

Analysis of shot probability in NBA games using Bayesian modeling

Tuomas Haapasalo, Aatu Selkee, Timi Turpeinen

December 9, 2025

Abstract

In this project, we utilize Bayesian modeling to analyze shot efficiency in the NBA, focusing on both temporal and spatial factors. Using publicly available shot-by-shot data [1], we develop two primary models. One is designed to assess the time-based effects of shot efficiency and the second assesses the spatial effects of the shot efficiency on the court.

One of the main findings of this project is that the time plays little to no effect on the probability of making a shot. On the other hand, spatial effects influence the probability a lot. We find that there are three distinct efficient areas that maximize the expected points of a shot. These areas are the close proximity of the hoop, the corner three point zone, and the middle three point zone. The long two point shots should also be avoided, as they provide the worst expected number of points. These findings are meant to provide guidance to analysts and coaches in the NBA and other leagues on what shots are the best to take and which shots should not be taken. We also perform this analysis out of scientific curiosity to apply Bayesian methods in difficult tasks, with a lot of data.

Contents

1	Introduction	3
1.1	Motivation	3
1.2	Problem	3
2	Data and Analysis	3
2.1	Illustration of the dataset	3
2.1.1	Spatial data	4
2.1.2	Temporal data	5
2.2	Usage of data in this study	5
3	Analysis of temporal effects	6
3.1	Prior specification	6
3.2	Convergence diagnostics using Hamiltonian Monte Carlo	7
3.3	Posterior predictive checks and model comparison	8
3.4	Overtime analysis	9
3.5	Remarks	10
4	Analysis of spatial effects	10
4.1	Prior specification	11
4.2	MCMC setup and convergence	13
4.3	Prior sensitivity analysis	14
4.4	Posterior predictive checks	14
4.5	Model comparison	15
4.6	Predictive performance	16
4.7	The most optimal shot locations	18
5	Conclusion	19
5.1	Interpretation of results	19
5.2	Limitations and potential improvements	20

1 Introduction

1.1 Motivation

Basketball analytics has become an integral part of modern sports science, offering insights into player behavior, team strategy, and performance optimization. The National Basketball Association (NBA) provides an ideal environment for such analysis, as detailed player tracking data enables the study of spatial and temporal patterns underlying shot outcomes. Understanding how shot success varies across court locations and over game time is of both practical and theoretical importance. From a practical standpoint, coaches and analysts can leverage such findings to inform defensive plans and optimize offensive plays. From a scientific perspective, modeling spatial and temporal shooting tendencies provides a valuable application of Bayesian methods in the real-world.

Traditionally, shot analysis has focused on descriptive summaries such as shot charts and raw shooting percentages by zone. By employing Bayesian data analysis, we can estimate a continuous surface of shot success probabilities across all court coordinates, yielding richer inference about a player's efficiency and preferred shooting zones. Furthermore, player performance is known to fluctuate during games and across time due to factors such as fatigue, rhythm, or psychological momentum. Investigating whether players exhibit "hot-hand" or "warm-up" effects allows us to quantify temporal dependencies and understand the dynamic nature of player performance.

1.2 Problem

The goal of this study is to develop a Bayesian framework for analyzing NBA shot data. We aim to model how spatial and temporal factors jointly influence shot success probabilities.

The **research questions** of this study are:

1. What regions of the basketball court are most favorable for shot success among NBA superstars and the entire league?
2. How does the probability of making a shot evolve with respect to the game clock? Does there exist any warm-up or fatigue effects?
3. How do spatial (location-based) and temporal (time-based) components interact to influence overall shooting efficiency?

2 Data and Analysis

The NBA player shot dataset was obtained from a GitHub repository[1] containing all regular-season field-goal attempts from the 2004–2025 seasons. This dataset was originally extracted from NBA.com. For illustration, the dataset includes players with distinct shooting profiles: LeBron James, who primarily attempts shots near the rim; Stephen Curry, known for his high volume of three-point attempts; and James Harden, whose shot distribution falls between these two extremes.

Previous studies using this dataset have focused mainly on descriptive analyses and have not employed Bayesian modeling. In contrast, this study adopts a Bayesian regression framework to model shot-success probabilities while incorporating both spatial and temporal covariates.

2.1 Illustration of the dataset

To better understand how spatial and temporal effects affect shot probability prior to fitting any models, we first perform classical data-analysis, illustrated in the Figures of this section [2].

Each season's dataset includes variables such as PLAYER_ID (NBA's unique player identifier), LOC_Y (vertical court location), LOC_X (horizontal court location), QUARTER (game quarter), MINS_LEFT (minutes remaining in the quarter), SECS_LEFT (seconds remaining in the quarter), SHOT_MADE (shot outcome), and SHOT_DISTANCE (distance from the hoop). For the present analysis, only these variables are retained.

2.1.1 Spatial data

Figure 1 illustrates shooting patterns of three players, namely Lebron, Harden and Curry. We have the shot charts of made and missed attempts during the 2023 NBA season that aid in understanding how shots are distributed across the field. In the second Figure 2, we have histograms of made shots relative to the distance from the basket for the same three players.

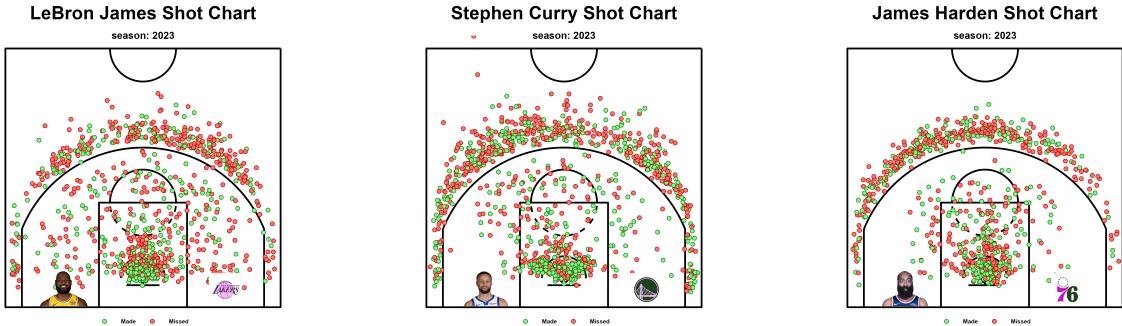


Figure 1: Shot chart of elite players - LeBron, Curry and Harden in the 2023 NBA season.

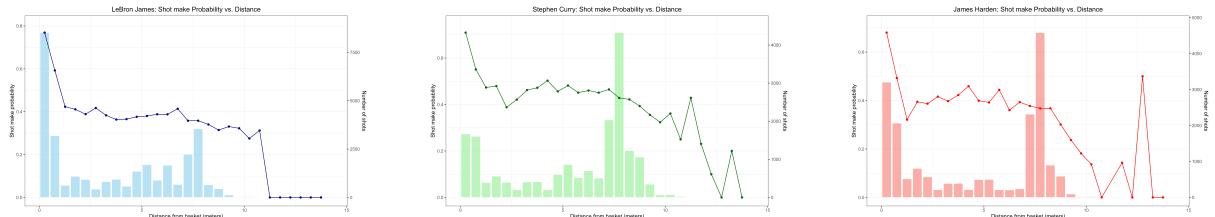


Figure 2: Distributions of made shots for LeBron, Curry, and Harden with respect to distance from the hoop in the 2004-2025 NBA season.

Figures 1 and 2 indicate a partially bimodal spatial distribution of successful shots, with high-density regions both near the basket and along the three-point line. In particular, Figure 2 shows that the success probability of made shots is an almost non-decreasing function of distance from the hoop up to the three-point line. Notably, shot success decreases sharply very close to the rim but remains relatively stable within the area inside the three-point line. This pattern suggests that a substantial portion of made shots occurs either near the basket or just inside the three-point line, consistent with typical NBA shot-selection tendencies. Beyond the three-point line, success probability declines more rapidly, as expected, since long-distance attempts are less common and more difficult.

Thus, our initial hypothesis is that the **probability of a successful shot follows a non-decreasing spatial distribution**, with areas of highest success probability occurring near the basket, while decreasing beyond the three-point line.

At the player level, Figure 2 further illustrates distinct spatial preferences: LeBron James's attempts cluster near the basket, James Harden's shots concentrate around the three-point arc, and Stephen Curry's distribution lies between these extremes. These results confirm our initial hypothesis regarding the players' differing spatial shooting profiles.

2.1.2 Temporal data

We further examine how shot success probabilities evolve throughout a game [2]. The following bar plots illustrate the probability of made shots relative to all attempted shots during the 2021-2025 NBA seasons for all players, using different temporal bin widths.

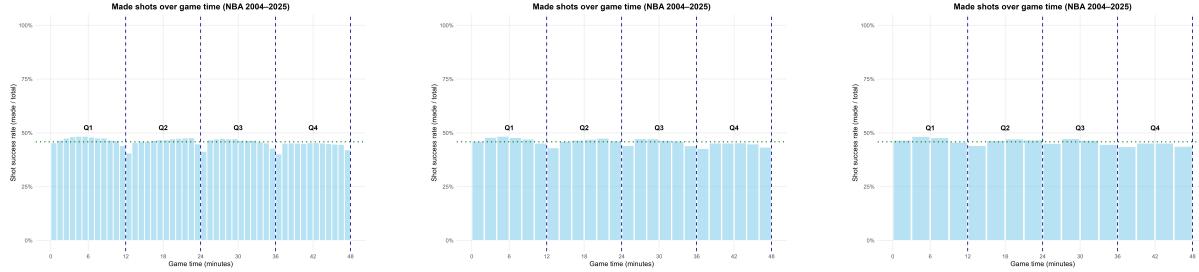


Figure 3: Probability of shot success on game time during the 2004-2025 NBA seasons with varying bin widths.

Figure 3 illustrates the shot success probability as a function of game time during the 2004-2025 NBA seasons, shown for varying temporal bin widths. Blue dashed lines mark quarter boundaries, and the green line indicates the overall mean success probability.

The bar plots shows that **shot success probability remains relatively uniform over game time**. Deviations from uniformity appear minor and can be attributed to random fluctuations (white noise) or structured interruptions such as quarter breaks. Overall, the distribution suggests that shot success is largely time-invariant at the game level.

However, figure 3 aggregates data across all players and thus does not capture potential player-specific temporal effects. To address this, figure 4 presents the same analysis separately for players LeBron, Curry, and Harden, using 2-minute time bins.

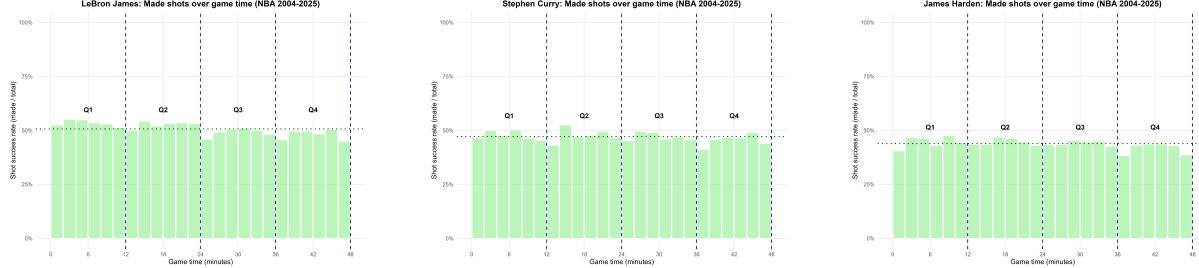


Figure 4: Player-specific shot success probability on game time during the 2004-2025 NBA seasons, using 2-minute bins.

Figure 4 indicates that the temporal patterns of **shot success are highly similar across players**. Minor differences can be explained by player rotations, rest periods, and stochastic variation. Overall, Figures 3 and 4 suggest that **shot success probability remains approximately constant over game time for both aggregated and player-level data**. In the following section, we implement a Bayesian model to formally test this hypothesis.

2.2 Usage of data in this study

Due to computational reasons, we can not use the whole dataset. The dataset contains every shot taken after 2004 in the NBA, and hence there is a lot of data in the dataset. Therefore, in each section of this study, we take a subset of the entire dataset. In section 3, we use 1000 randomly sampled shots from each season totaling to 21 000 data points [2]. On the other hand,

in section 4, we use less data as our models are more complex [2]. In total, we use 3000 data points from seasons 2021, 2022 and 2023 in the training set and 3000 additional data points from seasons 2024 and 2025 in the testing set.

3 Analysis of temporal effects

In this section, we test the hypothesis that shot success probability remains constant over game time by using 1000 datapoints from each season, both in training and testing. As established in Section 2, the temporal patterns of shot success are highly similar across players, so we do not impose any hierarchical structure on the models. To formally evaluate our hypothesis given the available observations, we employ Bayesian modeling and compare two models:

Model 1: $Y_i \sim \text{Ber}(\theta_i)$, $\text{logit}(\theta_i) = a$, and

Model 2: $Y_i \sim \text{Ber}(\theta_i)$, $\text{logit}(\theta_i) = f(t_i) = \beta_0 + \sum_{k=1}^K \beta_k B_k(t_i)$,

where β_k are coefficients of f and $B_k(t_i)$ are the basis functions of the model. In Model 1, a is a real-valued intercept, and $\theta_i \in (0, 1)$ denotes the success probability of the i th shot. The logit transformation constrains a and $f(t_i)$ to the probability scale. In model 2, the success probability θ_i is modeled coming from a deterministic function (a spline) $f : [0, \infty) \rightarrow [0, 1]$ that varies with the shot time t_i . This function serves as an unknown link function to be inferred through Bayesian analysis by trying to find proper coefficients β_k , $k = 0, 1, 2, \dots$. Accordingly, model 1 assumes a constant success probability across game time, while model 2 allows success probability to vary as a function of time.

3.1 Prior specification

We employ weakly informative priors for both models to regularize estimation without exerting excessive influence on the posterior distribution [2]. The priors are specified as follows:

- For **Model 1**: $a \sim N(0, 2^2)$, and
- for **Model 2**: $\beta_0 \sim N(0, 2^2)$, and $\beta_k \sim N(0, \tau^2)$ with $\tau \sim \text{Half-Cauchy}(0, 2)$.

In model 1, the choice $\alpha \sim N(0, 2^2)$ serves as a weakly informative prior: the mean $\mu = 0$ on the logit scale corresponds to a success probability of $\theta_i = 0.5$, while the variance $\sigma^2 = 4$ covers a broad, plausible range of success probabilities θ as $[\text{logit}^{-1}(-4), \text{logit}^{-1}(4)] \approx [0.018, 0.982]$. Similarly, model 2 includes an intercept prior $\beta_0 \sim N(0, 2^2)$. The remaining spline coefficients β_k are drawn from Gaussian distributions with variances τ^2 , where $\tau \sim \text{Half-Cauchy}(0, 2)$. This structure provides a standard weakly informative prior for scale parameters τ in Bayesian hierarchical models, which discouraging excessive non-smoothness in the spline as $\beta_k \sim N(0, \tau^2)$, and thus reducing the risk of overfitting.

Using the `brms` package in R, models 1 and 2 with the specified priors can be expressed as:

```
# Formula of model 1:
model1_formula <- bf(SHOT_MADE ~ 1, family = bernoulli(link = "logit"))

# Priors of model 1:
priors_model1 <- c(prior(normal(0, 2), class = "Intercept"))

# Formula of model 2:
model2_formula <- bf(SHOT_MADE ~ s(TIME), family = bernoulli(link = "logit"))

# Prior of model 2
```

```
prior_model2 <- c(prior(normal(0, 2), class = "Intercept"),
prior(cauchy(0, 2), class = "sds"))
```

By default, `brms` uses $K = 10$ basis functions with thin plate regression splines. For model 2, it is unnecessary to explicitly define priors for the spline coefficients, as `brms` assumes Gaussian priors by default. Only the hyperprior for τ must be specified.

3.2 Convergence diagnostics using Hamiltonian Monte Carlo

We employ Hamiltonian Monte Carlo (HMC) sampling to estimate the posterior distributions of models 1 and 2 and to assess how well they capture shot success probabilities over time.

The `brms` package provides an user-friendly interface to Stan, which implements HMC with the No-U-Turn Sampler (NUTS). NUTS is particularly well-suited for complex Bayesian models, especially hierarchical and high-dimensional ones, as it uses gradient information to efficiently explore the parameter space rather than relying on random-walk proposals. Since our model 2 is a Bernoulli likelihood with a logit link function, it is non-conjugate model, and thus lacking a closed-form posterior distribution that can be easily sampled. Hence, HMC is an appropriate and efficient choice, eliminating the need for manually derived conditional distributions as required of complex models, e.g., in Gibbs sampling. While model 1 one could easily derive manually the conditional distribution, for model 2 this stays a dream.

We fit each model using four independent chains with 2000 iterations each, including 1000 warm-up iterations. During warm-up, NUTS adapts its step size and mass matrix to efficiently explore regions of high posterior density. Running multiple chains allows a reliable assessment of convergence. For model 1, we use the default `brms` target acceptance rate of 0.80, while for model 2 we increase this to 0.90 to improve stability by forcing smaller step sizes for more accurate, albeit slower, sampling in high-density regions.

The HMC setup for both models in `brms` is as follows:

```
fit1 <- brm(
  formula = model1_formula,
  data = df_regular,
  family = bernoulli(link = "logit"),
  prior = priors_model1,
  chains = 4,           # Number of MCMC chains
  cores = 4,            # Number of CPU cores to use
  iter = 2000,          # Total iterations per chain
  warmup = 1000         # Warmup/burn-in iterations
)

model_2_fit <- brm(formula = model2_formula, data = df_regular,
family = bernoulli(link = "logit"), prior = priors_model2, chains = 4, cores = 4,
iter = 2000, warmup = 1000, control = list(adapt_delta = 0.95))
```

The table below summarizes key convergence diagnostics: the potential scale reduction factor (\hat{R}) and the Effective Sample Size (ESS) for the intercept parameter, along with the number of divergent transitions [2].

Table 1: HMC convergence diagnostics for Model 1 and Model 2

Model	Parameter	\hat{R} (Target ≤ 1.01)	Bulk ESS	Tail ESS
Model 1 <i>(Time-Invariant)</i>	Intercept	1.00	1310	1938
	Divergent Transitions	0 after warmup		
Model 2 <i>(Time-Varying Spline)</i>	Intercept	1.00	4292	2774
	sds(time)	1.00	1341	986
Divergent Transitions		27 after warmup		

From Table 1, all \hat{R} values are 1.0, indicating good chain mixing and convergence to a common posterior distribution. Model 1 exhibited no divergent transitions, suggesting stable sampling, while Model 2 produced 27 divergent transitions. Although such divergences suggest localized difficulties in exploring the posterior, their small number relative to total iterations indicates that NUTS adequately captured the main posterior structure. Thus, we conclude that we can proceed the analysis of temporal effects with these models.

3.3 Posterior predictive checks and model comparison

Because the target variable in both models follows a Bernoulli distribution, conventional posterior predictive checks provide limited insight, as they reveal little about model performance. Instead, we evaluate the models by comparing predicted shot success probabilities over time with 95 % confidence intervals for the posterior mean against observed success rates, aggregated into 2-minute bins [2]. The resulting plots are shown below.

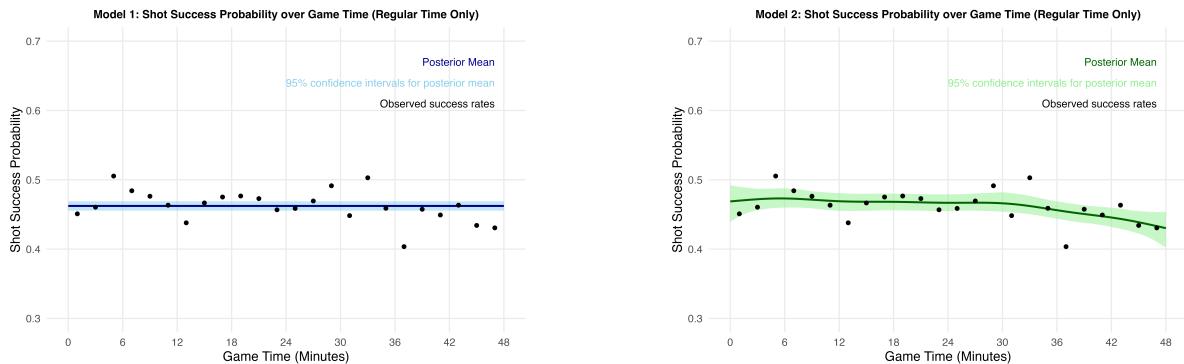


Figure 5: Predicted shot success probabilities over regular game time for Model 1 and Model 2, with 95 % credible intervals for the mean and observed success rates (2-minute bins)

Figure 5 shows that both models fit the data reasonably well, accounting for inherent noise in shot outcomes. Model 2, which incorporates a spline over time, appears to provide only a marginally better fit than Model 1. This suggests that the simpler, time-invariant Model 1 adequately captures the underlying pattern, supporting the hypothesis that shot success probability remains constant over game time.

To formally compare model performance, we apply Leave-One-Out Cross-Validation (LOO-CV) method, which estimates the expected log predictive density (elpd_{loo}) to assess out-of-sample predictive accuracy. The results are summarized below.

From Table 2, the time-varying model exhibits a slightly higher elpd_{loo} (-15088.7) than the time-invariant model (-15084.8), indicating marginally better predictive performance. The

Table 2: LOO Comparison of model 1 and model 2

Model	elpd _{loo}	seloo	elpd _{diff}	se _{diff}
Model 1 (Time-Invariant)	−15088.7	11.2	−3.92	3.31
Model 2 (Time-Varying Spline)	−15084.8	11.6	0.00	0.00

difference in expected log predictive density ($\text{elpd}_{\text{diff}} = -3.92$) which magnitude is slightly bigger than its standard error (3.31), suggesting that this difference is not statistically significant. Therefore, both models generalize equally well to unseen data.

Overall, the results from figure 5 and the LOO-CV comparison support the hypothesis that **shot success probability is generally time-invariant**. This implies that game time does not meaningfully influence spatial shot success, and that shot efficiency among NBA superstars remains stable throughout the game — providing no strong evidence of “hot-hand” or “warm-up” effects.

3.4 Overtime analysis

Previously, we compared the models only over regular game time; however, it remains unclear how they predict shot success probabilities during overtime.

NBA overtime consists of five-minute periods. If the score remains tied, additional five-minute periods are played until a winner is determined. Across the 2004–2025 seasons, the data show a maximum of four overtime periods per game.

We fit the same two models to the overtime subset of the 2004–2025 NBA seasons. Table 4 summarizes the convergence diagnostics [2]. Both models exhibit good chain mixing and convergence, with \hat{R} values near 1.00 and adequate bulk and tail ESS.

Table 3: HMC convergence diagnostics for Model 1 and Model 2 for game overtime

Model	Parameter	\hat{R} (Target ≤ 1.01)	Bulk ESS	Tail ESS
Model 1 <i>(Time-Invariant)</i>	Intercept	1.01	1335	1673
	Divergent Transitions	0 after warmup		
Model 2 <i>(Time-Varying Spline)</i>	Intercept	1.00	3308	2371
	sds(time)	1.00	1335	1544
	Divergent Transitions	1 after warmup		

Posterior predictive checks were performed as before. Figure 6 displays the predicted success probabilities over overtime with 95 % confidence intervals for the posterior mean, alongside observed success rates aggregated into 2-minute bins.

Figure 6 shows that model 1 yields a slightly lower and time-invariant prediction during overtime and fails to capture one observed spike at 65 minutes. Model 2 exhibits greater temporal variability and its 95 % interval captures more observed success rates. This indicates better alignment with the data, although the spline appears more variable than in figure 5, suggesting potential overfitting in the smaller overtime dataset.

We formally compare the models using LOO-CV method. Table 4 reports the results of the method.

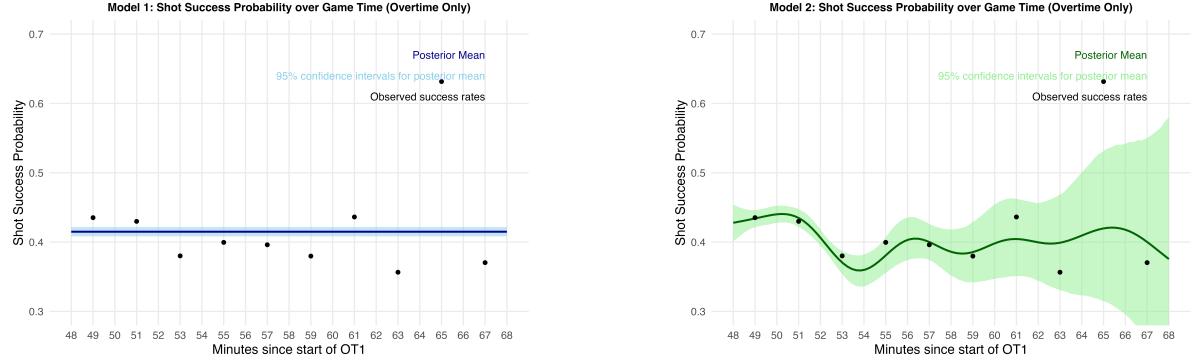


Figure 6: Predicted shot success probabilities over game overtime for Model 1 and Model 2, with 95 % confidence intervals and observed success rates (2-minute bins)

Table 4: LOO Comparison of model 1 and model 2 over game overtime

Model	elpd _{loo}	se _{loo}	elpd _{diff}	se _{diff}
Model 1 (Time-Invariant)	-14832.9	25.0	0.00	0.00
Model 2 (Time-Varying Spline)	-14805.5	26.1	-27.4	7.48

Model 2 shows better predictive performance in overtime, as the improvement in expected log predictive density ($\text{elpd}_{\text{diff}} = -27.4$) exceeds twice its standard error, indicating a meaningful difference. This may reflect the increased importance of remaining time during overtime play, which model 2 can accommodate.

However, because overtime events are relatively rare compared to regular-time play, the overall impact on season-long success probabilities is limited. Thus, although model 2 performs better within overtime, the time-invariant model 1 remains more robust model for capturing overall shot success probabilities across both regular and overtime play.

3.5 Remarks

We do not pursue temporal analysis any further as we conclude the temporal effects to be negligible. Therefore, we omit further analysis of the models such as prior sensitivity or posterior predictive checks.

4 Analysis of spatial effects

In this section, we find out, if and how, the probability of a successful throw varies in space. Specifically, our aim is to find out, which positions on the court provide the best opportunities to score given a field goal may award either two or three points. We hypothesize that the most profitable shots are the ones taken very close to the basket, as well as short three point shots in terms of expected points. We test this hypothesis with two different models:

$$\begin{aligned} \textbf{Model 1: } Y_i &\sim \text{Ber}(\theta_i), \quad \text{logit}(\theta_i) = \alpha_0 + \beta_1 \phi_1(d_i) + \beta_2 \phi_2(d_i) + \beta_3 \phi_3(d_i) + f(x_i, y_i), \text{ and} \\ \textbf{Model 2: } Y_{ij} &\sim \text{Ber}(\theta_{ij}), \quad \text{logit}(\theta_{ij}) = \alpha_0 + \beta_1 \phi_1(d_i) + \beta_2 \phi_2(d_i) + \beta_3 \phi_3(d_i) + f(x_i, y_i) \\ &\quad + \alpha_j + \beta_{1,j} \phi_1(d_i) + \beta_{2,j} \phi_2(d_i) + \beta_{3,j} \phi_3(d_i), \end{aligned}$$

where (x_i, y_i) is the location where the shot was taken from and d_i is the distance to the basket. Moreover, here $\phi_k(d_i)$ stands for the k -th orthonormal polynomial of distance d_i obtained using

Gram-Schmidt method in R, and similarly to the temporal models, the function f stands for a spline for the spatial component of our model. As the location on the court is two dimensional, we use a two dimensional spline with basis functions B_k and B_n instead. That is, our spline is

$$f(x_i, y_i) = \sum_{k=0}^{K_x} \sum_{n=0}^{K_y} \gamma_{kn} B_k(x_i) B_n(y_i),$$

where B_k, B_n are the basis functions of the spline and by default $K_x = K_y = 10$ in brms.

In the second model, we use hierarchy and assume that the distance does not effect the probability of making the shot in the same way for each player j . Hence, in addition to the overall coefficients, each player has their own coefficients for the distance features.

Justification for model selection:

The effect of distance on shot success probability maybe inferred to be non-linear as seen in Figure 7 that encapsulates the probabilities of making a shot. Moreover, we observe that the effect of distance to shot success probability varies between players but the non-linear trend is visible in every plot as can be seen from Figure 2. Thereby we use 3rd degree orthonormal polynomials in model 1 and 2.

We do not infer hierarchy for positional (x_i, y_i) terms to keep the models simple and effective with good running time efficiency when fitting the models to our data [1].

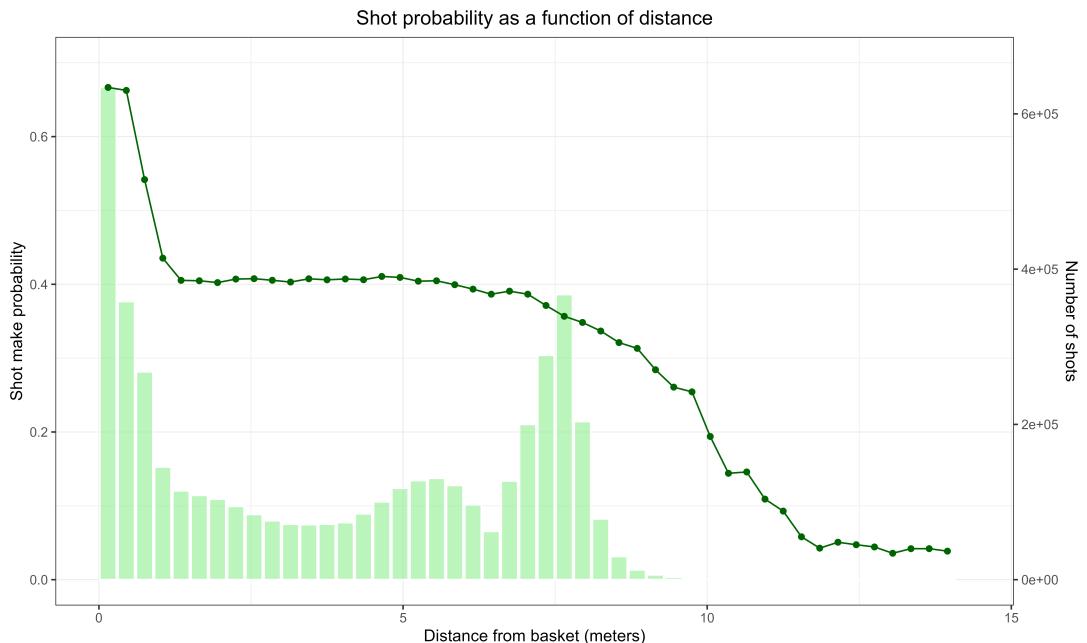


Figure 7: Shot probability as a function of distance from the hoop and the frequency of shots. NBA seasons 2004-2025.

4.1 Prior specification

Once again, we use weakly informative priors for both models. The priors are as follows:

Model 1: $\alpha_0 \sim N(0, 2^2)$, $\beta_l \sim N(0, 100^2)$, $l = 1, 2, 3$ and

$$\gamma_{kn} \sim N(0, \sigma_s^2), \sigma_s \sim \text{Student-}t(3, 0, 2.5)$$

Model 2: $\alpha_0 \sim N(0, 2^2)$, $\beta_l \sim N(0, 100^2)$, $l = 1, 2, 3$ and

$$u_j \sim N(0, C), C = \text{Diag}(\sigma) \cdot \Sigma \cdot \text{Diag}(\sigma), \Sigma \sim \text{LKJ}(2), \sigma_m \sim \text{Student-}t(3, 0, 2.5), m = 1, 2, 3, 4$$

$$\text{and } \gamma_{kn} \sim N(0, \sigma_s^2), \sigma_s \sim \text{Student-}t(3, 0, 2.5),$$

where $u_j = (\alpha_j, \beta_{1,j}, \beta_{2,j}, \beta_{3,j})^\top$ is the player-specific random-effects vector of j th player, $\sigma = (\sigma_\alpha, \sigma_{\beta_1}, \sigma_{\beta_2}, \sigma_{\beta_3})^\top$ is the group-level standard deviations, and Σ the group-level correlation matrix.

We use these priors, as they are good standard choices that do not yield too much information to the model, while still regularizing the model to not overfit the data. The likely values for the intercept α_0 in both models cover the believable range of success probabilities, and none of the other parameters dominate the others as we believe all effects are important in the final model. In addition, by default brms draws the spline coefficients from Gaussian distribution with variances σ_s^2 , where the variance is drawn from student-t distribution with 3 degrees of freedom, and mean 0 and standard-deviation 2.5 [3]. This structure also provides a standard weakly informative prior for scale parameters σ_s in Bayesian models, discouraging excessive non-smoothness in the spline.

Lastly, in the hierarchical model (model 2), each player has a vector of four random coefficients collected in u_j . The elements of the group-level standard deviation vector σ are given the same Student- $t(3, 0, 2.5)$ prior as in the spline component, providing a weakly informative prior on the scale of between-player variation. The correlation matrix Σ governing the dependence among the random effects is assigned an LKJ(2) prior, i.e., $p(\Sigma | 2) \propto \det(\Sigma)^{2-1}$, which favors modest correlations while ensuring that Σ remains a valid correlation matrix. Together with the group-level standard deviations, this yields the covariance matrix $C = \text{Diag}(\sigma)\Sigma\text{Diag}(\sigma)$, which describes how the player-specific coefficients are allowed to vary across players.

The following code captures these models with given priors using brms in R:

```
model_1_formula <- bf(
  SHOT_MADE ~ 1 + poly(SHOT_DISTANCE_FT, degree = 3) + s(LOC_X, LOC_Y))

# priors
prior_model_1 <- c(
  prior(normal(0, 2), class = "Intercept"),
  prior(normal(0, 100), class = "b"))

# model 2 formula
model_2_formula <- bf(
  SHOT_MADE ~ 1 + poly(SHOT_DISTANCE_FT, degree = 3)
  + (1 + poly(SHOT_DISTANCE_FT, degree = 3) | PLAYER_ID) + s(LOC_X, LOC_Y))

# priors
prior_model_2 <- c(
  # Fixed effects
  prior(normal(0, 2), class = "Intercept"),
  prior(normal(0, 100), class = "b"),

  # Hierarchical effects
  prior(student_t(3, 0, 2.5), class = "sd"),
  prior(lkj_corr(2), class = "cor"))
```

4.2 MCMC setup and convergence

Similarly to chapter 3, we use HMC to sample the posterior distribution. In particular, we use NUTS as in chapter 3. For both models we use a similar sampling parameters. We set the amount of iterations to 2000, the first 1000 of which are warm up iterations. We use four chains as we want to make sure the \hat{R} value is accurately computed and the chains have converged. Lastly, we use a slightly higher target acceptance rate of 0.95 to stabilize training and make sure the amount of divergences stays low.

```
model_fit <- brm(
  formula = model_formula,
  data = df_train,
  family = bernoulli(link = "logit"),
  prior = priors_model_1,
  chains = 4, cores = 4, iter = 2000,
  warmup = 1000, control = (list(adapt_delta = 0.95))
)
```

The most important convergence metrics are summarized in the table below as in the chapter 3 [2].

Table 5: HMC convergence diagnostics for spatial models. Only some parameters are presented.

Model	Parameter	\hat{R} (Target ≤ 1.01)	Bulk ESS	Tail ESS
Model 2 <i>(Pooled)</i>	α_0	1.00	5155	3483
	σ_s	1.01	439	983
Divergent Transitions		27 after warmup		
Model 2 <i>(Hierarchical)</i>	α_0	1.00	5155	3483
	σ_s	1.00	395	978
	σ_α	1.00	704	956
Divergent Transitions		18 after warmup		

Model 1 (Pooled): Model 1 exhibited excellent mixing and high sampling efficiency across the majority of its parameters. All regression coefficients, including the intercept and the orthonormal polynomial terms, achieved an \hat{R} value less than 1.01, indicating successful convergence across the MCMC chains. The Bulk and Tail Effective Sample Sizes (ESS) were sufficiently large for all parameters, confirming reliable estimation of posterior summaries. However, the model produced 27 divergent transitions after warmup by table 5. While this suggests localized difficulties in exploring the posterior, the low number relative to the total iterations (4000) and the stability of the \hat{R} values for all parameters indicate that the Hierarchical Monte Carlo (HMC) sampler successfully captured the main structure of the posterior distribution.

Model 2 (Hierarchical): Model 2 also demonstrated strong evidence of convergence based on the \hat{R} metric; all parameters, including the fixed effects, the smoothing standard deviation, and all multilevel standard deviations and correlations, achieved \hat{R} values of 1.01 or less. The ESS metrics were generally sufficient, although the lowest Bulk ESS values (as low as 395 for the smoothing term) suggest these specific hyperparameters were the most challenging to sample. Similar to model 1, 18 divergent transitions were reported as shown in table 5.

Given the overall stability of the primary convergence metrics and the complexity inherent in hierarchical models, we conclude that both models, fitted with the specified priors, have converged sufficiently well to proceed with the primary analysis.

4.3 Prior sensitivity analysis

As the choice of priors affects the convergence of our models, we performed prior sensitivity analysis on the models [2]. We chose weakly-informative priors for both the pooled (model 1) and hierarchical (model 2) structures to ensure the posterior inferences were primarily data-driven and robust.

Table 6: Powerscale sensitivity summary for spatial Models. Only some key parameters are presented, as there are way too many parameters.

Model	Parameter	Prior	Likelihood	Diagnosis ($\frac{\text{Prior}}{\text{Likelihood}}$)
Model 1 <i>(Pooled)</i>	α_0	0.0049	0.1065	0.046
	β_1	0.0019	0.3302	0.006
	σ_s	0.0031	1.3653	0.002
Model 2 <i>(Hierarchical)</i>	α_0	0.0169	0.1033	0.164
	β_1	0.0242	0.2690	0.090
	σ_s	0.0195	1.7656	0.011
	Σ	0.3708	0.5062	0.732

The powerscale sensitivity diagnostic from summary 6 indicated that for the main fixed effects and the smoothing scale, the likelihood dominated the prior, suggesting the posterior estimates were primarily data-driven (e.g., the likelihood contribution for the intercept was ≈ 0.10 while the prior was ≈ 0.005 for model 1). After refitting with alternative priors, the posterior means and credible intervals for the primary fixed effects and the global smoothing parameters varied by less than 5 %. Most importantly, the resulting predicted shot success surfaces in [2] showed no significant difference, leading us to conclude that our primary findings regarding spatial shooting efficiency are robust to our choice of priors.

However, we noticed that the correlation matrix (Σ), exhibited a much higher prior influence (ratio ≈ 0.732). This is expected as estimating a large number of correlation parameters in a high-dimensional hierarchy (we have 620 players in the data) is challenging, and the *LKJ_Corr(2)* prior is actively regularizing these off-diagonal terms to prevent sampling instability and overfitting. The prior is essential here for identifying the hierarchical model.

We tested more informative priors for the most influential fixed effects, $\beta_l \sim N(0, 10^2)$, to assess the limits of sensitivity. These tighter priors were found to be inferior to the original choice. The $\sim N(0, 10^2)$ specification resulted in multiple potential prior-data conflict's for the distance polynomial terms. Crucially, the prior's contribution for several key parameters, such as model 1's first distance polynomial term, dominated the likelihood (Prior 0.2832 versus Likelihood 0.1092). This outcome confirmed that the original, flatter priors were superior in yielding data-driven and robust model estimates.

Overall, we may conclude that our prior choices were reasonable and resulted in more robust and data-driven models.

4.4 Posterior predictive checks

For models with a binary target variable (shot outcome), conventional continuous posterior predictive checks are not informative. Consequently, we employ reliability diagrams (Figure 8) to rigorously evaluate the calibration of our models [2].

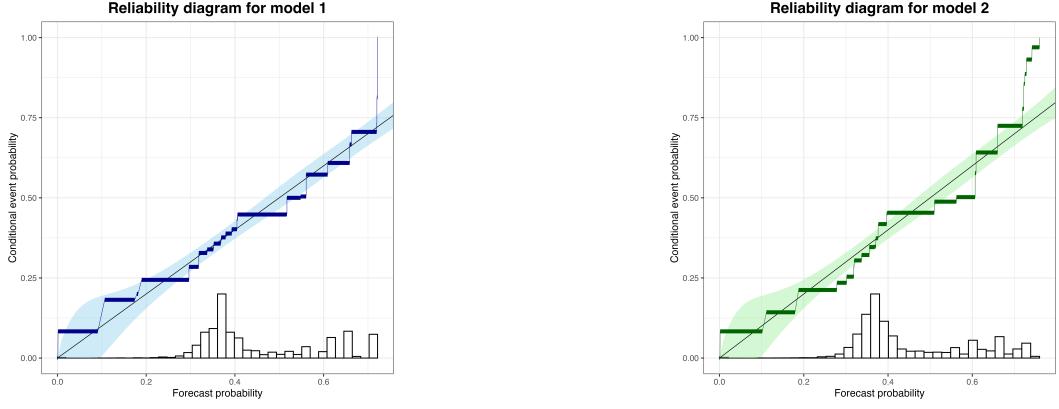


Figure 8: The reliability diagrams of pooled and hierarchical models

A reliability diagram plots the predicted probability of the event (forecast probability) on x -axis against the observed frequency of the event at that prediction level (conditional event probability) on y -axis. Perfect calibration is indicated by the black diagonal line, where the model's predicted probability perfectly matches the observed rate. The shaded band represents the uncertainty (typically the 95% credible interval) around this ideal line, while the step function illustrates the binned empirical observations.

The reliability diagrams confirm that both the pooled (model 1) and hierarchical (model 2) spatial models are well calibrated. The observed frequency (step function) for both models lies primarily within the 95% credible interval throughout the majority of the probability range (0.20 to 0.60), indicating that there is no evidence of fundamental model misspecification.. However, a minor weakness is observed at the extreme ends of the distribution: for high predicted probabilities (above ≈ 0.65), both models show a slight miscalibration where the observed success rate is actually higher than the predicted probability. This suggests the models are slightly under-predicting the probability of high-success shots. Given that the histogram at the bottom of the plots shows a low frequency for predictions in these high-probability bins, this slight deviation is most likely an artifact of limited sample size in the far right bins rather than a systematic flaw in the model structure. We conclude that both models are well-calibrated and suitable for subsequent analysis.

4.5 Model comparison

Next, we will compare our two spatial models - the hierarchical model and pooled model - by performing LOO-CV as in chapter 3. We estimate the expected log predictive density (elpd_{loo}) to assess out-of-sample predictive accuracy [2]. Table 7 summarizes the results of the LOO comparison.

Table 7: LOO Comparison of pooled and hierarchical model

Model	elpd_{loo}	se_{loo}	$\text{elpd}_{\text{diff}}$	se_{diff}
Model 1 (Pooled)	-5870.47	27.6	-1.435546	1.954007
Model 2 (Hierarchical)	-5869.03	27.7	0.00	0.00

From Table 7, the hierarchical model (model 2) exhibits a slightly higher elpd_{loo} value, implying marginally better predictive performance. However, the magnitude of the expected log predictive density difference ($\text{elpd}_{\text{diff}} \approx 1.44$) is smaller than its standard error ($\text{se}_{\text{diff}} \approx 1.95$). This finding suggests that the difference in out-of-sample predictive accuracy between the two models is

not statistically significant. Consequently, based solely on the LOO-CV metric, both models generalize equally well to unseen data.

We also visually compare the predicted shot make probability surfaces (see figure 9) [2]. Both models successfully predict the established NBA shooting patterns: the probability is highest near the hoop and decreases steeply outside the three-point line.

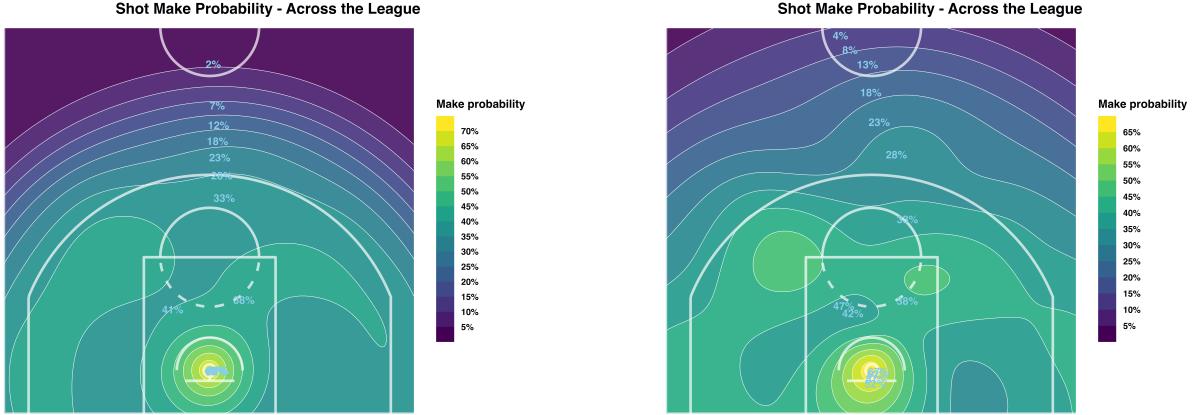


Figure 9: Contour plot of spatial shot make probabilities for pooled (left) and hierarchical models (right).

The primary visual difference occurs in the mid-range region between the hoop and the three-point line. The hierarchical model 2 appears to capture slightly greater localized variation in the probability contours, likely due to its ability to model player-specific non-linear effects of distance.

The non-significant gain in predictive performance from model 2 comes at the cost of increased complexity (many more parameters) and reduced computational stability (higher divergence rate than model 1). Given that both models predict the data well and yield similar elpd_{loo} values, the simpler model 1 presents a more robust choice unless the interpretive value of the hierarchical parameters in model 2 is deemed essential.

To fully resolve which spatial model is preferable, the following chapter will evaluate the predictive capabilities of the two models using classification accuracy.

4.6 Predictive performance

To evaluate the predictive capabilities of the two models, we compute out-of-sample and in-sample classification accuracies [2]. For each observation in the training and test sets, we obtain posterior expected predictions using `posterior_epred`, which returns the posterior mean of the Bernoulli success probability. This quantity serves as the model-based estimate of the shot-make probability.

For both models, a predicted class is obtained by thresholding the posterior mean probability at 0.5, that is,

$$\hat{y} = \begin{cases} 1, & \text{if } \mathbb{E}[y | x] > 0.5, \\ 0, & \text{otherwise.} \end{cases}$$

We then compare these classifications against the observed outcomes to compute accuracy. Figure 10 displays the distribution of correct and incorrect classifications for model 1 and model 2.

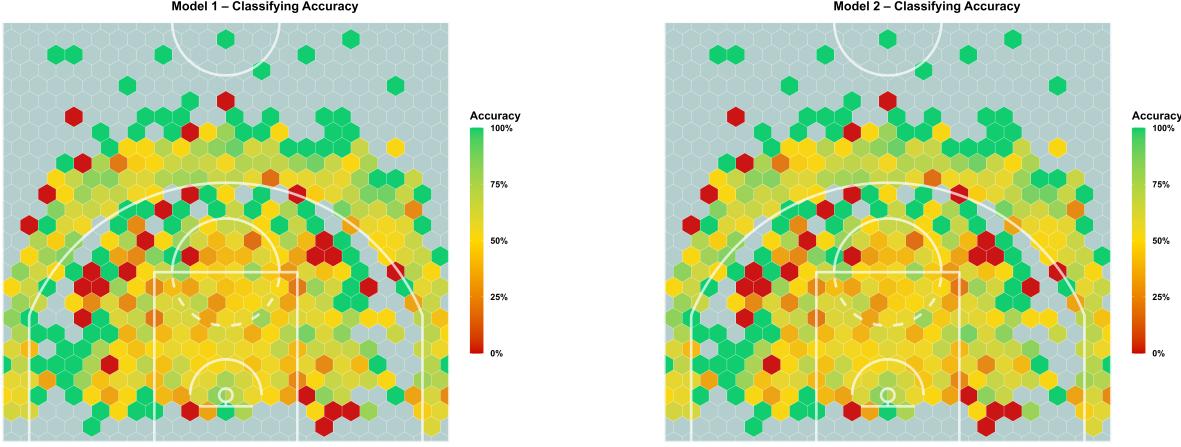


Figure 10: Classification accuracy for the two models.

Overall, both models produce nearly identical classification behavior. This similarity persists when accuracy is examined as a function of shot distance, as shown in Figure 11. No systematic divergence is observed, and in all distance bins the performance of the two models is essentially indistinguishable.

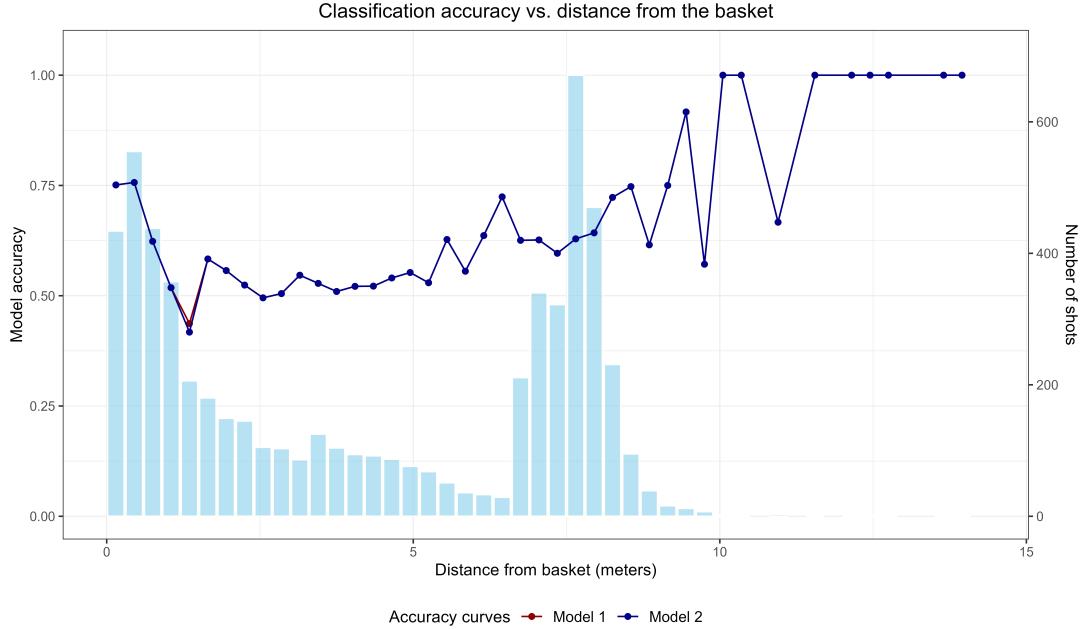


Figure 11: Classification accuracy across shot-distance bins for both models.

Table 8 reports the overall in-sample and out-of-sample accuracy. Model 1 and model 2 achieve virtually identical performance in both settings.

Table 8: Overall classification accuracy for both models.

	Model 1	Model 2
Training accuracy	0.6283	0.6283
Test accuracy	0.6208	0.6202

To assess whether the models differ at specific shot distances, we compute accuracy within 1-meter distance bins [2]. Table 9 summarizes these results. The column **Difference** reports the

absolute difference in accuracy between model 1 and model 2 for each bin. Across all distances, the difference is exactly zero or extremely small, indicating no meaningful improvement provided by the additional hierarchical structure of Model 2 for binary accuracy prediction.

Distance bin	n	Acc. M1	Acc. M2	Difference
0	1784	0.675	0.675	0.000
1	535	0.520	0.512	0.007
2	353	0.510	0.510	0.000
3	409	0.526	0.526	0.000
4	255	0.537	0.537	0.000
5	155	0.568	0.568	0.000
6	273	0.637	0.637	0.000
7	1803	0.626	0.626	0.000
8	365	0.718	0.718	0.000
9	35	0.771	0.771	0.000
10	6	0.833	0.833	0.000
11	2	1.000	1.000	0.000
12	5	1.000	1.000	0.000
13	1	1.000	1.000	0.000
14	6	0.833	0.833	0.000
15	5	1.000	1.000	0.000
16	2	1.000	1.000	0.000
18	2	1.000	1.000	0.000
20	1	1.000	1.000	0.000
21	2	1.000	1.000	0.000
23	1	1.000	1.000	0.000

Table 9: Classification accuracy as a function of shot distance for both models.

In summary, model 2’s performance on binary classification of made vs. missed shots is statistically indistinguishable from that of model 1. Overall, we see that the problem is very difficult and more information is needed to make better predictions than either of our models.

4.7 The most optimal shot locations

To address the first research question posed in Section 1.2, we construct a hexagonal grid over the court and compute the expected points associated with each location [2]. After fitting the models, we use posterior draws of the expected value of the posterior predictive distribution (`posterior_epred`) to estimate the probability of a made shot within each hex tile. For every tile, we then determine whether the corresponding shot would be worth two or three points, and compute expected points as the product of the make probability and the shot value.

Since model 2 demonstrated a slightly better predictive performance (based on LOO comparison), we base our analysis on its expected-points surface. The resulting map is shown in Figure 12.

Our initial hypothesis was that the most profitable locations would be either in very close proximity to the basket or efficient three-point locations. The figure supports this expectation. The highest expected values appear directly around the hoop, where both shot difficulty and distance are minimized. The next most profitable regions are the deep-corner three-pointers near the baseline, which combine a shorter three-point distance with the additional point value. Beyond these, the general arc three-pointers form the third-most efficient region.

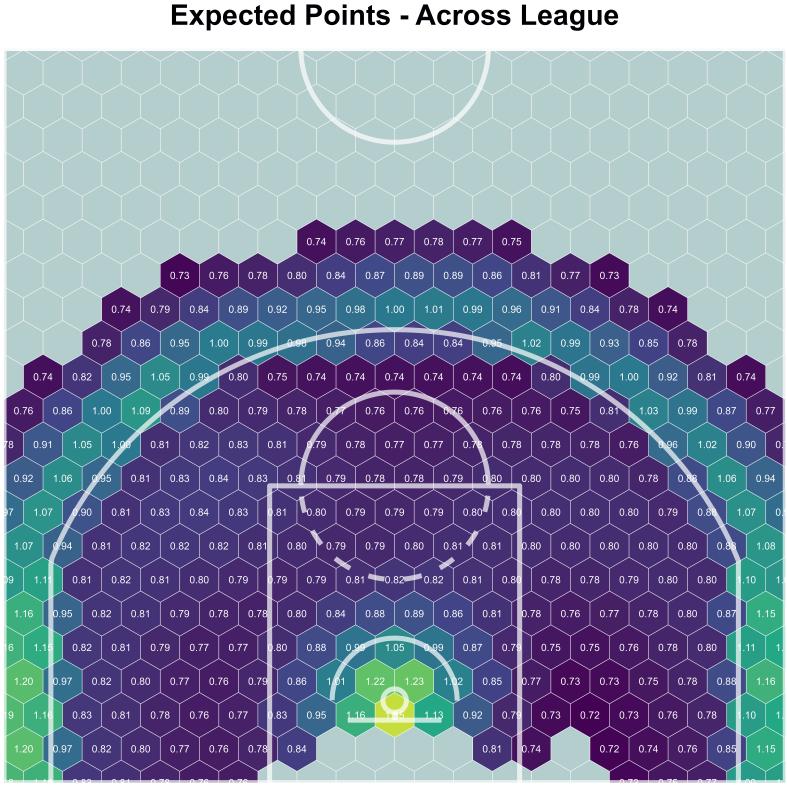


Figure 12: Expected number of points from each location.

Short mid-range attempts along the baseline and non-corner long two point shots offer substantially lower expected value. The least profitable locations are very deep three-pointers or shots from the mid-range area between the free-throw box and the arc, which produce both low make probability and limited point reward.

Overall, the expected points surface suggests to prioritize shots at the rim and efficient corner threes, while avoiding long twos and deep threes as expected in NBA games.

5 Conclusion

5.1 Interpretation of results

The primary achievement of this project is the construction of the expected points surface, which provides a definitive map of offensive efficiency. The expected-points surface as shown in figure 12 clearly delineates three highly profitable regions — shots near the hoop, deep corner three-pointers, and the general arc three-pointers — while confirming the inefficiency of mid-range and non-corner long two-point shots. This result strongly support the prevailing strategic wisdom in modern basketball: the highest-value shots are those that maximize proximity to the basket and those that leverage the three-point line, particularly the short-distance corner three.

Furthermore, the temporal analysis provided insight into how shot success probability may shift under specific game conditions, such as the increased pressure or fatigue during overtime, which can be critical for late-game decision-making.

5.2 Limitations and potential improvements

Despite the model’s robustness, several limitations present opportunities for future improvement. Firstly, the spatial model assumes a degree of exchangeability across all shots taken from the same location, ignoring the influence of defensive pressure. Secondly, the model does not currently account for contextual factors such as fast-break versus half-court plays, which dramatically affect shot difficulty. Incorporating metrics like the distance to the nearest defender, or time remaining on the shot clock could further refine the model’s predictive accuracy even though in analysis of temporal effects we concluded that time is not statistically significant factor in game play.

Thirdly, while we compared different spatial basis functions, exploring more complex smoothing techniques, such as Gaussian processes, might offer a better representation of the true underlying shot probability surface. In particular, one could try to experiment with different kernel choices in Gaussian processes, which encode different assumptions about the shot probability on the court. In addition, variational inference could provide additional efficiency benefits compared to MCMC, as we found the sampling time to be the most important factor in why we could not increase the amount of data points in the training set.

Several factors were identified as potential threats to the validity of our findings. The assumption of independence between successive shots is a theoretical threat, as the outcome of one possession often influences the subsequent play. We attempted to mitigate this by focusing on modeling location/time effects rather than sequential dependencies. Another potential threat is the inherent structure of the recorded data, which may suffer from selection bias if players only attempt certain types of low-probability shots from necessity. To address these issues, we implemented several robustness checks. The prior sensitivity analysis (section 4.3) confirmed that the posteriors were not overly dependent on the chosen prior distributions, increasing confidence in the model’s stability. Model comparison using Leave-One-Out cross-validation (LOO-CV) confirmed that the pooled spatial model is preferable, as it delivered nearly the same out-of-sample predictive performance as the more complex hierarchical alternatives. Nevertheless, we proceeded with the hierarchical model for the subsequent analysis, even though the statistical evidence suggests that the simpler pooled model would have been sufficient.

References

- [1] Dominic Samangy, 2025, [dataset](#).
- [2] Aatu Selkee, Timi Turpeinen, Tuomas Haapasalo, 2025, [code](#).
- [3] Paul-Christian Bürkner, [Prior Definitions for brms Models](#).