{C}$  for some finite set  $\mathcal{C}$ . We populate the states  $c \in \mathcal{C}$  with summaries of the full resolution data so that transitions between these states represent meaningful events in a basketball possession. We illustrate the coarsened process  $C$  in Figure 3.

First, there are 3 “bookkeeping” states, denoted  $\mathcal{C}_{\text{end}}$ , that categorize the end of the possession, so that  $C_T \in \mathcal{C}_{\text{end}}$  and for all  $t < T, C_t \notin \mathcal{C}_{\text{end}}$  (shown in the bottom row of Figure 3). These are  $\mathcal{C}_{\text{end}} = \{\text{made 2 pt, made 3 pt, end of possession}\}$ . These three states have associated point values— 2 points for a made 2 point shot, 3 points for a made 3 point shot, and 0 points for the generic possession end state (which can be reached by turnovers, defensive rebounds, etc.). Thus, there is a map  $h : \mathcal{C}_{\text{end}} \rightarrow \{0, 2, 3\}$  allowing us to rewrite the EPV equation in terms of the coarsened process:  $\nu_t = \mathbb{E}[h(C_T)|\mathcal{F}_t^{(Z)}]$ .

---

<sup>1</sup>The time of a possession is bounded, even for pathological examples, by the 12-minute length of a quarter; yet we do not leverage this fact and simply assume that possession lengths are almost surely finite.

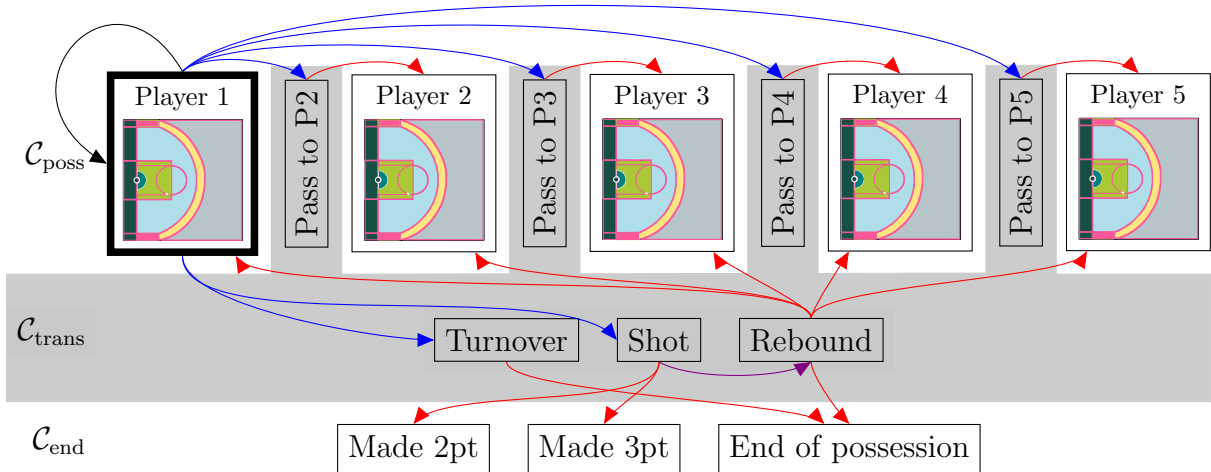


Figure 3: Schematic of the coarsened possession process  $\mathcal{C}$ , with states (rectangles) and possible state transitions (arrows) shown. The unshaded states in the first row compose  $\mathcal{C}_{\text{poss}}$ . Here, states corresponding to distinct ballhandlers are grouped together (Player 1 through 5), and the discretized court in each group represents the player’s coarsened position and defended state. The gray shaded rectangles are macrotransition states  $\mathcal{C}_{\text{trans}}$ , while the rectangles in the third row represent the end states  $\mathcal{C}_{\text{end}}$ . Blue arrows are the beginnings of macrotransition that can result when a player (WLOG Player 1 in this figure) possesses the ball. Red arrows are macrotransition exits. The purple arrow (between Shot and Rebound) carries the same macrotransition (beginning with the shot attempt) into the rebound state (from which it will exit). The black arrow is a microtransition, as the ballcarrier is unchanged.

Next, whenever a player possesses the ball at time  $t$ , we assume  $C_t = (\text{ballcarrier ID at } t) \times (\text{court region at } t) \times (\text{defended at } t)$ , having defined seven disjoint regions of the court and classifying a player as defended at time  $t$  by whether there is a defender within 5 feet of him. The possible values of  $C_t$ , if a player possesses the ball at time  $t$ , thus live in  $\mathcal{C}_{\text{poss}} = \{\text{player ID}\} \times \{\text{region ID}\} \times \{\mathbf{1}[\text{defended}]\}$ . These states are represented by the unshaded portion of the top row of Figure 3, where the differently colored regions of the court diagrams reveal the court space discretization.

Finally, if a player has initiated an annotated action currently in progress, we define  $C_t$  to take a “transition” state value. These states encapsulate constrained motifs in a possession, for example, when the ball is in the air traveling between players in a pass attempt. Explicitly, denote  $\mathcal{C}_{\text{trans}} = \{\text{shot attempt from } c \in \mathcal{C}_{\text{poss}}, \text{ pass attempt toward } c' \in \mathcal{C}_{\text{poss}} \text{ from } c \in \mathcal{C}_{\text{poss}}, \text{ turnover in progress, rebound in progress}\}$  (listed in the gray shaded portions of Figure 3). These transition states carry information about the possession path, such as the most recent ballcarrier, or the target of the pass, while the ball is in the air during shot attempts and passes.

Note that due to limitations of the data, this construction of  $\mathcal{C} = \mathcal{C}_{\text{poss}} \cup \mathcal{C}_{\text{trans}} \cup \mathcal{C}_{\text{end}}$  excludes several notable basketball events, such as fouls and violations, balls going out of



bounds without a change of possession, and other stoppages of play. In the case of turnovers, the event labels do not discriminate among steals, intercepted passes, or lost balls out of bounds, thus we treat this as a single category.

### 3.2 Multiresolution Conditioning

When the coarsened process  $C_t$  transitions from a state in  $\mathcal{C}_{\text{poss}}$  to one in  $\mathcal{C}_{\text{trans}}$ , we call this transition between coarsened states a *macrotransition*.

**Definition** If  $C_t \in \mathcal{C}_{\text{poss}}$  and  $C_{t+\epsilon} \in \mathcal{C}_{\text{trans}}$ , then  $C_t \rightarrow C_{t+\epsilon}$  is a *macrotransition*.

Macrotransitions, which include all ball movements (passes, shot attempts, turnovers), mark large-scale shifts that form the basis of offensive basketball play. The term carries a double meaning, as a macrotransition describes both a transition among states in our coarsened process,  $C_t \rightarrow C_{t+\epsilon}$ , as well as a transition of ballcarrier identity on the basketball court. By construction, for a possession that is in a state in  $\mathcal{C}_{\text{poss}}$  to proceed to a state in  $\mathcal{C}_{\text{end}}$  or a state in  $\mathcal{C}_{\text{poss}}$  corresponding to a different ballhandler, a macrotransition must occur as possession passes through a transition state in  $\mathcal{C}_{\text{trans}}$  (see possible transition paths illustrated in Figure 3).

This structure reveals that at any time  $t$  during a possession, we are guaranteed to observe the *exit state* of a future (or current, if  $C_t \in \mathcal{C}_{\text{trans}}$ ) macrotransition. Specifically, let  $\delta = \min\{s : s > t, C_{s-\epsilon} \in \mathcal{C}_{\text{trans}} \text{ and } C_s \notin \mathcal{C}_{\text{trans}}\}$  denote the time the possession reaches the state *after* the next (or current, if  $C_t \in \mathcal{C}_{\text{trans}}$ ) macrotransition after time  $t$ . Thus, if the possession is currently in a macrotransition,  $\delta$  is the first time at which a new possession or end state is occupied (ending the macrotransition), while if a player currently possesses the ball,  $\delta$  is the time at which the possession reaches the exit state of a future macrotransition.  $\delta$  is a bounded stopping time, so we can condition on  $C_\delta$  to rewrite EPV (1) as

$$\nu_t = \sum_{c \in \mathcal{C}} \mathbb{E}[h(C_T) | C_\delta = c, \mathcal{F}_t^{(Z)}] \mathbb{P}(C_\delta = c | \mathcal{F}_t^{(Z)}). \quad (2)$$

It is helpful to expand the second term in (2),  $\mathbb{P}(C_\delta = c | \mathcal{F}_t^{(Z)})$ , by conditioning on the start of the macrotransition that corresponds to the exit state  $C_\delta$ . Denote  $M(t)$  as the event that a macrotransition begins in  $(t, t + \epsilon]$ , and let  $\tau = \min\{s : s > t, M(s)\}$  be the time at which the macrotransition ending in  $C_\delta$  begins. Thus,  $\tau$  and  $\delta$  bookend the times during which the possession is in the next (or current, but ongoing) macrotransition, with  $C_\tau$  being the state in  $\mathcal{C}$  immediately *prior* to the start of this macrotransition and  $C_\delta$  the state immediately succeeding it. Like  $\delta$ , at any time  $t < T$ ,  $\tau$  is a bounded stopping time; however, note that if a macrotransition is in progress at time  $t$  then  $\tau < t$ , and, having been observed,  $\tau$  has a degenerate distribution. Defining  $\tau$  allows us to write:

$$\begin{aligned} \mathbb{P}(C_\delta = c | \mathcal{F}_t^{(Z)}) &= \sum_{c \in \mathcal{C}} \int_t^\infty \int_{\mathcal{Z}} \mathbb{P}(C_\delta = c | M(\tau), Z_\tau = z, \tau = s, \mathcal{F}_t^{(Z)}) \\ &\quad \times \mathbb{P}(M(\tau), Z_\tau = z, \tau = s | \mathcal{F}_t^{(Z)}) dz ds. \end{aligned} \quad (3)$$

We make one additional expansion to the terms we have introduced for calculating EPV. The second factor in (3),  $\mathbb{P}(M(\tau), Z_\tau = z, \tau = s | \mathcal{F}_t^{(Z)})$ , models the location and time of

the next macrotransition—implicitly averaging over the intermediate path of the possession in the process. This is the critical piece of our multiresolution structure that connects the full-resolution process  $Z$  to the coarsened process  $C$ , and the component of our model that fully utilizes multiresolution conditioning. We expand this term using our macro- and microtransition models.

**Definition** The *macrotransition model* is  $\mathbb{P}(M(t)|\mathcal{F}_t^{(Z)})$ .

**Definition** The *microtransition model* is  $\mathbb{P}(Z_{t+\epsilon}|M(t)^c, \mathcal{F}_t^{(Z)})$ , where  $M(t)^c$  is the complement of  $M(t)$ . *Microtransitions* are instantaneous changes in the full resolution data  $Z_t \rightarrow Z_{t+\epsilon}$  over time windows where a macrotransition is not observed; thus, only location components (and not event annotations) change from  $Z_t$  to  $Z_{t+\epsilon}$ .

Multiresolution transition models allow us to sample from  $\mathbb{P}(\tau, Z_\tau|\mathcal{F}_t^{(Z)})$ , enabling Monte Carlo evaluation of (3). The basic idea is that we use the macrotransition model to draw from  $\mathbb{P}(M(t)|\mathcal{F}_t^{(Z)})$  and if  $M(t)^c$  and no macrotransition occurs in  $(t, t + \epsilon]$ , we use the microtransition model to draw from  $\mathbb{P}(Z_{t+\epsilon}|M(t)^c, \mathcal{F}_t^{(Z)})$ . Iterating this process, we alternate draws from the macro- and microtransition models until observing  $(\tau, Z_\tau)$ —of course, this also yields  $M(\tau)$  as a consequence of our definition of  $\tau$ . Parametric forms for these macro- and microtransition models are discussed explicitly in Sections 4 and 5 respectively, while Section 6 provides additional details on the Monte Carlo integration scheme.

Expanding EPV by conditioning on intermediate values in principle does not ease the problem of its evaluation. However several of the components we have introduced motivate reasonable conditional independence assumptions that simplify their evaluation. Only by writing EPV as an average over additional random variables defined in the probability space of our possession can we articulate such assumptions and leverage them to compute EPV.

### 3.3 Conditional Independence Assumptions

Our expansions of  $\nu_t = \mathbb{E}[h(C_T)|\mathcal{F}_t^{(Z)}]$  introduced in the previous subsection (2)–(3) express EPV in terms of three probability models:

$$\begin{aligned} \nu_t = \sum_{c \in \mathcal{C}} E[h(C_T)|C_\delta = c, \mathcal{F}_t^{(Z)}] & \left( \int_t^\infty \int_{\mathcal{Z}} \mathbb{P}(C_\delta = c|M(\tau), Z_\tau = z, \tau = s, \mathcal{F}_t^{(Z)}) \right. \\ & \left. \times \mathbb{P}(M(\tau), Z_\tau = z, \tau = s|\mathcal{F}_t^{(Z)}) dz ds \right). \end{aligned} \quad (4)$$

The multiresolution transition models sample from  $\mathbb{P}(M(\tau), Z_\tau, \tau|\mathcal{F}_t^{(Z)})$ , eliminating the need to evaluate the third term in (4) explicitly when computing  $\nu_t$  via Monte Carlo. The second term in (4) is actually quite easy to work with since  $C_\delta$  is categorical, and given  $Z_\tau$  the space of possible values it can take is relatively small. This is due to the manner in which macrotransitions constrain the spatiotemporal evolution of the possession. Given  $Z_\tau$ , we can obtain the location and separation from the defense of all four possible pass recipients given a pass in  $(\tau, \tau + \epsilon]$ , so only a subset of states in  $\mathcal{C}_{\text{poss}}$  are possible for  $C_\delta$ . Similarly, if a shot attempt occurs in this time window,  $Z_\tau$  indicates whether a successful shot would

yield 2 or 3 points, further subsetting the possible values of  $C_\delta$ . Modeling  $C_\delta$  thus reduces to predicting the type of macrotransition corresponding to  $M(\tau)$ —a pass, shot attempt, or turnover. We discuss this in Section 4 in the context of our macrotransition model.

The first term in (4),  $E[h(C_T)|C_\delta = c, \mathcal{F}_t^{(Z)}]$  provides the expected point value of the possession given the (coarsened) result of the next macrotransition. Prima facie, this term seems as difficult to evaluate as it has the same essential structure as EPV itself, requiring integration over the future trajectory of the possession after time  $\delta$ . However, we make a key assumption that frees subsequent evolution of the possession, after time  $\delta$ , from dependence on the full-resolution history  $\mathcal{F}_t^{(Z)}$ :

$$\mathbb{E}[h(C_T)|C_\delta, \mathcal{F}_t^{(Z)}] = \mathbb{E}[h(C_T)|C_\delta]. \quad (5)$$

This assumption is intuitive for two reasons. First, by constraining the possession to follow a restricted spatiotemporal path, it is reasonable to assume that the macrotransition exit state itself contains sufficient information to characterize the future evolution of the system. Secondly, because macrotransitions play out over much longer timescales than the resolution of the data (i.e., several seconds, as opposed to 1/25th of a second), it is reasonable to assume that fine-scale spatial detail before the start of the macrotransition has been “mixed out” by the time the macrotransition ends.

An additional, reasonable conditional independence assumption is that the coarsened state sequence  $C_t, t > 0$  is marginally a semi-Markov process; that is, denoting  $\mathcal{F}_t^{(C)} = \sigma(\{C_s^{-1}, 0 \leq s \leq t\})$  as the history of the coarsened process, for all  $t' > t$  and  $c \in \mathcal{C}$ , we assume  $\mathbb{P}(C_{t'} = c | \mathcal{F}_t^{(C)}) = \mathbb{P}(C_{t'} = c | C_t)$ . A semi-Markov process generalizes a continuous time Markov Chain in that sojourn times need not be exponentially distributed. We associate with this semi-Markov process an embedded discrete, homogeneous Markov Chain: denote  $C^{(0)}, C^{(1)}, \dots, C^{(K)}$  as the sequence of consecutive states  $c \in \mathcal{C}$  visited by  $C_t$  during the possession  $0 < t \leq T$ . Thus,  $C^{(K)} = C_T$ , and  $K$  records the length of the possession in terms of the number of transitions between states in  $\mathcal{C}$ , which like  $T$  is random.

Combining these assumptions, the first term in (4),  $\mathbb{E}[h(C_T)|C_\delta, \mathcal{F}_t^{(Z)}]$ , can be computed easily from the transition probability matrix of the homogeneous Markov chain embedded in  $C_t$ . As  $C^{(K)}$  is an absorbing state, ending the possession, we can rewrite (5) as  $\mathbb{E}[h(C^{(K)})|C_\delta]$ . This is easily obtained by solving a linear system of equations deriving from the transition probability matrix of  $C^{(0)}, C^{(1)}, \dots, C^{(K)}$ . Estimating this transition probability matrix is also discussed in Section 4, where we show that it actually derives from the macrotransition model.

Compared to using discrete, homogeneous Markov Chains alone to calculate EPV, the multiresolution approach we take ultimately leverages much of the same computational advantages while remaining attenuated to the full-resolution data, responding smoothly as the possession evolves over space and time.

#### 4. MACROTRANSITION MODEL

Macrotransitions play a fundamental role in our EPV framework. Intuitively, they represent the strategies, schemes, and decision-making that characterize basketball offense. Mathematically, macrotransitions are part of our multiresolution conditioning scheme used to evaluate

EPV at any time during a possession given its history. Introduced in the previous section, the macrotransition model is  $\mathbb{P}(M(t)|\mathcal{F}_t^{(Z)})$ . More generally, we consider a family of macrotransition models  $\mathbb{P}(M_j(t)|\mathcal{F}_t^{(Z)})$ , where  $j$  indexes the type of macrotransition corresponding to  $M(t)$ . At any given moment when a player possesses the ball, there are six possible categories of macrotransition, corresponding to 4 pass options, a shot attempt, or a turnover, which we index by  $j \in \{1, \dots, 6\}$ . Without loss of generality, assume  $j \leq 4$  correspond to pass events,  $j = 5$  is a shot attempt and  $j = 6$  a turnover. Thus,  $M_j(t)$  is the event that a macrotransition of type  $j$  begins in the time window  $(t, t + \epsilon]$ , and  $M(t) = \bigcup_{j=1}^6 M_j(t)$ .

We now introduce the parameterization of the macrotransition models  $\mathbb{P}(M_j(t)|\mathcal{F}_t^{(Z)})$ , and also discusses how other components of our EPV equation (4) derive from these models.

#### 4.1 Macrotransition Entry Model

As  $M_j(t)$  denotes the start of a macrotransition in  $(t, t + \epsilon]$ , we refer to  $\mathbb{P}(M_j(t)|\mathcal{F}_t^{(Z)})$ , for all  $j$ , as the macrotransition entry model. This is specified using competing risks (Prentice, Kalbfleisch, Peterson Jr, Flournoy, Farewell & Breslow 1978): assuming player  $\ell$  possesses the ball at time  $t > 0$  during a possession, then denote

$$\lambda_j^\ell(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathbb{P}(M_j(t)|\mathcal{F}_t^{(Z)})}{\epsilon} \quad (6)$$

as the hazard for macrotransition  $j$  at time  $t$ , or the cause-specific hazard. As events  $M_1(t), \dots, M_6(t)$  are disjoint, it follows that the total macrotransition hazard is the sum of the cause-specific hazards,

$$\lim_{\epsilon \rightarrow 0} \frac{\mathbb{P}(M(t)|\mathcal{F}_t^{(Z)})}{\epsilon} = \sum_j \lambda_j(t)$$

We assume the cause-specific hazards are log-linear,

$$\log(\lambda_j^\ell(t)) = [\mathbf{W}_j^\ell(t)]' \boldsymbol{\beta}_j^\ell + \xi_j^\ell(\mathbf{z}_\ell(t)) + \left( \tilde{\xi}_j^\ell(\mathbf{z}_j(t)) \mathbf{1}[j \leq 4] \right), \quad (7)$$

where  $\mathbf{W}_j^\ell(t)$  is a  $p_j \times 1$  vector of time-varying covariates,  $\boldsymbol{\beta}_j^\ell$  a  $p_j \times 1$  vector of coefficients,  $\mathbf{z}_\ell(t)$  is the ballcarrier's 2D location on the court (denote the court space  $\mathbb{S}$ ) at time  $t$ , and  $\xi_j^\ell : \mathbb{S} \rightarrow \mathbb{R}$  is a mapping of the player's court location to an additive effect on the log-hazard, providing spatial variation. The last term in (7) only appears for pass events ( $j \leq 4$ ) to incorporate the location of the receiving player for the corresponding pass:  $\mathbf{z}_j(t)$  (which slightly abuses notation) provides his location on the court at time  $t$ , and  $\tilde{\xi}_j^\ell$ , analogously to  $\xi_j^\ell$ , maps this location to an additive effect on the log-hazard. All spatial effects  $\xi$  are assumed to be realizations of Gaussian processes; a detailed discussion of the structure and estimation of these spatial effects is included in Appendix A.

The macrotransition model (6)–(7) represents the ballcarrier's decision-making process as an interpretable function of the unique basketball predicaments he faces. For example, in considering the hazard of a shot attempt, the time-varying covariates ( $\mathbf{W}_j^\ell(t)$ ) we use are the distance between the ballcarrier and his nearest defender (transformed as  $\log(1 + d)$ )

to moderate the influence of extremely large or small observed distances), an indicator for whether the ballcarrier has dribbled since gaining possession, and a constant representing a baseline shooting rate (this is not time-varying)<sup>2</sup>. The spatial effects  $\xi_j^\ell$  reveal locations where player  $\ell$  is more/less likely to attempt a shot in a small time window, holding fixed the time-varying covariates  $\mathbf{W}_j^\ell(t)$ . Such spatial effects (illustrated in Figure 4) are well-known to be nonlinear in distance from the basket and asymmetric about the angle to the basket (Miller, Bornn, Adams & Goldsberry 2013).

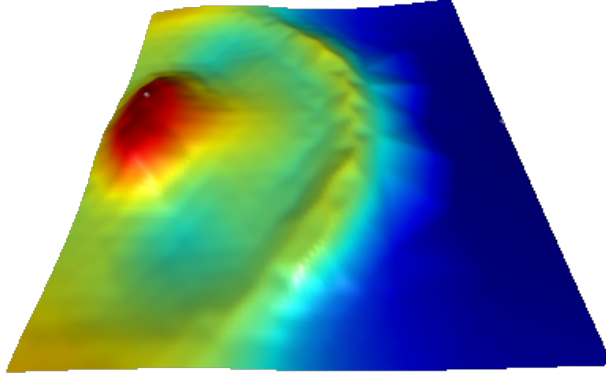


Figure 4: Estimated  $\xi_j^\ell$  for LeBron James’ shot-taking hazard ( $j = 5$ ). For each location on the court, the color and height of the surface are proportional to the additive effect on the log hazard of attempting a shot at that location.

For pass events, the time-varying covariates, their coefficients, and the spatial effect  $\xi_j^\ell$  vary for the same ballcarrier ( $\ell$ ) across his different passing options  $j = 1, \dots, 4$ . This reflects the fact that pass events between two players depend on those two players’ positions and roles on the team. For instance, a point guard will pass to a center in different situations than those in which he passes to a shooting guard. Thus in some sense, we are modeling the pass events independently for every passer-receiver pair; main effects from the passer’s and receiver’s identities are not explicitly modelled, though hierarchical models allow information sharing among the different pass models associated with player  $\ell$  (Appendix A introduces such hierarchical structure, which we use in our parameter estimation). It is conceptually useful to think of  $(\xi_j^\ell, \tilde{\xi}_j^\ell) : \mathbb{R}^4 \rightarrow \mathbb{R}$  as jointly providing a single spatial effect for the 4D location of the passer/receiver pair, and of the factorization

$$(\xi_j^\ell, \tilde{\xi}_j^\ell)(\mathbf{z}_\ell(t), \mathbf{z}_j(t)) = \xi_j^\ell(\mathbf{z}_\ell(t)) + \tilde{\xi}_j^\ell(\mathbf{z}_j(t))$$

as an assumption designed to reduce the computational complexity of fitting such a model.

#### 4.2 Macrotransition Exit Model

In Section 3, we noted that the model for the macrotransition exit state conditional on a macrotransition occurring,  $\mathbb{P}(C_\delta | M(\tau), Z_\tau, \tau, \mathcal{F}_t^{(Z)})$ , derives from the macrotransition model.

---

<sup>2</sup>Full details on all covariates used for all macrotransition types are included in Appendix A.2

We show this by noting that

$$\mathbb{P}(C_\delta | M(\tau), Z_\tau, \tau, \mathcal{F}_t^{(Z)}) = \sum_j \mathbb{P}(C_\delta | M_j(\tau), Z_\tau, \tau, \mathcal{F}_t^{(Z)}) \mathbb{P}(M_j(\tau) | M(\tau), Z_\tau, \tau, \mathcal{F}_t^{(Z)}). \quad (8)$$

Probabilities constituting the second term in (8) are proportional to  $\lambda_j(t)$ , thus they derive from our family of competing risks macrotransition models. For  $j \neq 5$  (not a shot attempt), the first term in (8) is actually degenerate. For each pass (corresponding to  $j \leq 4$ ), the location and position relative to the defense of the pass recipient are given by  $Z_\tau$ , thus yielding only one possible macrotransition exit state  $C_\delta$  for each pass option. Note that while players' positions change in the time window  $(\tau, \delta)$ , while a pass is airborne, we do not expect the pass recipient's location in the coarsened space  $\mathcal{C}$  to change (though our model could be augmented to incorporate this). Similarly, if  $j = 6$  and a turnover occurs at  $\tau$ , then  $C_\delta$  is the turnover state in  $\mathcal{C}_{\text{end}}$  with probability 1.

However, if  $j = 5$  and a shot is attempted in  $(\tau, \tau + \epsilon]$ , then  $C_\delta$  has two possible values depending on the shooter's location in  $Z_\tau$ : a made or missed 2 point shot, or made or missed 3 point shot. This motivates a shot probability model, predicting the probability of success given a shot attempt at time  $t$  and the associated full resolution data. The parametric form of our shot probability model is exactly the same as our macrotransition model, though we use a logit link function as we are modeling a probability instead of a hazard. Specifically, for player  $\ell$  possessing attempting a shot at time  $t$ , let  $p^\ell(t)$  represent the probability of the shot attempt being successful (resulting in a basket). We assume

$$\text{logit}(p^\ell(t)) = [\mathbf{W}^\ell(t)]' \boldsymbol{\beta}^\ell + \xi^\ell(\mathbf{z}_\ell(t)) \quad (9)$$

with components in (9) having the same interpretation as their  $j$ -indexed counterparts in the competing risks model (7); that is,  $\mathbf{W}^\ell$  is a vector of time-varying covariates (we use distance to the nearest defender—transformed as  $\log(1 + d)$ —an indicator for whether the player has dribbled, and a constant to capture baseline shooting efficiency) with  $\boldsymbol{\beta}^\ell$  a corresponding vector of coefficients, and  $\xi^\ell$  is a smooth spatial effect, assumed to be a realization of a Gaussian process.

#### 4.3 Transition Probability Matrix for Coarsened Process

The last component of the EPV calculation supplied by the macrotransition model is the transition probability matrix for the embedded Markov chain corresponding to the coarsened process  $C^{(0)}, C^{(2)}, \dots, C^{(K)}$ . This transition probability matrix is used to compute terms  $\mathbb{E}[h(C_T) | C_\delta]$  that appear in EPV equations (4)–(5). We shall denote the transition probability matrix as  $\mathbf{P}$ , where  $P_{qr} = \mathbb{P}(C^{(i+1)} = c_r | C^{(i)} = c_q)$  for any  $c_q, c_r \in \mathcal{C}$ .

Without any other probabilistic structure assumed for  $C^{(i)}$  other than Markov, for all  $i, j$ , the maximum likelihood estimator of  $P_{qr}$  is the observed transition frequency  $\frac{\#\{c_q \rightarrow c_r\}}{\#\{\text{visits to } c_q\}}$ . Of course, this estimator has undesirable performance if the number of visits to any particular state  $c_q$  is small, as the estimated transition probabilities from that state may be degenerate. One common approach is to model transition probability matrices hierarchically, possibly in a Bayesian fashion (Lee, Judge & Zellner 1968; Meshkani & Billard 1992).

Under our multiresolution model for basketball possessions, however, transition proba-

bilities between many coarsened states  $C^{(i)}$  can be computed as summaries of the macrotransition model. To show this, for any arbitrary  $t > 0$  let  $M_j^r(t)$  be the indicator

$$M_j^r(t) = \mathbf{1}[\mathbb{P}(M_j(t) \text{ and } C_{t+\epsilon} = c_r | \mathcal{F}_t^{(Z)}) > 0].$$

Thus  $M_j^r(t) = 1$  if it is possible for a macrotransition of type  $j$  into state  $c_q$  to occur in  $(t + t + \epsilon]$ . Now, for any  $c_q$  such that  $c_q \rightarrow c_r$  is a macrotransition, we can write

$$\begin{aligned} P_{qr} &= \mathbb{P}(C_{t+\epsilon} = c_r | C_t = c_q) \\ &= \mathbb{P}(M_j(t) | C_t = c_q, M_j^r(t) = 1) \\ &= \epsilon \mathbb{E}[\lambda_j(t) | C_t = c_q, M_j^r(t) = 1], \end{aligned} \tag{10}$$

where the last equality follows simply from iterated expectation, noting that  $C_t$  and  $M_j^r(t)$  are both  $\mathcal{F}_t^{(Z)}$ -measureable. Since we assume  $C_t$  is semi-Markov, (10) holds for any  $t$ .

The integrating measure in (10), which conditions on  $C_t$  and  $M_j^r(t)$ , is not immediately available from the multiresolution models without an onerous set of assumptions, so we substitute the empirical distribution of possession paths that occupy  $c_q$  at some time point. This yields a simple (unnormalized) estimator  $\tilde{P}_{qr} = \sum_{t \in \mathcal{T}^q} \epsilon \lambda_j(t)$  for each  $r$  such that  $c_q \rightarrow c_r$  is a macrotransition for some  $j$ , where  $\mathcal{T}^q$  is the set of (discretized at resolution  $\epsilon$ ) times for which  $C_t = c_q$ . Thus, we estimate the transition probability by accumulating the appropriate transition hazard  $\lambda_j(t)$ . This method leverages the parametric structure of our macrotransition model, and by propagating the shrinkage and temporal smoothing in the macrotransition model to the estimates of  $P_{qr}$ , we achieve greater precision than with the naïve MLE.

Transition probabilities corresponding to macrotransition exits are often degenerate (either 0 or 1); this is because for passes and turnovers, the exit state is encoded in the definition of the transition state—for instance, each pass state in  $\mathcal{C}_{\text{trans}}$  transitions to a single possession state in  $\mathcal{C}_{\text{poss}}$  with probability 1. The exception to this is the exit state of a shot attempt. Recalling that shot attempt states  $c_q$  are indexed by the state from which the shot originated, denoted  $c_{q'}$ , then the next state  $c_r$  is either a made or missed 2 point shot, or a made or missed 3 point shot, depending on whether the location corresponding to  $c_{q'}$  is behind the three point line. Given the potential point value of the shot, we determine its success probability following a similar line of reasoning:  $\tilde{P}_{qr} = \sum_{t \in \mathcal{T}^{q'}} \epsilon \lambda_5(t) p(t)$  when  $r$  represents a successful shot, and  $\tilde{P}_{qr} = \sum_{t \in \mathcal{T}^{q'}} \epsilon \lambda_5(t) (1 - p(t))$  when  $r$  represents a missed shot.

For all other transitions, where  $c_q \rightarrow c_r$  is not a macrotransition entry or exit, we simply use observed transitions  $\tilde{P}_{qr} = \sum_{t \in \mathcal{T}^q} \mathbf{1}[C_{t+\epsilon} = c_r]$ . Then the unnormalized transition rates yield estimated transition probabilities for all  $q, r$ :

$$\hat{P}_{qr} = \frac{\tilde{P}_{qr}}{\sum_{r'} \tilde{P}_{qr'}}. \tag{11}$$

For transitions from a state not associated with a particular player, we group observed transitions separately by team. The only example of this is the rebound state in  $\mathcal{C}_{\text{trans}}$ ,

which with some probability transitions to a defensive rebound (thus ending the possession) and in the case of an offensive rebound transitions into one of the possession states.

## 5. MICROTRANSITION MODEL

While the macrotransition model in Section 4 models ball movements, the microtransition model describes player movement with the ballcarrier held constant. In the periods between transfers of ball possession (including shots), all players on the court move in order to influence the character of the next ball movement (macrotransition). For instance, the ballcarrier might drive toward the basket to attempt a shot, or move laterally to gain separation from a defender, while his teammates move to position themselves for passes or rebounds, or to set screens and picks. The defense moves correspondingly, attempting to deter easy shot attempts or passes to certain players while simultaneously anticipating a possible turnover.

As defined in Section 3, the microtransition model supplies  $\mathbb{P}(Z_{t+\epsilon}|M(t)^c, \mathcal{F}_t^{(Z)})$ , giving the small-scale evolution of the current possession conditional on the ballcarrier staying the same in the next  $\epsilon$  time. Separate models are assumed for offensive and defensive players, as we shall describe.

### 5.1 Offensive Movement

Predicting the motion of offensive players over a relatively short time window is driven by the players' dynamics (velocity, acceleration, etc.). Let the location of an offensive player (the ballcarrier, for instance) at time  $t$  be  $\mathbf{z}(t) = (x(t), y(t))$ . Assuming the player's position is differentiable, a Taylor series expansion shows  $x(t + \epsilon) = x(t) + \dot{x}(t)\epsilon + e_x(t)$  where the innovations  $e_x(t)$  depend on higher derivatives of position (acceleration, jerk, etc.) and possibly involve white noise, as there is small measurement error associated with the position tracking in our data. The velocity  $\dot{x}(t)$  is unobserved, yet it is natural to replace this with  $(x(t) - x(t - \epsilon))/\epsilon$ , acknowledging that doing so adds additional structure to the residual  $e_x(t)$  term. We are now left with

$$x(t + \epsilon) = x(t) + a_x[x(t) - x(t - \epsilon)] + e_x(t), \quad (12)$$

where  $a_x = 1$  in theory, but this is relaxed because  $a_x < 1$  in practice provides some predictive stability, as the sequence of differences  $x(t) - x(t + \epsilon)$  becomes a stationary AR(1) process if  $e_x(t)$  is white noise. Note that (12) defines  $x(t)$  as an ARI(1,1) process (or equivalently, as an ARIMA(1,1,0) process).

We also assume spatial structure for the innovations,  $e_x(t) \sim \mathcal{N}(\mu_x(\mathbf{z}(t)), \sigma_x^2)$ , where  $\mu_x$  maps the player's two-dimensional location on the court to an additive effect in (12), which has the interpretation of an acceleration effect. Players' future motion is informed not only by their current dynamics, but also their position on the court. Players within the perimeter, for instance, may be more likely to accelerate towards the basket as they get closer, eventually decelerating to attempt a shot. Also, players will accelerate away from the edges of the court as they approach these, in order to stay in bounds (see Figure 5 for an illustration). These behaviors motivate the inclusion of  $\mu_x$  in the model (12). For the  $y(t)$ , we construct (12) analogously in terms of  $a_y$  and  $e_y(t) \sim \mathcal{N}(\mu_y(\mathbf{z}(t)), \sigma_y^2)$ .



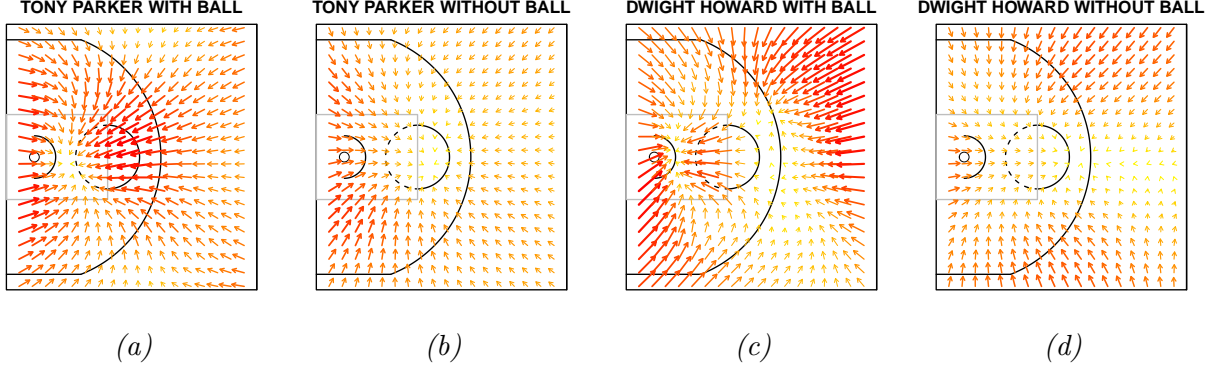


Figure 5: Acceleration fields  $(\mu_x(\mathbf{z}(t)), \mu_y(\mathbf{z}(t)))$  for Tony Parker (a)–(b) and Dwight Howard (c)–(d) with and without ball possession. The arrows point in the direction of the acceleration at each point on the court’s surface, and the size and color of the arrows are proportional to the magnitude of the acceleration. Comparing (a) and (c) for instance, we see that when both players possess the ball, Parker more frequently attacks the basket from outside the perimeter. Howard does not accelerate to the basket from beyond the perimeter, and only tends to attack the basket inside the paint.

## 5.2 Defensive Movement

The defensive components of  $P(Z_{t+\epsilon}|M(t)^c, \mathcal{F}_t^{(Z)})$ , corresponding to the positions of the five defenders, are easier to model conditional on the evolution of the offense’s positions. Following Franks, Miller, Bornn & Goldsberry (2014), we assume each defender’s position is centered on a linear combination of the basket’s location, the ball’s location, and the location of the offensive player he is guarding. Franks et al. (2014) use a hidden Markov model (HMM), based on this assumption, to learn which offensive players each defender is guarding, as well as the coefficients of this linear combination. The location coefficients they estimate are 0.62 for the offensive player being guarded, 0.11 for the ball, and 0.27 for the basket; that is, conditional on defender  $i$  guarding offender  $j$  his location  $\mathbf{z}_i(t)$  should be normally distributed with mean  $\mathbf{m}_j^{\text{opt}}(t) = 0.62\mathbf{z}_j(t) + 0.11\mathbf{z}_{\text{ball}} + 0.27\mathbf{z}_{\text{ball}}(t)$ .

Of course, the dynamics (velocity, etc.) of defensive players’ are still hugely informative for predicting their locations within a small time window. Thus our microtransition model for defenders balances these dynamics with the mean path induced by the player each is guarding:

$$x(t + \epsilon)|m_{j,x}^{\text{opt}}(t) = x(t) + a_x[x(t) - x(t - \epsilon)] + b_x[m_{j,x}^{\text{opt}}(t + \epsilon) - m_{j,x}^{\text{opt}}(t)] + \mathcal{N}(0, \tau_x^2). \quad (13)$$

Rather than implement the HMM procedure used in Franks et al. (2014), we simply assume each defender is guarding at time  $t$  whichever offensive player  $j$  yields the smallest residual  $\|\mathbf{z}(t) - \mathbf{m}_j^{\text{opt}}(t)\|$ . Note that more than one defender may be guarding the same offender (as in a “double team”). Thus, conditional on the locations of the offense at time  $t + \epsilon$ , (13) provides a distribution over the locations of the defense at time  $t + \epsilon$ . As in estimating the microtransition components for the offense, we fit (13) separately for all defenders, for both the  $x$  and  $y$  component of the position.

## 6. INFERENCE

To this point, we have expressed EPV using multiresolution conditioning, introduced parameterizations for our macro- and microtransition models, and derived all necessary probability components for evaluating EPV (such as the shot probability model and the transition probability matrix for  $C_t$ ) from parameters of these models. This section outlines the pipeline for estimating these model parameters and computing the derived EPV estimates that yield actual results, such as those highlighted in Section 2.1. Our estimates are based on all games from the 2013-14 season up until February 7, 2014, though only 90% of these games were used in model fitting (the remaining 10% were used to assess out of sample predictive performance or our model components).

### 6.1 Likelihood Inference

We estimate multiresolution transition models by Bayesian inference. Note that the multiresolution transition framework helps us rewrite the data generating process by conditioning on macrotransition events. The likelihood of a possession  $\{Z_t, 0 \leq t \leq T\}$  observed at a temporal resolution of  $\epsilon$  can be written

$$\begin{aligned} \prod_{t=0}^{T-\epsilon} \mathbb{P}(Z_{t+\epsilon} | \mathcal{F}_t^{(Z)}) &= \left( \prod_{t=0}^{T-\epsilon} \mathbb{P}(Z_{t+\epsilon} | M(t)^c, \mathcal{F}_t^{(Z)})^{\mathbf{1}_{[M(t)^c]}} \prod_{j=1}^6 \mathbb{P}(Z_{t+\epsilon} | M_j(t), \mathcal{F}_t^{(Z)})^{\mathbf{1}_{[M_j(t)]}} \right) \\ &\quad \times \left( \prod_{t=0}^{T-\epsilon} \mathbb{P}(M(t)^c | \mathcal{F}_t^{(Z)})^{\mathbf{1}_{[M(t)^c]}} \prod_{j=1}^6 \mathbb{P}(M_j(t) | \mathcal{F}_t^{(Z)})^{\mathbf{1}_{[M_j(t)]}} \right), \end{aligned} \quad (14)$$

where the first term models  $Z_{t+\epsilon}$  conditional on a macrotransition (or lack thereof) in the window  $(t, t + \epsilon]$ , and the second term models such macrotransition events. Following this factorization, we consider two separate models for macrotransitions and microtransitions, decomposing (14) into partial likelihoods (Cox 1975b) that inform the parameters of the macro- and microtransitions independently. The macrotransition model is estimated using the second term in the likelihood (14), whereas the microtransition model is fit using the term  $\left( \prod_{t=0}^{T-\epsilon} \mathbb{P}(Z_{t+\epsilon} | M(t)^c, \mathcal{F}_t^{(Z)})^{\mathbf{1}_{[M(t)^c]}} \right)$ . Under mild conditions, this inferential procedure leads to consistent and asymptotically well-behaved estimators (Wong 1986).

The reader is referred to Appendix A for explicit expressions of the partial likelihood terms for the macro and microtransition model in terms of the corresponding model parameters. Prior distributions are also given for these parameters, such that the inference is partially Bayesian (Cox 1975a). All spatial effects in our models ( $\xi$  in the macrotransition model and  $\mu$  in the microtransition model) are represented using functional bases whose loadings are given a unique prior structure that shares information across space and across players. This not only provides more precise inference with better out-of-sample predictive performance (see Table 1), but it also offers substantial computational advantages. Appendix A also outlines the computational requirements for parameter inference using (14).

### 6.2 EPV Algorithm

Given all parameter values for our multiresolution transition models, denoted  $\Theta$ , then EPV at any time during a possession  $\nu_t$  can be evaluated deterministically—though this may

require Monte Carlo integration depending on the current state of the possession. Algorithm 1 illustrates this process explicitly. EPVDRAW obtains a draw from the distribution of  $X = h(C_T)$  given  $\mathcal{F}_t^{(Z)}$ , and repeated draws yield an arbitrarily accurate estimate of  $\nu_t$ .

Inside the function EPVDRAW, we repeatedly iterate draws from the macro- and micro-transition models in order to simulate a future possession path up until a macrotransition occurs. From this sample path, we then draw the macrotransition exit state  $C_\delta$ , and compute its value  $\mathbb{E}[h(C_T)|C_\delta]$  using the transition probability matrix  $\mathbf{P}$ . This procedure mirrors the multiresolution conditioning equations (4) given in Section 3. Note that it is computationally necessary to work with a compressed version of  $\mathbf{P}$  corresponding to only the states accessible from  $C_t$ . This dramatically reduces the dimension of  $\mathbf{P}$ , as most states in  $\mathcal{C}$  are possession or transition states belonging to players not on the same team as the ballcarrier, meaning they are not accessible from  $C_t$ .

## 7. RESULTS

Applying Algorithm 1 using our parameter estimates for the multiresolution transition model, we can plot EPV ( $\nu_t$ ) throughout the course of any possession in our data. We view EPV curves as the main contribution of our work, and their behavior and potential inferential value has been introduced in Section 2.1. Analysts may also find meaningful aggregations of EPV curves that summarize players’ behavior over a possession, game, or season in terms of EPV—we offer two such aggregations in Appendix B.

### 7.1 Possession Inference from Multiresolution Transitions

Understanding the calculation of EPV in terms of multiresolution transitions is also a valuable exercise for a basketball analyst, as these model components reveal precisely how the EPV estimate derives from the spatiotemporal circumstances of the time point considered. Figure 6 diagrams four moments during our example possession (introduced originally in Figures 1 and 2) in terms of multiresolution transition probabilities. These diagrams illustrate equation (4) by showing EPV as a weighted average of the value of the next macrotransition. Potential ball movements representing macrotransitions are shown as arrows, with their respective values and probabilities graphically illustrated by color and line thickness (this information is also annotated explicitly). Microtransition distributions are also shown, indicating distributions of players’ movement over the next two seconds. Note that the possession diagrammed here was not used in our model fitting.

Analyzing Figure 6, we see that our model estimates largely agree with basketball intuition. For example, players are quite likely to take a shot when they are near to and/or moving towards the basket, as shown in panels A and D. Additionally, because LeBron James is a better shooter than Norris Cole, the value of his shot attempt is higher, even though in the snapshot in panel D he is much farther from the basket than Cole is in panel A. While the value of the shot attempt averages over future microtransitions, which may move the player closer to the basket, when macrotransition hazards are high this average is dominated by microtransitions on very short time scales.

We also see Ray Allen, in the right corner 3, as consistently one of the most valuable pass options during this possession, particularly when he is being less closely defended as in panels A and D. In these panels, though, we never see an estimated probability of him receiving

---

**Algorithm 1** Calculating EPV ( $\nu_t$ ).

---

**Require:** Player  $\ell$  possess the ball at time  $t$

```
function MACRO( $\mathcal{F}_s^{(Z)}, \Theta$ ) ▷ Simulates a possible macrotransition in  $(s, s + \epsilon]$ 
  for  $j$  in  $1, \dots, 6$  do
    Set  $M_j(s) = 1$  with probability  $\min\{1, \lambda_j^\ell(s)\}$ 
  end for
  if  $\sum_j M_j(s) > 1$  then
    Keep only one  $j$  such that  $M_j(s) = 1$ , choosing it proportional to  $\lambda_j^\ell(s)$ 
  end if
  return  $\{M_j(s), j = 1, \dots, 6\}$ 
end function

function EPVDRAW( $\mathcal{F}_t^{(Z)}, \Theta$ ) ▷ Gets EPV from single simulation of next macro
  Initialize  $s \leftarrow t$ 
  Initialize  $M_j(s) \leftarrow \text{MACRO}(\mathcal{F}_s^{(Z)}, \Theta)$ 
  while  $M_j(s) = 0$  for all  $j$  do
    Draw  $Z_{s+\epsilon} \sim \mathbb{P}(Z_{s+\epsilon} | M(s)^c, \mathcal{F}_s^{(Z)})$ 
     $\mathcal{F}_{s+\epsilon}^{(Z)} \leftarrow \{\mathcal{F}_t^{(Z)}, Z_{s+\epsilon}\}$ 
     $s \leftarrow s + \epsilon$ 
     $M_j(s) \leftarrow \text{MACRO}(\mathcal{F}_s^{(Z)}, \Theta)$ 
  end while
  Draw  $C_\delta \sim \mathbb{P}(C_\delta | M_j(s), \mathcal{F}_s^{(Z)})$ 
   $\nu_t \leftarrow \mathbb{E}[h(C_T) | C_\delta]$ 
  return  $\nu_t$ 
end function

function EPV( $N, \mathcal{F}_t^{(Z)}, \Theta$ ) ▷ Averages over simulations of next macrotransition
  Initialize  $\nu_t \leftarrow 0$ 
  for  $i$  in  $1, \dots, N$  do
     $\nu_t \leftarrow \nu_t + \text{EPVDRAW}(\mathcal{F}_t^{(Z)}, \Theta)$ 
  end for
  return  $\nu_t / N$ 
end function
```

---

a pass above 0.05, most likely because he is being fairly closely defended for someone so far from the ball, and because there are always closer passing options for the ballcarrier. Similarly, while Chris Bosh does not move much during this possession, he is most valuable as a passing option in panel C where he is closest to the basket and without any defenders in his lane. Lastly, while we estimated the probability of Lewis passing to James in panel C at 0.61 (by far Rashard Lewis' most likely passing option), we only estimated the probability of the pass from Cole to Lewis (panel B) at 0.04 (the pass actually happens a fraction of a second after the situation in panel B). More generally, our model anticipated a layup from

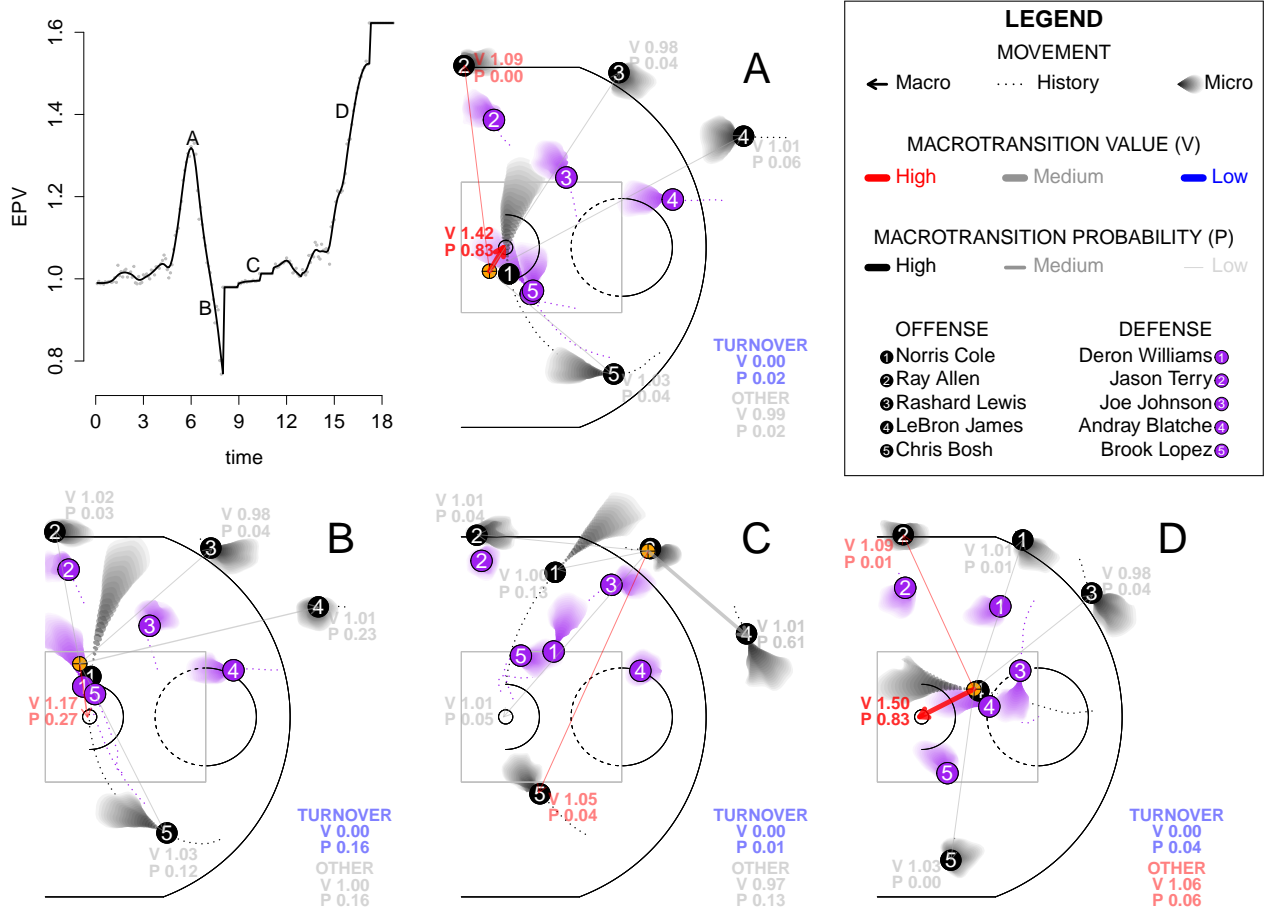


Figure 6: Detailed diagram of EPV as a function of multiresolution transition probabilities for four time points (labeled A,B,C,D) of the possession featured in Figures 1–2. Two seconds of microtransitions are shaded (with forecasted positions for short time horizons darker) while macrotransitions are represented by arrows, using color and line thickness to encode the value and probability of such macrotransitions. The value and probability of the “other” category represents the case that no macrotransition occurs during the next two seconds.

Cole, instead of his path underneath the basket and eventual pass to Lewis. These issues aside, the estimated probabilities and values of the macrotransitions highlighted in Figure 6 match well with basketball intuition.

The analysis presented here could be repeated on any of tens (hundreds) of thousands of possessions available in a season of optical tracking data. EPV plots as in Figure 2 and diagrams as in Figure 6 provide powerful insight as to how players’ movements and decisions contribute value to their team’s offense. With this insight, coaches and analysts can formulate strategies and offensive schemes that make optimal use of their players’ ability—or, defensive strategies that best suppress the motifs and situations that generate value for the opposing offense.

## 7.2 Predictive Performance of EPV

Our paper introduces EPV, and as such there are no existing results to benchmark the predictive performance of our estimates. We can, however, compare the proposed implementation for estimating EPV with simpler models, based on lower resolution information, to verify whether our model captures meaningful features of our data. Assessing the predictive performance of an EPV estimator is difficult because the estimand is a curve whose length varies by possession. Moreover, we never observe any portion of this curve; we only know its endpoint. Therefore, rather than comparing estimated EPV curves between our method and alternative methods, we compare estimated transition probabilities. For any EPV estimator method that is stochastically consistent, if the predicted transitions are properly calibrated, then the derived EPV estimates should be as well.

As mentioned in Section 6, we use only 90% of our data set for parameter inference, with the remaining 10% used to evaluate the out-of-sample performance of our model. We also evaluated out-of-sample performance of alternative macrotransition models, which use varying amounts of information from the data. Table 1 provides the out-of-sample log-likelihood for the macrotransition model applied to the 10% of the data not used in model fitting for various hazard parameterizations. The most basic parameterization assumes constant hazards for each ballcarrier/macrotransition type. We also consider a hazard mode that is unique for each ballcarrier and macrotransition, yet includes only the time-referenced covariates (situational effects) used in our full model (and no spatial effect). Finally, we consider our full model, with unique situational and spatial effects for each ballcarrier/macrotransition, both with and without the hierarchical model we use for information sharing across players. This hierarchical model is discussed in Appendix A.4. Without any shrinkage, our full model performs in some cases worse than a model with no spatial effects included, but with shrinkage, it consistently performs the best of the configurations compared—this behavior motivates the novel hierarchical structure we discuss in Appendix A, which incorporates both spatial and between-player structure in a computationally efficient manner.

Macrotransition Model				
Macrotransition	Player	Covariates	Covariates + Spatial	Full
Pass1	-29399.68	-27659.72	-27251.40	-26433.76
Pass2	-24884.99	-23691.81	-23294.98	-22226.61
Pass3	-26326.99	-25199.52	-25335.92	-23909.71
Pass4	-20426.53	-20266.06	-24487.17	-18879.47
Shot Attempt	-48885.21	-46471.49	-40914.66	-40711.55
Made Basket	-6579.31	-6626.55	-5601.75	-5284.31
Turnover	-9311.80	-9075.60	-8990.61	-8390.85

Table 1: Out of sample log-likelihood for macrotransition models (and shot probability model) under various model specifications. “Player” assumes constant hazards for each player/event type combination. “Covariates” augments this model with situational covariates,  $\mathbf{W}(t)$  as given in (7). “Covariates + Spatial” adds a spatial effect, yielding (7) in its entirety. Lastly, “Full” implements this model with the hierarchical model discussed in Appendix A.

Our comparison of the predictive performance of the competing risks macrotransition model under several different parameterizations includes a notable case. Assuming constant hazards for each player/macrotransition type is equivalent to using only the discrete, homogeneous Markov chain  $C^{(0)}, C^{(1)}, \dots, C^{(K)}$  to compute EPV, using empirical transition frequencies to estimate the transition probability matrix. The superior predictive performance of our EPV model illustrates the value in modeling the full resolution data instead of simply relying on discrete summaries.

## 8. DISCUSSION

This paper introduces a new quantity, EPV, which represents a paradigm shift in the possibilities for statistical inferences about basketball. Using high resolution, optical tracking data, EPV reveals the value in many of the schemes and motifs that characterize basketball offenses but are omitted in the box score. For instance, as diagrammed in Figures 2 and 6, we see that EPV may rise as a player attacks the basket (more so for a strong scorer like LeBron James than for a bench player like Norris Cole), passes to a well-positioned teammate, or gains separation from the defense. Aside from simply tracking changes in EPV, analysts can understand why EPV changes by expressing its value as a weighted average of transition values (as done in Figure 6). Doing so reveals that the source of a high (or low) EPV estimate may come from alternate paths of the possession that were never realized, but were probable enough to have influenced the EPV estimate—an open teammate in a good shooting location, for instance. These insights, which can be reproduced for any valid NBA possession in our data set, have the potential to reshape the way we quantify players’ actions and decisions.

We make a number of assumptions—mostly to streamline and simplify our modeling and analysis pipeline—that could be relaxed and yield a more precise model. The largest assumption is that the particular coarsened view of a basketball possession that we propose here is marginally semi-Markov. While this serves as a workable first-order approximation, there are cases that clearly violate this assumption, for example, pre-set plays that string together sequences of runs and passes. Future refinements of the model could define a wider set of macrotransitions that encapsulate these motifs, effectively encoding this additional possession structure from the coach’s playbook. A number of smaller details could also be addressed. For instance, it seems desirable to model rebound outcomes conditional on high resolution information, such as the identities and motion dynamics of potential rebounders; we do not do this, however, and use a constant probability for each team of a rebound going to either the offense or defense. We also do not distinguish between different types of turnovers (steals, bad passes, ball out of bounds, etc.), though this is due to a technical feature of our data set. Indeed, regardless of the complexity and refinement of an EPV model, we stress that the full resolution data still omits key information, such as the positioning of players’ hands and feet, their heights when jumping, and other variables which impact basketball outcomes. As such, analyses based on EPV are best accompanied by actual game film and the insight of a basketball expert.

The computational requirements of estimating EPV curves (and the parameters that generate them) likely limit EPV discussions to academic circles and professional basketball teams with access to the appropriate resources. Our model nevertheless offers a case study

whose influence extends beyond basketball. High resolution spatiotemporal data sets are an emerging inferential topic in a number of scientific or business areas, such as climate, security and surveillance, advertising, and gesture recognition. Many of the core methodological approaches in our work, such as using multiresolution transitions and hierarchical spatial models, provide insight beyond the scope of basketball to other spatiotemporal domains.



## APPENDIX A. FULL HIERARCHICAL MODEL SPECIFICATIONS

In this appendix we provide full details on parametrizing and fitting the hierarchical models for the multiresolution models introduced in Sections 4 and 5. Our intent is to present our methodology with enough specificity that it could be implemented and reproduced by readers with access to the data (the data is not publicly available) and appropriate computational resources. EPV represents a unique inferential challenge, as our data is high dimensional and the parameter space of our multiresolution transition models—including spatial random effect surfaces for all ballcarrier and macrotransition type combinations—is extremely rich. Thus, even for readers not wishing to reproduce our results, this appendix may provide a valuable example of hierarchical spatiotemporal modeling.

### A.1 Macrotransition Partial Likelihood

As discussed in Section 6, parameters for macro and microtransition models are estimated separately using partial likelihoods (14). We now focus on inference for the macrotransition model, beginning with the partial likelihood. Following (14), the competing risks model (6)–(7) specifies a partial likelihood function for all unknown model components— $\beta_j^\ell$  and  $\xi_j^\ell$  for all players  $\ell$  and macrotransitions  $j$ , as well as  $\tilde{\xi}_j^\ell$  for  $j \leq 4$ . Let  $\mathcal{T}^\ell$  comprise the time intervals for which player  $\ell$  possesses the ball, with  $\mathcal{T}_j^\ell \subset \mathcal{T}^\ell$  the time points at which a macrotransition of type  $j$  occurs. Then the (partial) likelihood can be written

$$\begin{aligned} L(\beta, \xi, \tilde{\xi}) &= \prod_{\text{possessions}} \left( \prod_{t=0}^{T-\epsilon} \mathbb{P}(M(t)^c | \mathcal{F}_t^{(Z)})^{1[M(t)^c]} \prod_{j=1}^6 \mathbb{P}(M_j(t) | \mathcal{F}_t^{(Z)})^{1[M_j(t)]} \right) \\ &= \prod_{\ell} \prod_{j=1}^6 \left( \prod_{t \in \mathcal{T}_j^\ell} \lambda_j^\ell(t) \right) \exp \left( - \int_{\mathcal{T}^\ell} \lambda_j^\ell(s) ds \right), \end{aligned} \quad (\text{A.1})$$

with  $(\beta_j^\ell, \xi_j^\ell, \tilde{\xi}_j^\ell)$  parameterizing  $\lambda_j^\ell(t)$  as in (7). This likelihood is identical to that of a model that assumes macrotransition events occur according to an inhomogeneous Poisson Process with intensity  $\lambda_j^\ell(t)$  (Laird & Olivier 1981). Notice that (A.1) factors across player-macrotransition pairs, however exact likelihood inference is impossible for this model due to the infinite-dimensional spatial effect parameters  $\xi$  and  $\tilde{\xi}$  contained in  $\lambda$ , over which we need to integrate in the second term of (A.1).

Since the observed data is discretized to every 1/25th of a second, we do not observe players' locations (and consequently, do not observe some of the time-referenced covariates) continuously, making it appropriate to replace the integral in the rightmost term in (A.1) with a sum over the finite collection of times at which data is observed. Let  $\mathcal{T}_0^\ell$  index times at which data is observed but a macrotransition does not occur (in reality, macrotransitions occur almost surely between points at which data is observed, yet our data references them to the nearest time at which locations are recorded). Then the likelihood (A.1) is approximated by

$$L(\beta, \xi, \tilde{\xi}) = \prod_{\ell} \prod_{j=1}^6 \left( \prod_{t \in \mathcal{T}_j^\ell} \lambda_j^\ell(t) \exp(-\lambda_j^\ell(t)) \right) \exp \left( - \sum_{t \in \mathcal{T}_0^\ell} \lambda_j^\ell(t) \right), \quad (\text{A.2})$$

which yields (A.1) in the limit as the temporal resolution of the data increases. While the intractable integral in (A.1) has been replaced by a sum in (A.2), evaluation of the likelihood may be extremely computationally expensive if  $|\mathcal{T}_0^\ell|$  is large (depending on the functional form assumed for the spatial random effect surface). In our data set, for some players  $\ell$ ,  $|\mathcal{T}_0^\ell|$  is as large as 300000, which makes evaluating (A.2) impossible when assuming a Gaussian process prior for  $\xi$ , as discussed in Appendix A.3.

## A.2 Covariates

As revealed in (7), the hazards  $\lambda_j^\ell(t)$  are parameterized by spatial effects ( $\xi_j^\ell$  and  $\tilde{\xi}_j^\ell$  for pass events), as well as coefficients for situation covariates,  $\beta_j^\ell$ . The covariates used may be different for each macrotransition  $j$ , but we assume for each macrotransition type the same covariates are used across players  $\ell$ .

Among the covariates we consider, **dribble** is an indicator of whether the ballcarrier has started dribbling after receiving possession. **ndef** is the distance between the ballcarrier and his nearest defender (transformed to  $\log(1+d)$ ). **ball\_lastsec** records the distance traveled by the ball in the previous one second. **closeness** is a categorical variable giving the rank of the ballcarrier’s teammates’ distance to the ballcarrier. Lastly, **open** is a measure of how open a potential pass receiver is using a simple formula relating the positions of the defensive players to the vector connecting the ballcarrier with the potential pass recipient.

For  $j \leq 4$ , the pass event macrotransitions, we use **dribble**, **ndef**, **closeness**, and **open**. For shot-taking and turnover events, **dribble**, **ndef**, and **ball\_lastsec** are included. Lastly, the shot probability model (which, from (9) has the same parameterization as the macrotransition model) uses **dribble** and **ndef** only. All models also include an intercept term.

A prior distribution is assumed for each coefficient for each macrotransition model jointly across players; we provide this in A.4.

## A.3 Spatial Effects

The parameters  $\xi_j^\ell$  for all players  $\ell$  and macrotransitions  $j \in \{1, \dots, 6\}$ , as well as  $\tilde{\xi}_j^\ell$  for  $j \leq 4$ , are infinite-dimensional as they are functions from the court space  $\mathbb{S}$  to  $\mathbb{R}$ . We assume each is a realization of a Gaussian process (sometimes called a Gaussian random field in the 2-dimensional case). Generally speaking, if  $\xi$  is a 0 mean Gaussian process with covariance function  $C(\mathbf{z}, \mathbf{z}^*)$ , then for locations  $\mathbf{z}, \mathbf{z}^* \in \mathbb{S}$ ,

$$\begin{pmatrix} \xi(\mathbf{z}) \\ \xi(\mathbf{z}^*) \end{pmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{pmatrix} C(\mathbf{z}, \mathbf{z}) & C(\mathbf{z}, \mathbf{z}^*) \\ C(\mathbf{z}^*, \mathbf{z}) & C(\mathbf{z}^*, \mathbf{z}^*) \end{pmatrix} \right). \quad (\text{A.3})$$

The form of the joint distribution (A.3) extends to arbitrarily many  $\mathbf{z} \in \mathbb{S}$ , and naturally provides interpolation and uncertainty quantification for values of the spatial field at unobserved locations (see, e.g., Rasmussen (2006)). The covariance function,  $C(\mathbf{z}, \mathbf{z}^*)$ , is called isotropic if it is a function only of  $\Delta = \|\mathbf{z} - \mathbf{z}^*\|$ . A common choice of isotropic covariance function is the Matérn, where

$$C(\Delta) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} (\kappa\Delta)^\nu K_\nu(\kappa\Delta), \quad (\text{A.4})$$

with  $K_\nu$  being the modified Bessel function of the second kind and order  $\nu > 0$ ,  $\kappa > 0$  being a scaling parameter for the distance  $\Delta$ , and  $\sigma^2$  giving the marginal variance of any point  $\xi(\mathbf{z})$ . Typically,  $\nu$  is fixed by the analyst, and  $\kappa$  and  $\sigma^2$  are estimated from the data, perhaps in a Bayesian fashion (Neal 1997). Likelihood evaluations for parameters of the covariance function are  $\mathcal{O}(n^3)$  where  $n$  is the number of spatially referenced observations associated with a particular field (assuming different parameter values for each field), which is prohibitively expensive for large data sets such as that for our macrotransition model (A.2).

An alternative to specifying a form of covariance function is to represent Gaussian processes using functional bases; that is, for  $\phi_1, \dots, \phi_d : \mathbb{S} \rightarrow \mathbb{R}$  and any  $\mathbf{z} \in \mathbb{S}$ ,

$$\xi(\mathbf{z}) = \sum_{i=1}^d w_i \phi_i(\mathbf{z}), \quad (\text{A.5})$$

with  $\mathbf{w} = (w_1 \ w_2 \ \dots \ w_d)'$  and  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ . Because the  $\{\phi_i\}$  are fixed, this yields a finite representation of the process because the map  $\xi$  is completely determined by  $\mathbf{w}$ , which offers many computational advantages (Higdon 2002; Quiñonero-Candela & Rasmussen 2005), notably that the  $\mathcal{O}(n^3)$  cost of evaluating the likelihood of the covariance parameters reduces to at worst  $\mathcal{O}(n^2 d + d^3)$ , depending on the structure of  $\mathbf{\Sigma}$ . Denoting  $\phi(\mathbf{z}) = (\phi_1(\mathbf{z}) \phi_2(\mathbf{z}) \dots \phi_d(\mathbf{z}))'$ , the representation (A.5) yields the covariance function  $C(\mathbf{z}, \mathbf{z}^*) = \phi(\mathbf{z})' \mathbf{\Sigma} \phi(\mathbf{z}^*)$ . In general, for fixed  $d$ , it is not possible to find bases  $\phi_1, \dots, \phi_d$  and covariance matrix  $\mathbf{\Sigma}$  such that the representation (A.5) is equivalent to a process specified using a Matérn covariance function, however it is possible to obtain very accurate approximations.

Following Lindgren, Rue & Lindström (2011), we assume a functional basis  $\{\phi_i, i = 1, \dots, d\}$  induced by a triangular mesh of  $d$  vertices on the court space  $\mathbb{S}$  (in practice, the triangulation is defined on a larger region that includes  $\mathbb{S}$ , due to boundary effects). The mesh is formed by partitioning  $\mathbb{S}$  into triangles, where any two triangles share at most one edge or corner (see figure 7 for an illustration). With some arbitrary ordering of the vertices of this mesh,  $\phi_i : \mathbb{S} \rightarrow \mathbb{R}$  is the unique function taking value 0 at all vertices  $j \neq i$ , 1 at vertex  $i$ , and linearly interpolating between any two points within the same triangle used in the mesh construction. Thus, with this basis, fields  $\xi$  are piecewise linear on the triangles of the mesh.

As Lindgren et al. (2011) show, there are a couple advantages to using this particular functional basis for the spatial field representation (A.5) in addition to the generic computational advantages of having a discrete representation of an infinite-dimensional parameter. The first is that it is possible to find  $\mathbf{\Sigma} = \mathbf{\Sigma}(\nu, \kappa, \sigma^2)$  (for closed-form expression, see Lindgren et al. (2011)) with  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$  such that  $\xi(\mathbf{z}) = \sum_{i=1}^d w_i \phi_i(\mathbf{z})$  closely approximates a Gaussian random field with Matérn covariance (A.4). The second advantage is that the precision  $\mathbf{\Sigma}^{-1}$  is sparse, equivalent to a conditional independence structure for  $\mathbf{w}$ .

For the spatial fields in the macrotransition model,  $\xi_j^\ell$  for all players  $\ell$  and  $j \in \{1, \dots, 6\}$ , as well as  $\tilde{\xi}_j^\ell$  for  $j \leq 4$ , we assume the representation (A.5) and the functional basis illustrated in Figure 7. Reducing the spatial effects for each log-hazard model (7) to unknown  $d$ -vectors offers many computational benefits, and eases the implementation of hierarchical models that exploit the structural variation of our model parameters across players. To introduce the appropriate notation, let  $\xi_j^\ell(\mathbf{z}) = \phi(\mathbf{z})' \mathbf{w}_j^\ell$ , where  $\mathbf{w}_j^\ell \in \mathbb{R}^d$  and  $\phi(\mathbf{z}) = (\phi_1(\mathbf{z}) \ \phi_2(\mathbf{z}) \ \dots \ \phi_d(\mathbf{z}))'$

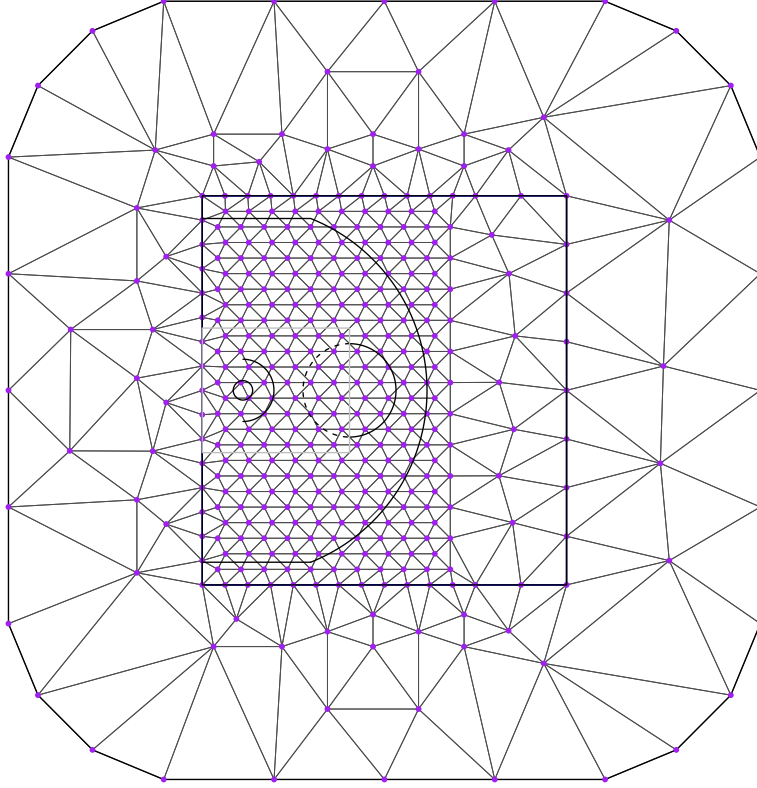


Figure 7: Triangulation of  $\mathbb{S}$  used to build the functional basis  $\{\phi_i, i = 1, \dots, d\}$ . Here,  $d = 383$ .

with  $\phi_i : \mathbb{S} \rightarrow \mathbb{R}$  for  $i = 1, \dots, d$ . Similarly define  $\tilde{\mathbf{w}}_j^\ell$  (loading the same basis functions  $\{\phi_i\}$ ). Note that the basis functions  $\phi_i$  are the same for each macrotransition  $j$  across all players, so that between-player variation in the spatial effects for each macrotransition components is provided by the weight vectors,  $\mathbf{w}_j^\ell$ .

This discrete representation of the spatial fields reduces the likelihood (A.2) to that of a Poisson regression, as

$$\log(\lambda_j^\ell(t)) = [\mathbf{W}_j^\ell(t)]' \boldsymbol{\beta}_j^\ell + \boldsymbol{\phi}(\mathbf{z}_\ell(t))' \mathbf{w}_j^\ell + \left( \tilde{\boldsymbol{\phi}}(\mathbf{z}_j(t))' \tilde{\mathbf{w}}_j^\ell \mathbf{1}[j \leq 4] \right) \quad (\text{A.6})$$

is now linear in all unknown parameters.

#### A.4 Between-Player Structure

Beyond the information in the data for estimating the unknown components of (A.6) (which are, for  $j \in \{1, \dots, 6\}$ ,  $\boldsymbol{\beta}_j^\ell$ ,  $\mathbf{w}_j^\ell$ , and for  $j \leq 4$ ,  $\tilde{\mathbf{w}}_j^\ell$ ), we have prior knowledge on their structure due to spatial smoothness and the natural clustering of players by team and by position. Our data includes position labels (e.g. center, point guard, power forward) for each player in the NBA. A player's position on a basketball team usually depends on his size and other physical attributes, as well as on his skill set and intended role in the team's strategic scheme. We therefore assume that parameters of players' macrotransition model cluster by position.

Rather than use the labeled positions in our data, we define position as a distribution of a player’s location during his time on the court. Specifically, we divide the offensive half of the court into 4-square-foot bins (575 total) and count, for each player, the number of data points for which he appears in each bin. Then we stack these counts together into a  $L \times 575$  matrix (there are  $L = 461$  players in our data), denoted  $\mathbf{G}$ , and take the square root of all entries in  $\mathbf{G}$  for normalization. We then perform non-negative matrix factorization on  $\mathbf{G}$  in order to obtain a low-dimensional representation of players’ court occupancy that still reflects variation across players (Miller et al. 2013). Specifically, this involves solving:

$$\hat{\mathbf{G}} = \underset{\mathbf{G}^*}{\operatorname{argmin}} \{D(\mathbf{G}, \mathbf{G}^*)\}, \text{ subject to } \mathbf{G}^* = \begin{pmatrix} \mathbf{U} \\ L \times r \end{pmatrix} \begin{pmatrix} \mathbf{V} \\ r \times 575 \end{pmatrix} \text{ and } U_{ij}, V_{ij} \geq 0 \text{ for all } i, j, \quad (\text{A.7})$$

where  $r$  is the rank of the approximation  $\hat{\mathbf{G}}$  to  $\mathbf{G}$  (we use  $r = 5$ ), and  $D$  is some distance function, such as a Kullback-Liebler type

$$D(\mathbf{G}, \mathbf{G}^*) = \sum_{i,j} G_{ij} \log (G_{ij}/G_{ij}^*) - G_{ij} + G_{ij}^*.$$

The rows of  $\mathbf{V}$  are non-negative basis vectors for players’ court occupancy distributions and the rows of  $\mathbf{U}$  give the loadings for each player. With this factorization,  $\mathbf{U}_i$  (the  $i$ th row of  $\mathbf{U}$ ) provides player  $i$ ’s “position”—a  $r$ -dimensional summary of where he spends his time on the court. Moreover, the smaller the difference between two players’ positions,  $\|\mathbf{U}_i - \mathbf{U}_j\|$ , the more alike are their roles on their respective teams, and the more similar we expect the parameters of their macrotransition models to be a priori.

Formalizing this, let  $\mathbf{H}$  be a  $L \times L$  matrix consisting of 0s, then set  $H_{ij} = 1$  if player  $j$  is one of the eight closest players in our data to player  $i$  using the distance  $\|\mathbf{U}_i - \mathbf{U}_j\|$  (the cutoff of choosing the closest eight players is arbitrary). This construction of  $\mathbf{H}$  does not guarantee symmetry, which is required for the expressions that follow, thus we set  $H_{ji} = 1$  if  $H_{ij} = 1$ . Let  $n_i = \sum_{j=1}^L H_{ij}$  count the number of neighbors for player  $i$ . For any parameter  $\boldsymbol{\theta}$ , with  $\theta_i$  being the value for player  $i$ , we assume a conditional autoregressive model (CAR) (Besag 1974):

$$\theta_i | \theta_{(-i)}, \tau \sim \mathcal{N} \left( \frac{1}{n_i} \sum_{j: H_{ij}=1} \theta_j, \frac{\tau^2}{n_i} \right). \quad (\text{A.8})$$

This can be expressed as a joint distribution for  $\theta$ ,

$$P(\boldsymbol{\theta} | \tau) \propto \tau^{-L} \exp \left( -\frac{1}{2\tau^2} \sum_{i,j: H_{ij}=1} (\theta_i - \theta_j)^2 \right), \quad (\text{A.9})$$

which is improper ((A.9) is constant for constant location shifts of  $\boldsymbol{\theta}$ ), though as a prior distribution will yield a proper posterior in typical applied settings (Besag, York & Mollié 1991). While  $\mathbf{H}$  derives from the data and may therefore seem problematic as a component of a prior distribution,  $\mathbf{H}$  is ancillary for the macrotransition parameters. As seen in (A.1), the macrotransition likelihood conditions on players’ locations, modeling only their decisions

at the locations they occupy. The CAR prior thus plays a key role in our estimation of the parameters of the macrotransition model, allowing for information sharing between players and easing inference for situations in which player-specific data is sparse.

#### A.5 Parameter Estimation for the Macrotransitions

With the likelihood (A.2), finite representation for the Gaussian random fields (A.5), and prior form (A.9) introduced, we now connect these components together and discuss inference.

Each unknown coefficient  $\beta$  of the macrotransition model is assumed the CAR prior structure. Specifically, we assume the vector  $\boldsymbol{\beta}_{j,i} = (\beta_{j,i}^1 \ \beta_{j,i}^2 \ \dots \ \beta_{j,i}^L)'$  has the distribution given in (A.9) for the generic parameter  $\boldsymbol{\theta}$ , and that  $\{\boldsymbol{\beta}_{j,i}, i = 1, \dots, p_j \text{ and } j = 1, \dots, 6\}$  are independent a priori. We may analogously define  $\mathbf{w}_{j,i} = (w_{j,i}^1 \ w_{j,i}^2 \ \dots \ w_{j,i}^L)'$ , the vector of the loadings across players for the  $i$ th basis function for the  $j$ th macrotransition model's spatial field. While we may expect, a priori, components of this vector to covary according to a CAR structure, we also assume spatial covariance among the loadings for any particular player-macrotransition model, i.e.,  $\mathbf{w}_j^\ell \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}(1, \kappa_j, \sigma_j^2))$  as discussed in Appendix A.3. This suggests a Kronecker structure for the prior covariance of the basis loadings across players and space.

The computation demands of fitting such a model are prohibitive. Even assuming prior independence of all parameters across macrotransition types,  $\mathbf{w}_j = (\mathbf{w}_j^1 \ \dots \ \mathbf{w}_j^L)'$  would enter the model as a  $Ld$ -dimensional random effect ( $Ld = 176563$  in our current specification), with an unknown covariance matrix itself parameterized by  $\tau_j^2, \kappa_j, \sigma_j^2$ . The conditional independence structure implied by the CAR model (A.8) does not hold when spatial structure is included, as all components of  $\mathbf{w}_j^\ell$  for all  $\ell$  depend on  $\kappa_j$  and  $\sigma_j^2$ . Coarsening the mesh that induces the functional basis  $\boldsymbol{\phi}$  reduces  $d$ , relieving some computational burden, but also impacting our ability to detect small-scale spatial variation.

Our approach is to consider  $h$ -vectors  $\mathbf{v}_j^\ell \in \mathbb{R}^h$  where the  $i$ th component is a linear combination of the  $d$  components of  $\mathbf{w}_j^\ell$ :  $v_{j,i}^\ell = \sum_{m=1}^d a_{i,m}^j w_{j,m}^\ell$ . The idea here is that  $\mathbf{v}_j^\ell$  is an  $h$ -dimensional representation of the  $d$ -dimensional vector of loadings  $\mathbf{w}_j^\ell$ . With a rough estimate of the matrix  $\mathbf{w}_j$ , the weights  $a_{i,m}^j$  can be estimated by matrix factorization, such as SVD or NMF. This is exactly the route we take. For each macrotransition  $j$ , we estimate  $\mathbf{w}_j^\ell$  independently for each player  $\ell$  by fitting the  $j$ th macrotransition model (A.2) and (A.6) using the spatial prior  $\mathbf{w}_j^\ell \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\nu = 1, \kappa, \sigma^2))$  and vague priors on  $\boldsymbol{\beta}_j^\ell, \kappa, \sigma^2$ . This was done using the software R-INLA ([www.r-inla.org](http://www.r-inla.org)), which uses integrated nested Laplace approximations (Rue, Martino & Chopin 2009) for approximate Bayesian inference for generalized linear models with latent Gaussian Markov parameters, along with possibly non-Gaussian hyperparameters. Estimates  $\mathbf{w}_j^1, \mathbf{w}_j^2, \dots, \mathbf{w}_j^L$  were stacked on top of each other as row vectors to form  $\mathbf{w}_j$ , which was then exponentiated and factored using NMF with KL loss for  $h = 10$  ( $r = 10$  in the notation of (A.7)). Using NMF instead of SVD gives slightly better out-of-sample predictive results for the final estimated macrotransition model; note also that components of  $\mathbf{w}_j$  enter the likelihood (and log-likelihood) exponentiated.

Our construction of  $\mathbf{v}_j^\ell$  implies a new functional basis representation for  $\xi_j^\ell$ . Since  $v_{j,i}^\ell = \sum_{m=1}^d a_{i,m}^j w_{j,m}^\ell$ , for any  $\mathbf{z} \in \mathbb{S}$  and  $i \in \{1, \dots, h\}$  we may write  $\psi_{j,i}(\mathbf{z}) = \sum_{m=1}^d a_{i,m}^j \phi_m(\mathbf{z})$ .

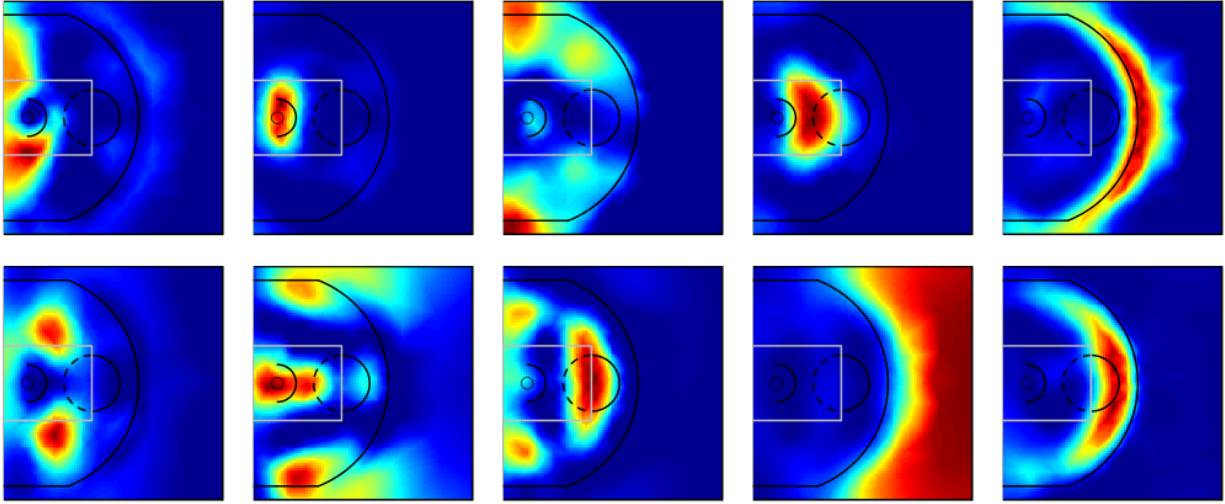


Figure 8: The functional bases  $\psi_{j,i}$  for  $i = 1, \dots, 10$  and  $j$  corresponding to the shot-taking macrotransition. Unlike SVD, there is no interpretation of the ordering of the bases. These bases correspond to well-known basketball motifs, such as layups (2), three point shots (3 and 5), and perimeter shots (6, 7, and 10).

$\boldsymbol{\psi}_j = (\psi_{j,1} \psi_{j,2} \dots \psi_{j,h})'$  now provides a new functional basis corresponding to the loadings  $\mathbf{v}_j^\ell$  for each player  $\ell$ —note that after incorporating the weights  $a_{i,m}^j$  into the basis functions, we no longer tether the definition of  $\mathbf{v}_j^\ell$  to these weights. Correspondingly, the basis functions  $\boldsymbol{\psi}_j$  depend on  $j$ —meaning they differ by macrotransition type—unlike the basis functions  $\boldsymbol{\phi}$ . Similar to Miller et al. (2013), these functional bases allow for spatial fields that are not simply locally smooth, but smooth across regions where players employ similar strategies—for an illustration of this, see Figure 8. Thus, they represent information sharing across players and across space. Note that for pass events, we analogously construct  $\tilde{\boldsymbol{\psi}}_j$  and  $\tilde{\mathbf{v}}_j^\ell$  for all  $\ell$  and  $j \leq 4$ .

Besides dimension reduction, another advantage to the new functional basis  $\boldsymbol{\psi}_j$  is that, due to obtaining weights  $a_{i,m}^j$  by NMF, the components of  $\mathbf{v}_j^\ell$  may be assumed to be uncorrelated a priori<sup>3</sup>, both within and across players  $\ell$ . The linear combinations that comprise  $\boldsymbol{\psi}_j$  already provide spatial covariation. This allows us to model  $\mathbf{v}_{j,i} = (v_{j,i}^1 \ v_{j,i}^2 \ \dots \ v_{j,i}^L)'$  identically to other model parameters  $\boldsymbol{\beta}_{j,i}$ ; namely, we assume  $\mathbf{v}_{j,i}$  have prior structure given by (A.9), and that  $\mathbf{v}_{j,i}$  are a priori independent across  $j$  and  $i$  (as are  $\tilde{\mathbf{v}}_{j,i}$ ).

To summarize, we rewrite the model components introduced in this section and concisely

<sup>3</sup>This is not a theoretical property of NMF, but approximately holds due to the similarity in results from NMF and SVD.

reveal the form of the posterior:

$$\begin{aligned}
(\text{likelihood}) \quad L(\boldsymbol{\beta}, \mathbf{v}, \tilde{\mathbf{v}}) &= \prod_{\ell} \prod_{j=1}^6 \left( \prod_{t \in \mathcal{T}_j^{\ell}} \lambda_j^{\ell}(t) \right) \exp \left( - \int_{\mathcal{T}^{\ell}} \lambda_j^{\ell}(s) ds \right) \\
\text{where} \quad \log(\lambda_j^{\ell}(t)) &= [\mathbf{W}_j^{\ell}(t)]' \boldsymbol{\beta}_j^{\ell} + \boldsymbol{\psi}_j(\mathbf{z}_{\ell}(t))' \mathbf{v}_j^{\ell} + \left( \tilde{\boldsymbol{\psi}}_j(\mathbf{z}_j(t))' \tilde{\mathbf{v}}_j^{\ell} \mathbf{1}[j \leq 4] \right), \\
(\boldsymbol{\beta} \text{ prior}) \quad P(\boldsymbol{\beta}_{j,i} | \tau_{j,i}^{\beta}) &\propto \left( \tau_{j,i}^{\beta} \right)^{-L} \exp \left( - \frac{1}{2(\tau_{j,i}^{\beta})^2} \sum_{q,r: H_{qr}=1} (\beta_{j,i}^q - \beta_{j,i}^r)^2 \right) \\
&\quad \text{for } i = 1, \dots, p_j \text{ and } j = 1, \dots, 6, \\
(\mathbf{v} \text{ prior}) \quad P(\mathbf{v}_{j,i} | \tau_{j,i}^v) &\propto \left( \tau_{j,i}^v \right)^{-L} \exp \left( - \frac{1}{2(\tau_{j,i}^v)^2} \sum_{q,r: H_{qr}=1} (v_{j,i}^q - v_{j,i}^r)^2 \right) \\
&\quad \text{for } i = 1, \dots, h \text{ and } j = 1, \dots, 6, \\
(\tilde{\mathbf{v}} \text{ prior}) \quad P(\tilde{\mathbf{v}}_{j,i} | \tau_{j,i}^{\tilde{v}}) &\propto \left( \tau_{j,i}^{\tilde{v}} \right)^{-L} \exp \left( - \frac{1}{2(\tau_{j,i}^{\tilde{v}})^2} \sum_{q,r: H_{qr}=1} (\tilde{v}_{j,i}^q - \tilde{v}_{j,i}^r)^2 \right) \\
&\quad \text{for } i = 1, \dots, h \text{ and } j = 1, \dots, 4, \\
\text{and } (\tau \text{ prior}) \quad P(\tau_{j,i}^*) &\propto \left( \tau_{j,i}^* \right)^{-2} \exp \left( \frac{-1}{\tau_{j,i}^*} \right) \text{ for all } j, i, \text{ and } * = \beta, v, \tilde{v}. \quad (\text{A.10})
\end{aligned}$$

This looks daunting, but it is just a mixed-effects Poisson regression, with a very particular structure for the random effects. However, due to the scale of the data, considerable computational resources are needed for inference. The posterior factors across macrotransition models  $j$ , thus we can estimate components corresponding to each macrotransition model separately. Each macrotransition model has over 10 million data points entering the likelihood (this is the size of  $\sum_{\ell} (|\mathcal{T}_j^{\ell}| + |\mathcal{T}_0^{\ell}|)$ ), and at least 6006 parameters, corresponding to  $(L+1)(p_j+h)$  for  $L=461$ ,  $p_j=3$  (all macrotransitions have  $p_j \geq 3$ ), and  $h=10$  (the  $L+1$  term accounts for  $\tau$  as well). Approximate Bayesian inference was performed using R-INLA.

## A.6 Parameter Estimation Microtransitions

Both  $\mu_x$  and  $\mu_y$  are assumed to be realizations of Gaussian processes with Matérn covariance (A.4) with  $\nu=1$ , though approximated with the functional basis (A.5) used in the macrotransition model and illustrated in Figure 7. Like the spatial fields in the macrotransition model, we use R-INLA to fit the microtransition model (12) independently for  $x(t)$  and  $y(t)$ . Note that for each player, separate models are fit to predict his motion during times he is the ballcarrier and times he is not the ballcarrier but still on offense. There are thus  $L \times 2 \times 2$  microtransition models of the form (12)—one for each of  $L$  players, two situations (carrying ball and not carrying ball) and two dimensions ( $x$  and  $y$ ). There is no information sharing between models; in principle we can imagine different components are connected a priori, yet the data is so informative that any appropriate prior would not be influential. Unlike the macrotransition model, where we may only observe a handful of macrotransitions depend-



ing on the player and macrotransition type (e.g. turnovers are fairly rare for all players), all players are constantly moving, so the data alone are sufficient for precise inference.

## APPENDIX B. EPV-DERIVED QUANTITIES

While detailed studies of EPV curves and multiresolution transitions, as in Section 7, are the most immediate and impactful application of EPV, we may also consider metrics that aggregate a season’s worth of EPV curves.

### B.1 EPV-Added

*EPV-Added* (EPVA) quantifies a player’s overall offensive value of all of his movements and decisions while handling the ball, relative to the estimated value contributed by a league-average player receiving ball possession in the same situations. The notion of *relative* value is important because the martingale structure of EPV ( $\nu_t$ ) prevents any meaningful aggregation of EPV across a specific player’s possessions; for instance,  $\mathbb{E}[\nu_t - \nu_{t+\epsilon}] = 0$  for all  $t$ , meaning that *on average* EPV does not change during any specific player’s ball handling. For instance, while we see the EPV skyrocket after LeBron James receives the ball and eventually attack the basket in Figure 2, the definition of EPV prevents such increases being observed on average. James does not always attack the basket given the spatial situation he encountered when receiving the ball, and even when he does, he does not always beat the defense and gain a clear lane to the basket.

To calculate the baseline EPV at any time point for a league average player, we start by considering an alternate version of the transition probability matrix between coarsened states  $\mathbf{P}$ . For each player  $\ell_1, \dots, \ell_5$  on offense, there is a disjoint subset of rows of  $\mathbf{P}$ , denoted  $\mathbf{P}_{\ell_i}$ , that correspond to possession states for player  $\ell_i$ . Each row of  $\mathbf{P}_{\ell_i}$  is a probability distribution over transitions in  $\mathcal{C}$  given possession in a particular state. Technically, since states in  $\mathcal{C}_{\text{poss}}$  encode player identities, players on different teams do not share all states which they have a nonzero probability of transitioning to individually. To get around this, we remove the columns from each  $\mathbf{P}_{\ell_i}$  corresponding to passes to players not on player  $\ell_i$ ’s team, and reorder the remaining columns according to the position (guard, center, etc.) of the associated pass recipient. Thus, the interpretation of transition distributions  $\mathbf{P}_{\ell_i}$  across players  $\ell_i$  is as consistent as possible. We create a baseline transition profile of a hypothetical league-average player by averaging these across all players: (with slight abuse of notation) let  $\mathbf{P}_r = \sum_{\ell=1}^L \mathbf{P}_{\ell}/L$ . Using this, we create a new transition probability matrix  $\mathbf{P}_r(\ell)$  by replacing player  $\ell$ ’s transition probabilities ( $\mathbf{P}_{\ell}$ ) with the league-average player’s ( $\mathbf{P}_r$ ). The baseline (league-average) EPV at time  $t$  is then found by evaluating  $\mathbb{E}_{\mathbf{P}_r(\ell)}[h(C_T)|C_t]$ , which contrasts with the coarsened approximation,  $\mathbb{E}_{\mathbf{P}}[h(C_T)|C_t]$ , to  $\nu_t$ . Denote this baseline (coarsened) EPV  $\nu_t^{r(\ell)} = \mathbb{E}_{\mathbf{P}_r(\ell)}[h(C_T)|C_t]$ .

If player  $\ell$  has possession of the ball at time  $t_1$  until time  $t_2$ , the quantity  $\nu_{t_2} - \nu_{t_1}^{r(\ell)}$  estimates the value contributed player by  $\ell$  relative to a league-average player during his ball possession. We calculate EPVA for player  $\ell$  (EPVA( $\ell$ )) by summing such differences over all a player’s touches (and dividing by the number of games played by player  $\ell$  to provide

standardization):

$$\text{EPVA}(\ell) = \frac{1}{\# \text{ games for } \ell} \sum_{\{t_s, t_e\} \in \mathcal{T}^\ell} \nu_{t_e} - \nu_{t_s}^{r(\ell)} \quad (\text{A.11})$$

where  $\mathcal{T}^\ell$  contains all intervals of form  $[t_s, t_e]$  that span player  $\ell$ 's ball possession.

It must first be noted that for any  $[t_s, t_e] \in \mathcal{T}^\ell$ ,  $\mathbb{E}[\nu_{t_e} - \nu_{t_s}^{r(\ell)}] = \mathbb{E}[\nu_{t_s} - \nu_{t_s}^{r(\ell)}]$  due to  $\nu_t$  being a martingale. However, as defined, EPVA( $\ell$ ) sums over terms with more variation, since  $\nu_t$  is more variable later in the possession. This is helpful for identifying particular plays in which players accrue very high (low) EPVA. Secondly, the choice to average over games implicitly rewards players who have high usage, even if their value added per touch might be low. Often, one-dimensional offensive players accrue the most EPVA per touch since they only handle the ball when they are uniquely suited to scoring; for instance, some centers (such as Miami's Chris Andersen) only receive the ball right next to the basket, where their height offers a considerable advantage for scoring over other players in the league. Thus, averaging by game—not touch—balances players' efficiency per touch with their usage and importance in the offense. Lastly, using coarsened EPV as a baseline  $\nu_t^{r(\ell)}$  exploits the fact that, when averaging possessions over the entire season, the results are (in expectation) identical to using full-resolution EPV, assuming corresponding multiresolution transition probability models for this hypothetical league-average player—a consequence of (10).

Table 2 provides a list of the top and bottom 10 ranked players by EPVA using our 2013-14 data, which is complete until February 7, 2014. Generally, players with high EPVA effectively adapt their decision-making process to the spatiotemporal circumstances they inherit when gaining possession. They receive the ball in situations that are uniquely suited to their abilities, so that on average the rest of the league is less successful in these circumstances. Players with lower EPVA are not necessarily “bad” players in any conventional sense; their actions simply tend to lead to fewer points than other players given the same options. Of course, EPVA provides a limited view of a player's overall contributions since it does not quantify players' actions on defense, or other ways that a player may impact EPV while not possessing the ball (though EPVA could be extended to include these aspects).

As such, we stress the idea that EPVA is not a best/worst players in the NBA ranking. Analysts should also be aware that the league-average player being used as a baseline is completely hypothetical, and we heavily extrapolate our model output by considering value calculations assuming this nonexistent player possessing the ball in all the situations encountered by an actual NBA player. The extent to which such an extrapolation is valid is a judgment a basketball expert can make. Alternatively, one can consider EPV-added over *specific* players (assuming player  $\ell_2$  receives the ball in the same situations as player  $\ell_1$ ), using the same framework developed for EPVA. Such a quantity may actually be more useful, particularly if the players being compared play similar roles on their teams and face similar situations (and the degree of extrapolation is minimized).

## B.2 Shot Satisfaction

Another EPV-derived player metric we consider is called *shot satisfaction*. For each shot attempt a player takes, we wonder how satisfied the player is with his decision to shoot—what was the expected point value of his most reasonable passing option at the time of the shot?

Player	EPVA	Player	EPVA
Dirk Nowitzki	6.08	Ricky Rubio	-0.07
Kevin Durant	6.08	Luke Ridnour	0.18
Jose Calderon	5.33	Tayshaun Prince	0.26
Damian Lillard	5.28	Shaun Livingston	0.38
Kevin Love	5.13	Beno Udrih	0.47
Stephen Curry	4.63	P.J. Tucker	0.55
Channing Frye	4.58	Al-Farouq Aminu	0.59
Kyle Lowry	4.50	Andre Miller	0.68
Paul George	4.40	Gerald Henderson	0.71
LeBron James	4.38	Cody Zeller	0.71

Table 2: Top 10 and bottom 10 players by EPV-added (EPVA) in 2013-14 (per game, minimum 500 touches during season).

If for a particular player, the EPV measured at his shot attempts is higher than the EPV conditioned on his possible passes at the same time points, then by shooting the player is consistently making the best decision for his team. On the other hand, players with pass options at least as valuable as shots should regret their shot attempts (we term “satisfaction” as the opposite of regret) as passes in these situations have higher expected value.

Player	Shot Satisfaction	Player	EPVA
Jose Calderon	0.34	Ricky Rubio	-0.01
Martell Webster	0.34	Tayshaun Prince	0.00
Spencer Hawes	0.33	DeMar DeRozan	0.03
Andre Iguodala	0.33	LaMarcus Aldridge	0.04
Channing Frye	0.32	Tyreke Evans	0.05
Kyle Lowry	0.32	Shaun Livingston	0.05
Mike Miller	0.31	Gerald Henderson	0.05
Marvin Williams	0.31	Kevin Garnett	0.06
Kyle Korver	0.30	Jarrett Jack	0.06
Jodie Meeks	0.29	Anthony Davis	0.07

Table 3: Top 10 and bottom 10 players by shot satisfaction in 2013-14 (per game, minimum 500 touches during season).

Specifically, we calculate

$$\text{SATIS}(\ell) = \frac{1}{|\mathcal{T}_{\text{shot}}^\ell|} \sum_{t \in \mathcal{T}_{\text{shot}}^\ell} \nu_t - \mathbb{E} \left[ h(C_T) \mid \bigcup_{j=1}^4 M_j(t) \right] \quad (\text{A.12})$$

where  $\mathcal{T}_{\text{shot}}^\ell$  indexes times a shot attempt occurs,  $\{t : M_5(t)\}$ , for player  $\ell$ . Recalling that macrotransitions  $j = 1, \dots, 4$  correspond to pass events (and  $j = 5$  a shot attempt),

$\bigcup_{j=1}^4 M_j(t)$  is equivalent to a pass happening in  $(\epsilon, t + \epsilon]$ . Unlike EPVA, pass satisfaction  $\text{SATIS}(\ell)$  is expressed as an average per shot (not per game), which favors player such as three point specialists, who often take fewer shots than their teammates, but do so in situations where their shot attempt is extremely valuable. Table 3 provides the top/bottom 10 players in shot satisfaction for our 2013-14 data. While players who attempt many three-pointers (e.g. Calderon, Miller, Korver) and/or players shots near the basket (e.g. Iguodala) have the most shot satisfaction, players who primarily take mid-range or long-range two pointers (e.g. Aldridge, Garnett) or poor shooters (e.g. Rubio, Prince) have the least. However, because almost all shot satisfaction numbers are positive, players still shoot relatively efficiently—almost every player generally helps his team by shooting rather than passing in the same situations, though some players do so more than others.

## REFERENCES

- Besag, J. (1974), “Spatial Interaction and the Statistical Analysis of Lattice Systems,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2), 192–236.
- Besag, J., York, J., & Mollié, A. (1991), “Bayesian Image Restoration, with two Applications in Spatial Statistics,” *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20.
- Bukiet, B., Harold, E. R., & Palacios, J. L. (1997), “A Markov Chain Approach to Baseball,” *Operations Research*, 45(1), 14–23.
- Burke, B. (2010), “Win Probability Added (WPA) Explained,” *www.advancedfootballanalytics.com*, (website).
- Cox, D. R. (1975a), “A Note on Partially Bayes Inference and the Linear Model,” *Biometrika*, 62(3), 651–654.
- Cox, D. R. (1975b), “Partial Likelihood,” *Biometrika*, 62(2), 269–276.
- Franks, A., Miller, A., Bornn, L., & Goldsberry, K. (2014), “Characterizing the Spatial Structure of Defensive Skill in Professional Basketball,” *arXiv preprint*, arXiv:1405.0231.
- Goldner, K. (2012), “A Markov Model of Football: Using Stochastic Processes to Model a Football Drive,” *Journal of Quantitative Analysis in Sports [online]*, 8(1).
- Higdon, D. (2002), “Space and Space-Time Modeling Using Process Convolutions,” in *Quantitative Methods for Current Environmental Issues*, New York, NY: Springer, pp. 37–56.
- Hollinger, J. (2005), *Pro Basketball Forecast, 2005-06*, Washington, D.C: Potomac Books.
- Ihler, A., Hutchins, J., & Smyth, P. (2006), “Adaptive Event Detection with Time-Varying Poisson Processes,” in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY: ACM, pp. 207–216.
- Laird, N., & Olivier, D. (1981), “Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques,” *Journal of the American Statistical Association*, 76(374), 231–240.

- Lee, T. C., Judge, G., & Zellner, A. (1968), “Maximum Likelihood and Bayesian Estimation of Transition Probabilities,” *Journal of the American Statistical Association*, 63(324), 1162–1179.
- Lindgren, F., Rue, H., & Lindström, J. (2011), “An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: the Stochastic Partial Differential Equation Approach,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 73(4), 423–498.
- Meshkani, M. R., & Billard, L. (1992), “Empirical Bayes Estimators for a Finite Markov Chain,” *Biometrika*, 79(1), 185–193.
- Miller, A., Bornn, L., Adams, R., & Goldsberry, K. (2013), “Factorized Point Process Intensities: A Spatial Analysis of Professional Basketball,” in *Proceedings of the 31st International Conference on Machine Learning*, pp. 235–243.
- Neal, R. M. (1997), “Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification,” *arXiv preprint*, arXiv:physics/9701026.
- Omidiran, D. (2011), “A New Look at Adjusted Plus/Minus for Basketball Analysis,” *MIT Sloan Sports Analytics Conference [online]*, 2011.
- Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V., & Breslow, N. (1978), “The Analysis of Failure Times in the Presence of Competing Risks,” *Biometrics*, pp. 541–554.
- Quiñonero-Candela, J., & Rasmussen, C. E. (2005), “A Unifying View of Sparse Approximate Gaussian Process Regression,” *The Journal of Machine Learning Research*, 6, 1939–1959.
- Rasmussen, C. E. (2006), *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT Press.
- Rue, H., Martino, S., & Chopin, N. (2009), “Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 71(2), 319–392.
- Shao, X., & Li, L. (2011), “Data-Driven Multi-Touch Attribution Models,” in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY: ACM, pp. 258–264.
- Thomas, A., Ventura, S. L., Jensen, S. T., & Ma, S. (2013), “Competing Process Hazard Function Models for Player Ratings in Ice Hockey,” *The Annals of Applied Statistics*, 7(3), 1497–1524.
- Wong, W. H. (1986), “Theory of Partial Likelihood,” *The Annals of Statistics*, pp. 88–123.
- Yang, T. Y., & Swartz, T. (2004), “A Two-Stage Bayesian Model for Predicting Winners in Major League Baseball,” *Journal of Data Science*, 2(1), 61–73.