Supplemental Material for Factorized Point Process Intensities: A Spatial Analysis of Professional Basketball

Andrew Miller ACM@SEAS.HARVARD.EDU
School of Engineering and Applied Sciences, Harvard University, Cambridge, USA

Luke Bornn BORNN@STAT.HARVARD.EDU
Department of Statistics, Harvard University, Cambridge, USA

Ryan Adams RPA@SEAS.HARVARD.EDU
School of Engineering and Applied Sciences, Harvard University, Cambridge, USA

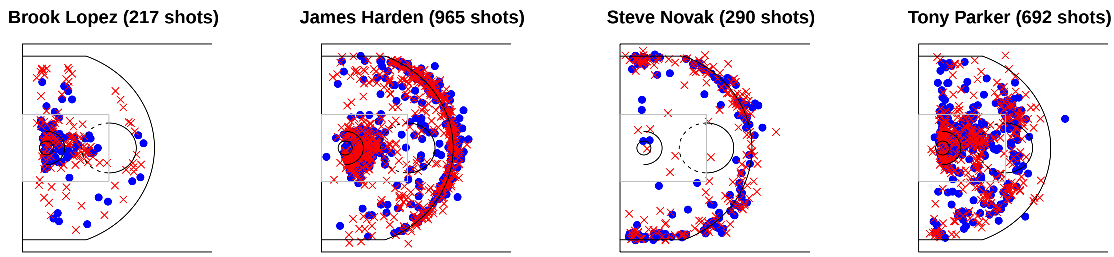
Kirk Goldsberry KGOLDSBERRY@FAS.HARVARD.EDU
Center for Geographic Analysis, Harvard University, Cambridge, USA

We use this space to present more visualizations produced by our method. First, we show a few more examples of individual player's raw data. We select a few players that differ in their habits and show how the LGCP-NMF method approximates their intensity surface. Shot charts, grids, LGCPs and LGCP-NMF reconstructions are in Figure 1.

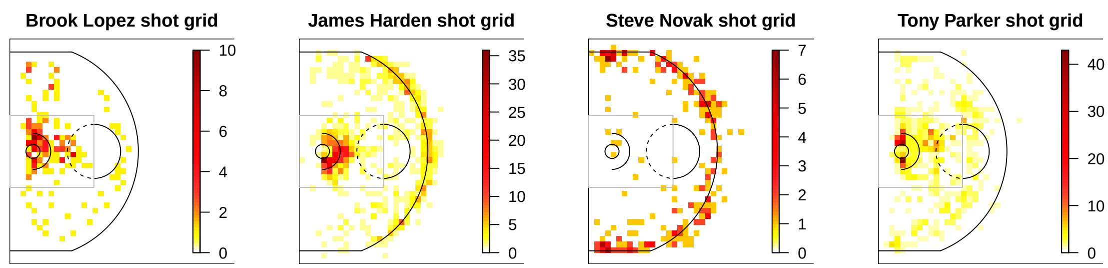
Secondly, we examine basis surfaces as a function of K. Figure 2 shows how LGCP-NMF naturally finds short, middle, and long range shot types when $K = 3$. Figures 3 and 4 visualize how those regions of the court are refined as K grows to 5 and then 10.

Figure 5 presents more empirical correlation visualizations, which allow us to examine how players tend habits tend to correlate over space.

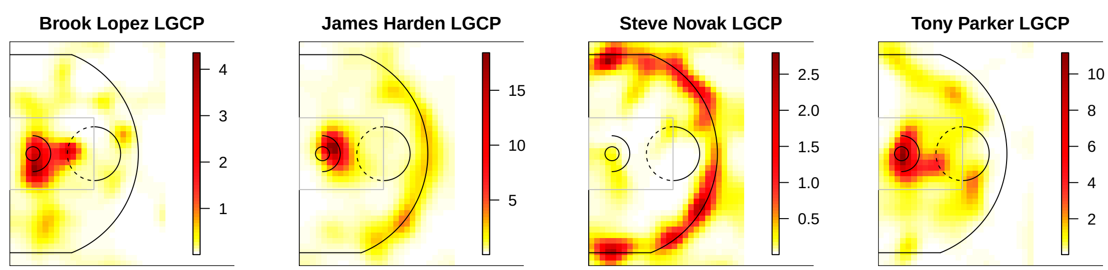
The remaining figures provide a breakdown of the spatial field goal percentage surfaces for a handful of players. We include posterior summaries for the global parameters $\beta_{0,1:K}$, $\sigma^2_{1:K}$, as well as posterior summaries for individual player parameters, $\beta_{\cdot,1:K}$. We also present the player's empirical shot chart for comparison.



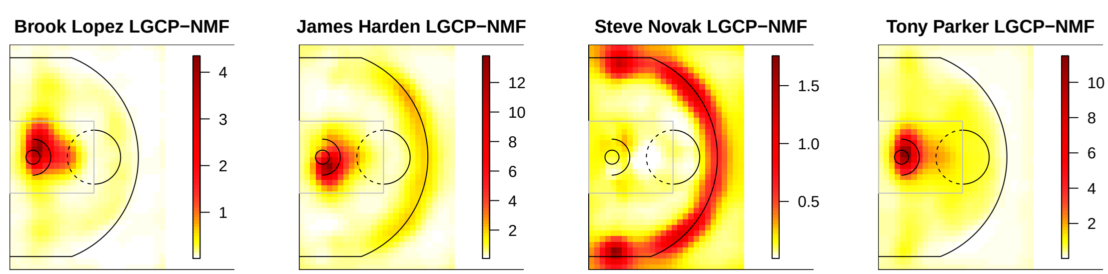
(a) points



(b) grid

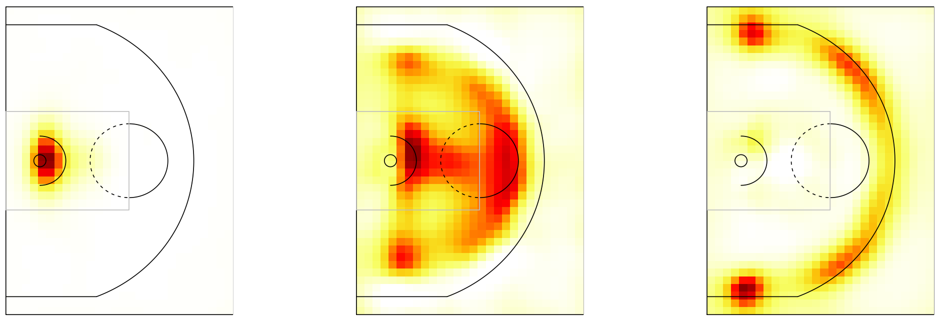


(c) LGCP



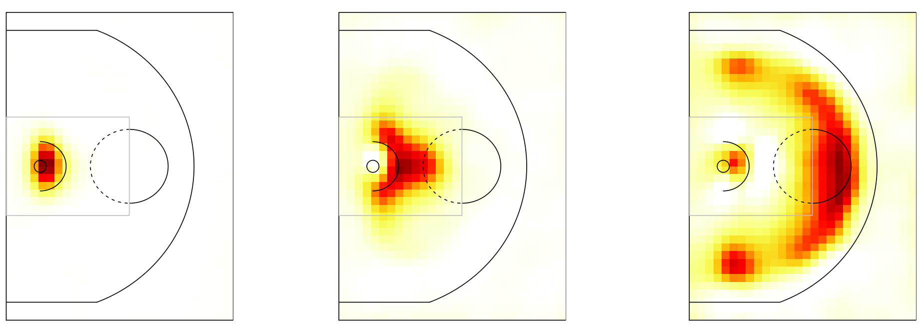
(d) LGCP-NMF

Figure 1. More examples of NBA shooters. Note the wide range of variation in shooting habits. The top two rows are the point process and discretized counts. The third row shows individual player LGCPs, and the fourth row shows the low rank reconstruction of each LGCP surface for $K = 10$ shot types. Made and missed shots are represented as blue circles and red $\times$'s, respectively.

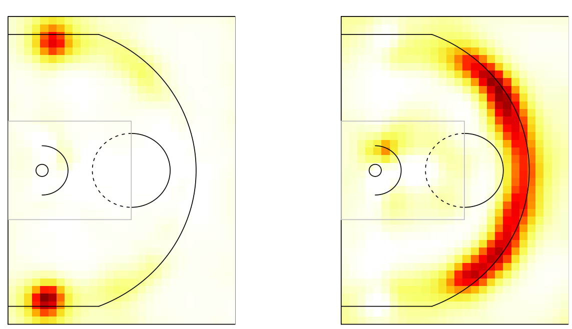


(a) short (b) mid (c) long

Figure 2. Basis given by LGCP-NMF for $K = 3$ shot types. LGCP-NMF immediately distinguishes between short range, mid range and long range shots.

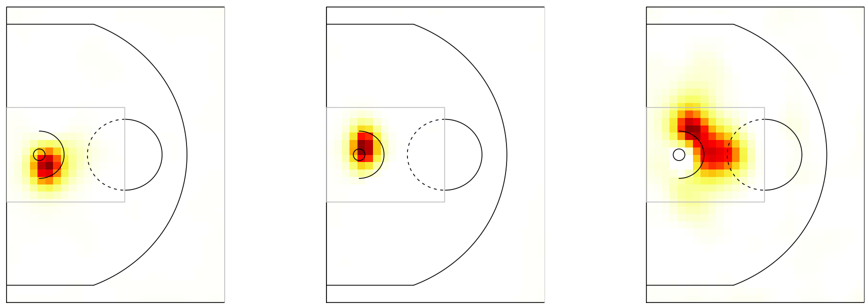


(a) short (b) mid

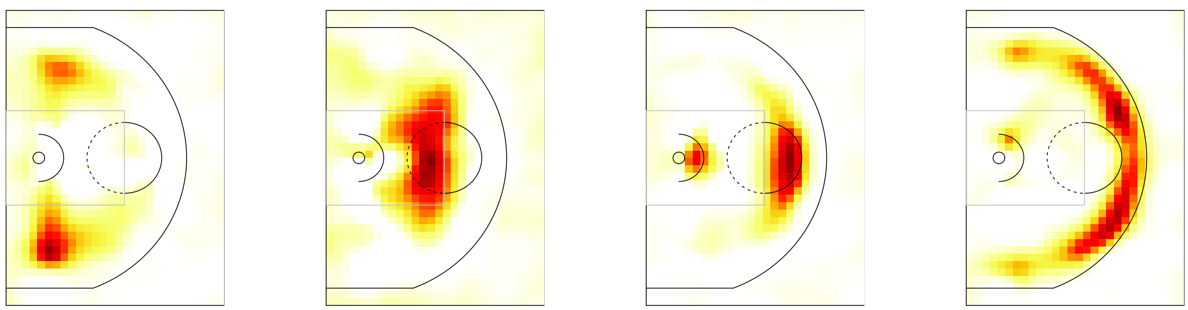


(c) long

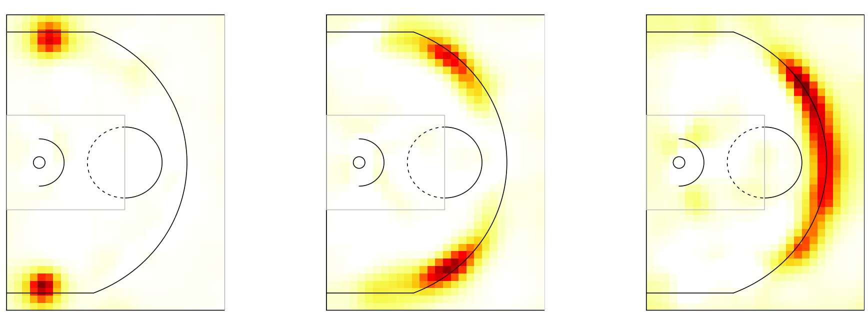
Figure 3. $K = 5$ basis vectors. LGCP-NMF separates short range shots into 'dunks' and layups. Long range shots have been separated into corner threes and other threes.



(a) short

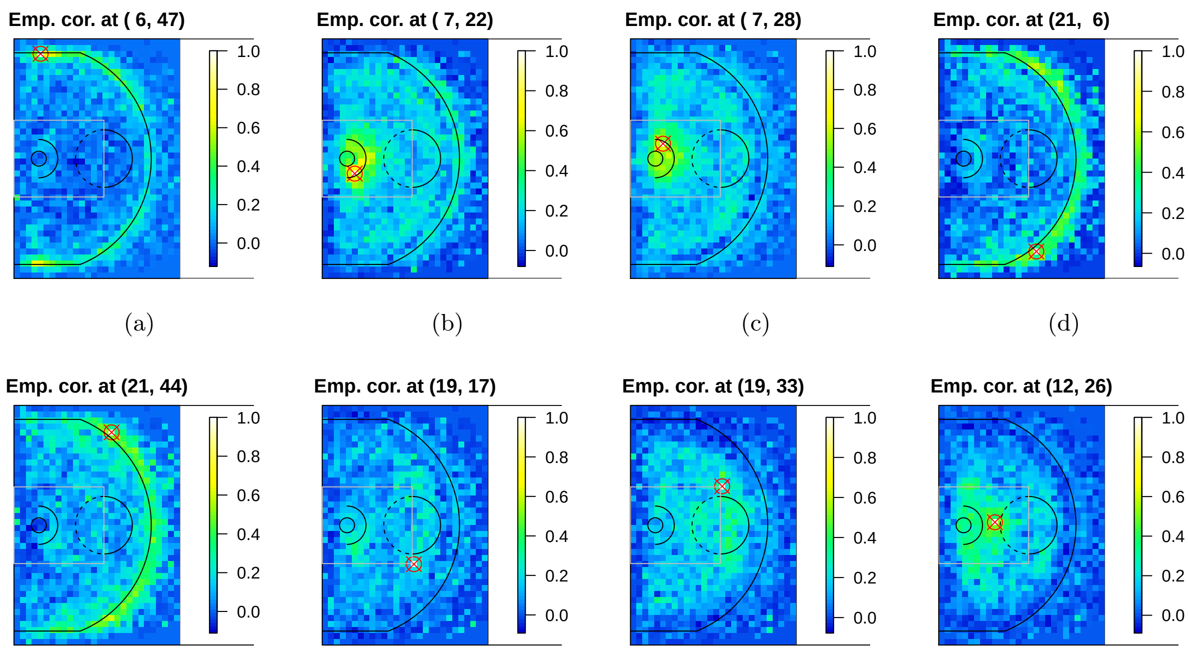


(b) mid



(c) long

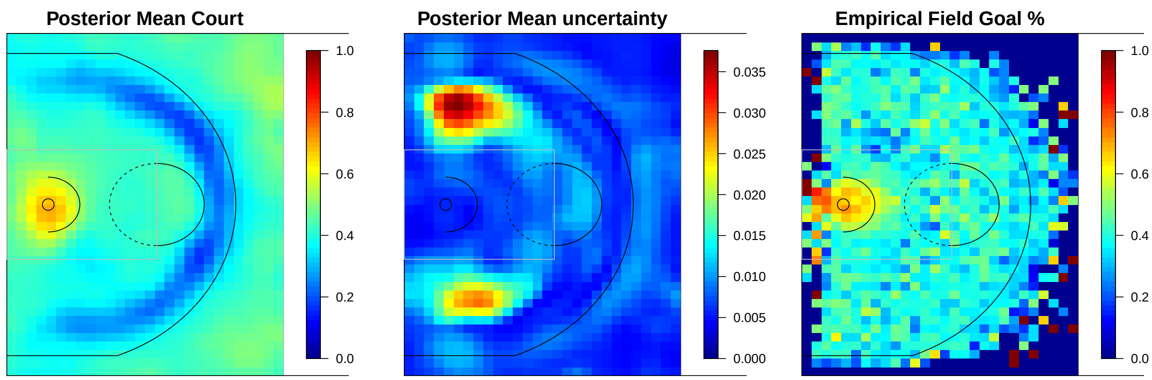
Figure 4. $K = 10$ basis vectors. LGCP-NMF has separated the layups into right and lefty (with righty taking some extra mass). Mid range shots have been expanded into four distinct shot types. Three pointers have been separated into corner, wing and top of the key.



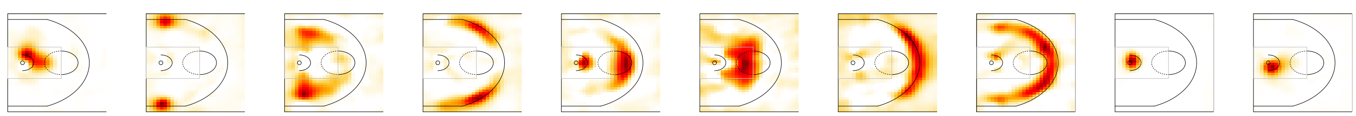
(a) (b) (c) (d)

(e) (f) (g) (h)

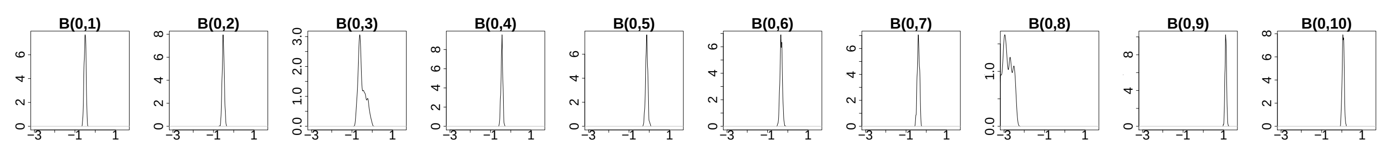
Figure 5. A broader picture of empirical correlation in this application. There clearly exist long range correlations corresponding to the three point shot. The NMF decomposition of individually fit stationarity surfaces not only reduces the dimension of our data, but also finds this global correlation structure in its basis.



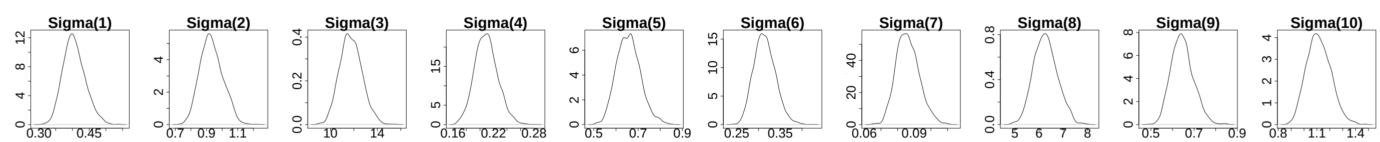
(a) posterior mean and uncertainty for $\beta_{0,:}$, and empirical efficiency surface



(b) basis

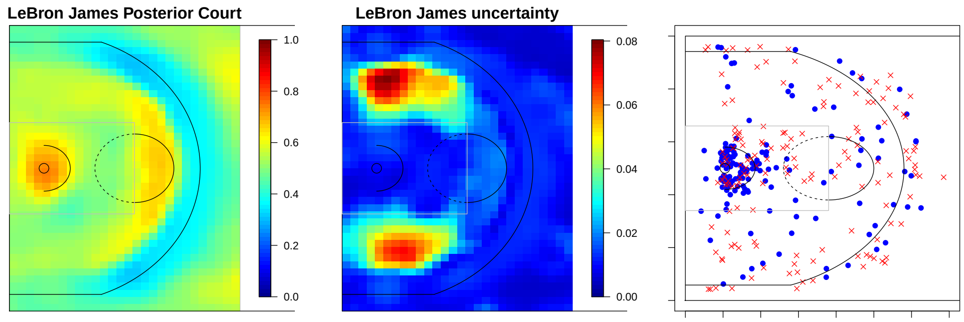


(c) $\beta_{0,:}$ posteriors

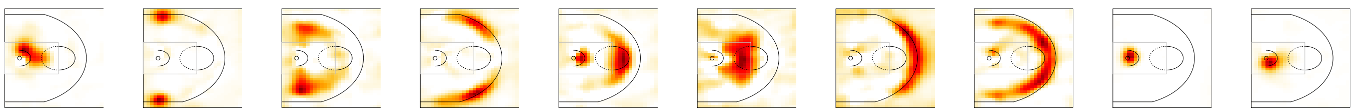


(d) σ_k^2 posteriors

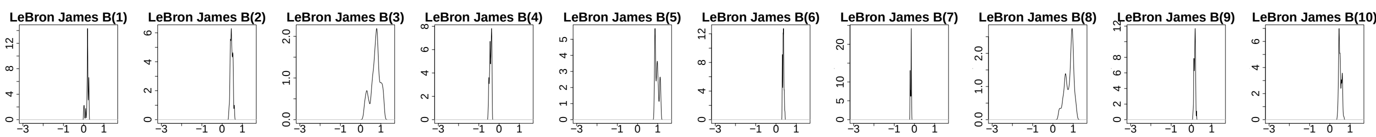
Figure 6. Field goal percentage surface global parameter fit. (a) graphically depicts the spatial surface for the mean $P(Y = 1|X)$, and its posterior uncertainty. (b) graphically depicts the shot types. (c) shows the posterior fits for $\beta_{0,:}$, which determine the average field goal percentage for each of the $K = 10$ shot types. (d) shows the fit for the σ_k^2 parameters, which determine how much the field goal percentage varies from each surface. The shots that vary the most are the longer range two point shots. We note that this large variation allows for an extremely low β value for basis 8, which appears lower than it ought to be.



(a) posterior mean and uncertainty, $\beta_{0,:}$

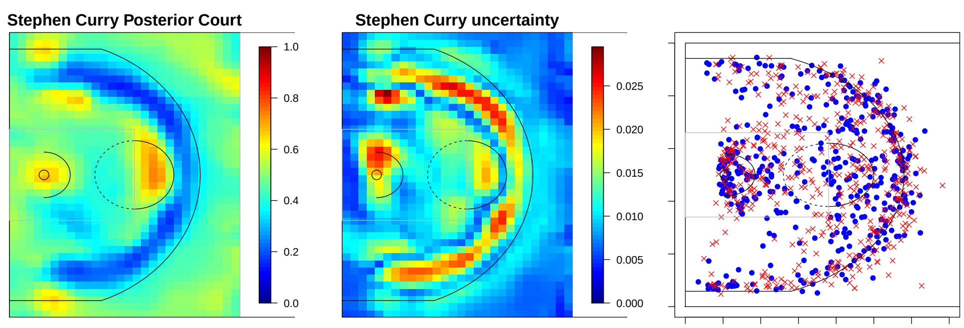


(b) basis

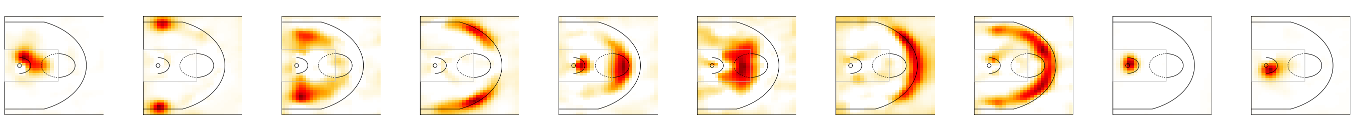


(c) LeBron's β_k posteriors

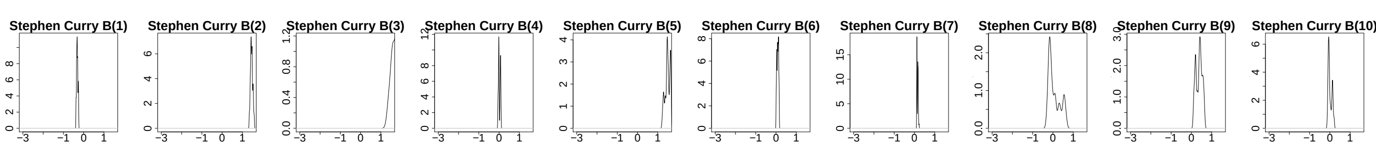
Figure 7. Spatial breakdown of LeBron James.



(a) posterior mean and uncertainty, and original shot chart

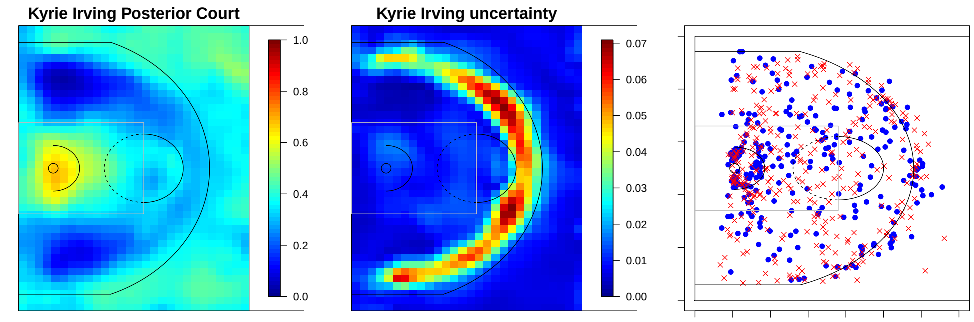


(b) basis

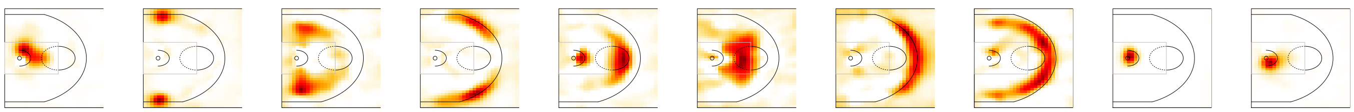


(c) Curry's β_k posteriors

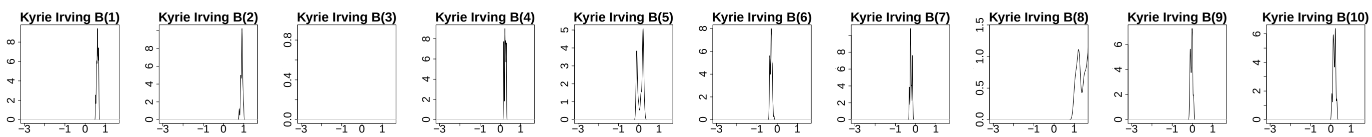
Figure 8. Spatial breakdown of Stephen Curry.



(a) posterior mean and uncertainty, $\beta_{0,:}$

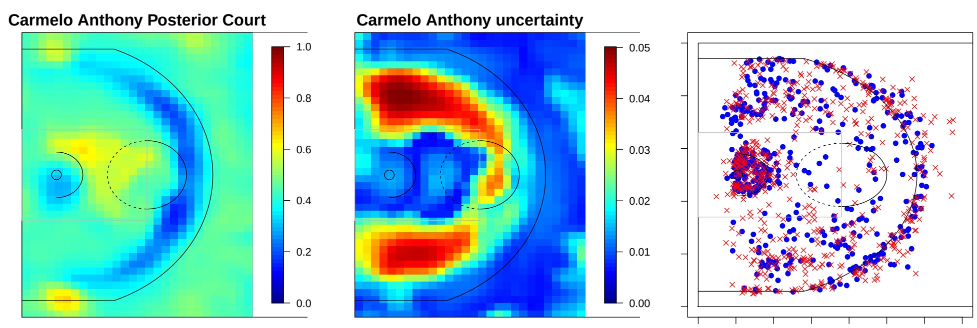


(b) basis

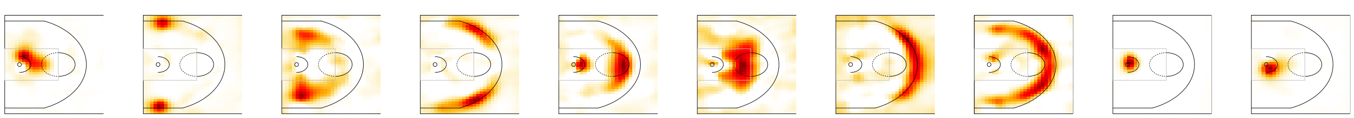


(c) Irving's β_k posteriors

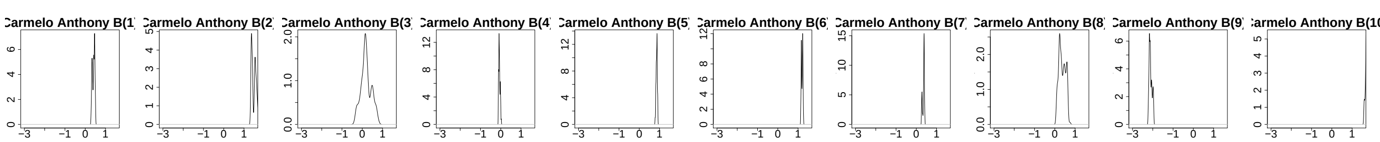
Figure 9. Spatial breakdown of Kyrie Irving.



(a) posterior mean and uncertainty, $\beta_{0,:}$



(b) basis



(c) Anthony's β_k posteriors

Figure 10. Spatial breakdown of Carmelo Anthony.