

---

# Análise do Estilo de Jogo do Time da NBA *Houston Rockets* na Temporada 2019/20

---

Daniel de Amaral da Silva<sup>1</sup> Marcos Aurélio Secundes Freire<sup>2</sup>

## Resumo

A modelagem estatística é uma ferramenta imprescindível à gerência de times, principalmente em franquias que estão em processo de reestruturação de elenco. Dos times que sofreram uma mudança drástica na última temporada 2019/20 - 2020/21, está o time texano do *Houston Rockets*, que passou de um time que brigou pelo título para um dos piores times na temporada 2020/21. O Presente estudo, buscou modelar, tendo em mãos dados extraídos do portal da NBA, o comportamento minuto a minuto da pontuação do time texano, através de um modelo de processos gaussianos, assim como a distribuição de localização dos acertos de quadra, pela aplicação de um modelo de misturas gaussianas. Ao final, obtendo resultados satisfatórios com  $R^2$  médio de 90% para o modelo de processos gaussianos.

## 1. Introdução

A análise quantitativa do desempenho no basquete, especialmente por meio de estatísticas de jogo, é amplamente utilizada entre treinadores, a fim de analisar decisões de jogo e otimizar táticas e escalações. Recentemente, no ano de 2019, o time da conferência leste *Orlando Magic* se tornou o primeiro time a utilizar modelos estatísticos para melhorar o processo de tomada de decisões, focando inicialmente na escolha de jogadores de base<sup>1</sup>.

Devido à enorme massa de dados disponível, e atualizada diariamente, a modelagem estatística se torna uma ferramenta imprescindível à gerência de times, principalmente em franquias que estão em processo de reestruturação de elenco. Dos times que sofreram uma mudança drástica na última

temporada 2019/20 - 2020/21, está o time texano do *Houston Rockets*, que bateu recordes de diferentes combinações de *lineups* (escalações) na temporada 2020/21, e passou de um time que brigou pelo título na temporada 2019/20 para um dos piores times na temporada 2020/21.

Criado em 1967, o *Houston Rockets* na verdade não nasceu na cidade de *Houston*, mas em *San Diego*. Contudo, em 1971 o time mudou sua sede para *Houston* que coincidentemente abriga o Centro Espacial Lyndon B. Johnson<sup>2</sup> da NASA. O time que ficou conhecido por grandes estrelas (Chris Paul, James Harden, Clint Capela, Russel Westbrook, etc) nos últimos anos, na temporada 2020/21 alterou boa parte de seu elenco, devido a trocas e vendas de jogadores. Essa alteração na escalação da equipe colocou o time no top 10 das franquias com maiores derrotas seguidas, com exatas 20 derrotas<sup>3</sup> na temporada 2020/21. Dessa forma, se faz necessário entender o comportamento do antigo *Houston Rockets* e seu estilo de jogo, para compreender como adaptar e reestabelecer a qualidade, e competitividade do time atual.

O objetivo do presente estudo é, portanto, modelar o comportamento minuto a minuto da pontuação do time texano, assim como a distribuição de localização dos acertos de quadra. Com uma visão geral desses resultados, incluindo informações relevantes como pontuação do time rival e o próprio local do jogo, poderemos fornecer ao treinador/pesquisador, informações sobre o desempenho da equipe na temporada, e servir de *baseline* para futuras alterações na equipe.

## 2. Fundamentação Teórica

Este capítulo expõe o conhecimento teórico necessário para compreensão deste trabalho. Inicialmente, é apresentado o conceito de Modelos de Misturas Gaussianas (*GMM*), logo em seguida o conceito de Processos Gaussianos (*GP*).

---

<sup>1</sup>Programa de Pós-Graduação em Engenharia de Teleinformática, Centro de Tecnologia, Universidade Federal do Ceará

<sup>2</sup>Programa de Pós-Graduação em Métodos Quantitativos, Centro de Ciências, Universidade Federal do Ceará. Contato: Daniel <danielamaral@alu.ufc.br>, Marcos <secundesfreire@bol.com.br>.

<sup>2</sup><https://spacecenter.org>

<sup>3</sup><https://www.dunkest.com/en/nba/news/31568/nba-longest-losing-streaks>

<sup>1</sup><https://basketball.realgm.com/wiretap/252765/Magic-Become-First-NBA-Team-To-Use-AI-Player-Tracking-Data>

## 2.1. Misturas Gaussianas

O Modelo de Mistura Gaussiana é uma ferramenta de modelagem estatística associada à reconhecimento de padrões na presença de variáveis latentes. Dado que a função de verossimilhança de um modelo *GMM* é de difícil otimização, então para este tipo modelo se usa o algoritmo *Expectation Maximization (EM)*.

Este algoritmo consta dos passos *E* e passo *M*. No passo *E* a estimação dos parâmetros, é baseada no cálculo do valor esperado condicional da função de verossimilhança com respeito à distribuição das variáveis latentes (variáveis não observáveis), no passo *M* tomando como dado o valor esperado das variáveis latentes, se calcula os parâmetros de interesse (Bishop, 2006).

O modelo de misturas gaussianas (*GMM*, do inglês, *Gaussian Mixture Model*) é a soma de funções gaussianas, cada uma dessas gaussianas é parametrizada por  $\theta_i$ , que é composto por um vetor de média  $\mu_i$ , uma matriz de covariância  $\Sigma_i$  e os pesos, onde  $i = 1, 2, \dots, k$ .

$$\theta = \{\alpha_1, \mu_1, \Sigma_1, \dots, \alpha_k, \mu_k, \Sigma_k\}$$

Cada componente de densidade gaussiana possui um peso, resultando numa soma ponderada (Zhang et al., 2015; Silva, 2014). A Equação abaixo apresenta a função ponderada das *M* componentes:

$$g(x|\theta) = \sum_{i=1}^k \alpha_i P(\mathbf{x}|\mu_i, \Sigma_i)$$

$\mathbf{x}$  é um vetor de características de dimensão  $d$ , os pesos das misturas são representados por  $\alpha_i$ , para  $i = 1, 2, \dots, k$ , e  $P(\mathbf{x}|\mu_i, \Sigma_i)$ ,  $i = 1, 2, \dots, k$ , representa as densidades das componentes gaussianas. Cada uma dessas componentes é uma função gaussiana  $d$ -variada. Os pesos das misturas devem respeitar a restrição  $\sum_{i=1}^k \alpha_i = 1$  (Silva, 2014).

O ajuste do modelo é realizado maximizando a verossimilhança dos dados de treinamento. Por exemplo, para  $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , tem-se que:

$$\theta^* = \arg \max_{\theta} g(\mathbf{X}|\theta)$$

Com

$$g(\mathbf{X}|\theta) = \prod_{i=1}^n g(x_i|\theta)$$

O treinamento pode ser realizado utilizando, por exemplo, o algoritmo iterativo *Expectation-Maximization (EM)*, usado para determinar os parâmetros do *GMM* para um conjunto de padrões. Por ser um algoritmo iterativo ele atualiza os valores dos parâmetros do *GMM* em cada iteração, assim, tornando-o cada vez mais correlacionado ao conjunto de observações.

Começando de um modelo inicial  $\theta_0$ , a cada iteração, um novo modelo  $\theta_n + 1$  relaciona-se com o modelo anterior  $\theta_n$  obedecendo a relação:

$$g(\mathbf{X}|\theta_{n+1}) \geq g(\mathbf{X}|\theta_n)$$

esse processo é repetido até que um limiar de convergência seja alcançado.

A inicialização do *GMM* exige um modelo inicial, isto para que seja possível a estimação de um novo modelo. Há duas formas principais para se obter esse modelo inicial (Silva, 2014):

- Inicialização por agrupamento: as médias são inicializadas selecionando o centro de cada grupo, o número de grupos deve ser igual à quantidade de componentes gaussianas do modelo. Os pesos são inicializados uniformemente e a matriz de covariância é a diagonalizada;
- Inicialização aleatória: as médias são obtidas através do conjunto de treinamento, escolhendo-se vetores característicos aleatórios para a inicialização. A matriz identidade é usada para inicializar a matriz de covariância e os pesos também são inicializados uniformemente.

O critério de parada acontece quando o algoritmo *EM* alcança um máximo local, para isso ou ele deve alcançar o número de iterações ou quando a diferença relativa entre o modelo atual e o anterior for maior que um determinado limiar. Isso significa que o algoritmo encontrou os melhores parâmetros do modelo.

## 2.2. Processos Gaussianos

Uma das grandes vantagens de Aprendizado de Máquina é conseguir estimar funções com relações complexas entre as variáveis e cheias de não linearidade. Se quisermos encarar esses problemas com modelos lineares simples, podemos entender a forma funcional estimada usando mais parâmetros, o que pode ser obtido, por exemplo, fazendo uma expansão polinomial. Sempre podemos aumentar o grau do polinômio estimado, tornando-o arbitrariamente complexo, para assim conseguir ajusta-lo a dados bastante não lineares.

De uma forma bastante simplista, podemos entender Processos Gaussianos como um interpolador de dados. Assumindo que não há ruído, um interpolador prevê sempre o valor observado onde há dados. A parte complicada então é saber o que prever entre um e outro ponto. A intuição diz que nossas previsões não mudam bruscamente se as variáveis explicativas não mudarem muito. Por exemplo, se eu observe quando então devo esperar que seja próximo de zero quando por próximo de zero, digamos. Na Figura 1 podemos ver como Processos Gaussianos estão de acordo com

essa intuição e fazem uma interpolação bastante interessante. A imagem mostra três amostras de um processo gaussiano

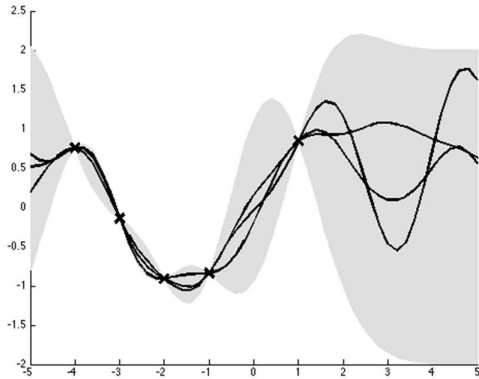


Figura 1. Processo Gaussiano.

Fonte: Murphy (2013).

definido pelas cinco observações marcadas com  $x$ . Note que, onde há dados, o valor previsto é o mesmo que o observado e não há incerteza.

Conforme nos afastamos dos dados, a incerteza da interpolação aumenta, denotada pelo alargamento do intervalo de confiança de 95% (em cinza).

Podemos definir um processo Gaussiano como um conjunto de variáveis que se desenvolvem no decorrer do tempo (ou espaço)  $f(x)$  é dito ser um Processo Gaussiano, se qualquer coleção finita retirado deste, possuem distribuição Normal Multivariada. Um Processo Gaussiano é completamente especificado por sua função de média e sua função de covariância, assim como a distribuição Normal Multivariada.

Denota-se um Processo Gaussiano por:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

onde  $m(x)$  é a função de média e  $k(x, x')$ , a função de covariância.

### 2.2.1. KERNEL

A noção de distância entre dois pontos é definida num Processo Gaussiano pelo que chamamos de kernel, uma função de dois pontos no espaço, geralmente não linear, que pode ser interpretada como a distância entre os seus pontos de entrada. Um kernel bastante popular é o exponencial quadrático, também conhecido como função de base radial ou kernel gaussiano (lembramos que existem muitos tipos de kernels):

$$k(x, x') = \sigma_f^2 \exp \left\{ -\frac{(x - x')^2}{2l^2} \right\}$$

Ele tem esse nome pois é igual a função gaussiana, removendo o fator de normalização. Isso lhe confere algumas propriedades de distância bastante interessantes.

Resumindo, uma função é um Processo Gaussiano se qualquer conjunto finito  $f(x_1), f(x_2), \dots, f(x_n)$  segue uma distribuição normal multivariada, onde tipicamente representam um vetor observação (linha) numa tabela de dados. Assim como uma gaussiana, um GP é definido por uma função de média  $m(x)$ , que é usualmente assumida como zero e por uma função de covariância, também conhecida por kernel.

A função de covariância é usada para construir uma matriz de covariância  $N \times N$ . Finalmente, podemos retirar amostras a priori de um GP com

$$f(x) \sim \mathcal{N}(m_x, K_{xx})$$

Em que  $m(x)$  é da mesma dimensão de  $x$  e tem dimensões  $X X^T$ , e como já havíamos dito antes geralmente assumimos que  $m(x)$  é zero.

## 3. Material e Métodos

### 3.1. Descrição dos Dados

Os dados utilizados são provenientes do seguinte *endpoint* de dados *pbp* (Play-by-Play) da NBA:

[data.nba.net/v2015/json/mobile\\_teams/nba/2019/scores/pbp/<GAME\\_ID>\\_full\\_pbp.json](https://data.nba.net/v2015/json/mobile_teams/nba/2019/scores/pbp/<GAME_ID>_full_pbp.json)

Onde a variável **GAME\_ID** é dada por uma combinação do código da etapa regular<sup>4</sup>, em nosso caso a temporada regular de 2019/20 teve código 002190, e do número do jogo com quatro dígitos. Por exemplo, o primeiro jogo da temporada regular de 2019/20 tem **GAME\_ID** = 0021900001 e url de consulta:

[data.nba.net/v2015/json/mobile\\_teams/nba/2019/scores/pbp/0021900001\\_full\\_pbp.json](https://data.nba.net/v2015/json/mobile_teams/nba/2019/scores/pbp/0021900001_full_pbp.json)

Os dados provenientes são arquivos do tipo *JSON* (JavaScript Object Notation) e contém diversas informações que vão desde a data do jogo, até as descrições dos lances dos jogos. Contudo, foi realizada uma “limpeza” e engenharia de atributos nesses dados a fim de remover variáveis inúteis para o estudo, por exemplo a data e horário do jogo, e criar novos recursos para os modelos utilizados.

Para a aplicação no modelo de misturas gaussianas o conjunto de dados resultante da limpeza e engenharia de atributos foi composto do:

<sup>4</sup>O código da temporada regular pode ser encontrado na busca de *boxscores* (resultados/placares) da temporada em questão: <https://www.nba.com/stats/teams/boxscores>

- *ID*: GAME.ID (Código do Jogo), combinação do código da temporada e número do jogo, com 10 dígitos;
- *Home*: Variável *Booleana* que indica se o time em estudo (*Houston Rockets*) está como Mandante na partida;
- *locX* e *locY*: Coordenadas do Jogador pontuador (*Houston Rockets*) na quadra.

Para o modelo de processos gaussianos, mais *features* foram adicionadas, como a variável de dependência temporal Minutos. Assim, o conjunto de dados foi composto do:

- *ID*: GAME.ID (Código do Jogo), combinação do código da temporada e número do jogo, com 10 dígitos;
- *Home*: Variável *Booleana* que indica se o time em estudo (*Houston Rockets*) está como Mandante na partida;
- *RivalName*: Nome da franquia (Time) rival;
- *Minute*: Minuto do jogo, compreendido entre 0 e 48;
- *Score*: Pontuação do time do *Houston Rockets* até o Minuto de referência;
- *RivalScore*: Pontuação do time Rival até o Minuto de referência.

### 3.2. Configurações dos Modelos e Ambiente

#### 3.2.1. MISTURAS GAUSSIANAS

O Modelo de misturas gaussianas requer a definição a priori do hiperparâmetro  $k$ , que define a dimensão da distribuição de misturas gaussianas. Para uma inicialização ótima, utilizamos um procedimento de otimização bayesiana da métrica de AIC, definido por:

$$k_{opt} = \arg \min_k AIC(L(k)) \quad (1)$$

em que  $L(k)$  é a função de verossimilhança do modelo dado  $k$ . Assim, como o critério de informação de Akaike (AIC) diminui conforme o modelo melhora (James et al., 2014), utilizamos um processo gaussiano para minimizar o problema definido na Equação 1. Ademais, o modelo de misturas gaussianas também admite a escolha da estrutura de covariância das  $k$  componentes, desse modo, escolhemos também a estrutura de covariância das componentes. A Figura 2 mostra a função de aquisição do processo de otimização bayesiana.

Como resultado do processo de otimização, foi escolhido o número de componentes  $k = 27$  e estrutura de covariância diagonal (cada componente tem sua própria matriz de covariância diagonal).

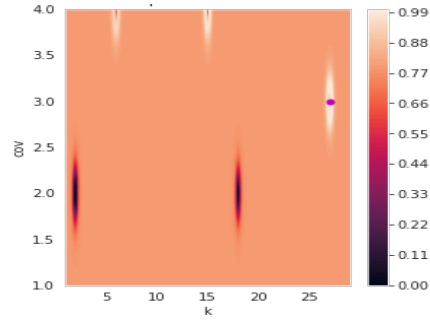


Figura 2. Função de Aquisição do Processo de Otimização Bayesiana do Modelo de Misturas Gaussianas.

#### 3.2.2. PROCESSOS GAUSSIANOS

Para uma melhor avaliação do modelo de processos gaussianos e a variabilidade exercida por cada tipo de kernel, foram realizados experimentos com dois tipos de kernel: *RBF* e *RBF+ExpSin*. Por termos uma forte tendência positiva nos dados (Pontuação Acumulada), optamos por esses dois tipos de kernel para modelar a tendência.

O treinamento do modelo foi realizado com partes compostas de 7 jogos seguidos, com o intuito de prever a série temporal completa de pontuação do próximo jogo. Para otimização dos parâmetros do kernel e do próprio modelo, foi utilizada o método *L-BFGS-B*, a Tabela 1 mostra os hiperparâmetros e suas restrições de suporte.

Hiperparâmetro	Suporte
$\sigma_y^2$	$[1e^{-30}, 100]$
$\sigma_{RBF}^2$	$[1e^{-30}, 100]$
$l_{RBF}^2$	$[1e^{-20}, 500]$
$\sigma_{ExpSin}^2$	$[1e^{-30}, 100]$
$l_{ExpSin}^2$	$[1e^{-20}, 500]$

Tabela 1. Hiperparâmetros e suportes associados para os modelos de processos gaussianos com kernel *RBF* e *RBF+ExpSin*

As variáveis independentes utilizadas para o treinamento do modelo foram a Pontuação dos *Rockets*, Pontuação do Rival, e a informação de Mandante (*Home*). Como janela de treinamento, foram escolhidos os sete jogos antecedentes ao jogo de predição.

Por fins de simplicidade, foram contabilizados apenas os 48 minutos regulares, ou seja, sem períodos de *OT* (*Over Time*, Acréscimos).

#### 3.2.3. AMBIENTE DO EXPERIMENTO

O experimento foi executado no ambiente Python do Google Colab<sup>5</sup>, com 13 GB RAM, processador Xeon 2.2 GHz

<sup>5</sup><https://colab.research.google.com>

e GPU Tesla T4 12 GB RAM. Para o framework de modelagem, foi escolhida a biblioteca do Scikit-learn<sup>6</sup> da linguagem Python. Para a tarefa de otimização bayesiana foi utilizada a biblioteca GPyOpt<sup>7</sup>.

#### 4. Resultados e Discussões

Para a avaliação de desempenho dos modelos de processos gaussianos com kernel *RBF* e *RBF+ExpSin*, foram escolhidos os últimos 65 jogos do time do *Houston Rockets* na temporada em questão. Assim, para cada jogo a ter sua pontuação minuto a minuto predita, foram treinados dois modelos de processos gaussianos com os kernels citados nos 7 jogos seguidos anteriores ao de predição. A Tabela 2 mostra o desempenho dos modelos, utilizando as métricas de  $R^2$  e *MAE* (*Mean Absolute Error*) nos 65 jogos preditos.

Métrica	Kernel	Min	$Q_1$	Média	Mediana	$Q_3$	Máx
$R^2$	RBF	0.5115	0.8890	0.9113	0.9449	0.9666	0.9907
	RBF+ExpSin	0.5106	0.8799	0.9075	0.9399	0.9639	0.9867
<i>MAE</i>	RBF	2.3574	4.7850	8.1478	6.1745	10.0557	25.8526
	RBF+ExpSin	2.3495	5.1299	10.8146	8.5890	13.2030	33.8727

Tabela 2. Métricas de desempenho para os modelos de processos gaussianos nos 65 cenários de treinamento da temporada regular 2019/20.

Pela análise da Tabela 2, aparentemente, o modelo com melhor desempenho preditivo é o de processos gaussianos com kernel *RBF*, possuindo maior  $R^2$  e menor *MAE*. Contudo, ao nível de confiança de 95%, o teste de *Kruskal-Wallis* mostrou que os dois modelos (GP *RBF* e GP *RBF+ExpSin*) possuem  $R^2$  estatisticamente iguais (*valor - p* < 0.01) e *MAE* estatisticamente iguais (*valor - p* < 0.01).

Para uma melhor visualização do potencial preditivo dos dois modelos, vamos avaliar as distribuições preditivas para quatro jogos arbitrários do *Houston Rockets* na Figura 3.

Visualmente as distribuições preditivas confirmam o resultado do teste de hipóteses da não existência de diferenças entre as métricas de desempenho para os dois modelos. Entretanto, o modelo mais simples (*RBF*) possui um leve melhor ajuste nos quatro jogos da Figura 3, comportamento visto na Tabela 2. Em relação ao ajuste minuto a minuto da pontuação, podemos verificar uma alta incerteza para ambos os modelos treinados, tendo intervalos de credibilidade abrangendo em média 40 pontos. Entretanto, os dois modelos conseguem, aparentemente, ter boa previsão média dos finais de cada partida.

Prosseguindo com a análise dos resultados, vamos avaliar o comportamento das distribuições de misturas gaussianas encontradas. A Figura 4 mostra a verossimilhança do modelo de misturas gaussianas, parametrizado conforme descrito na

Subseção 3.2.1, para o time do *Houston Rockets* em casa, i.e, como mandante do jogo.

A Figura 4 mostra uma alta densidade de pontos abaixo da cesta, no “garrafão”, o que nos dá uma visão de que o estilo de jogo do *Houston Rockets*, como mandante, estava bastante concentrado em *dunks* (enterradas) ou acertos de quadra de curta distância. Como um dos grandes contribuidores desse comportamento, podemos citar a excelente atuação do jogador pivô Clint Capela, com 62% de tentativas convertidas e incríveis 0 tentativas de lances fora do perímetro, durante toda a temporada regular, ou seja, a pontuação média por jogo de 13.9 se deve, exclusivamente, à sua atuação dentro do perímetro<sup>8</sup>. Contudo, outros jogadores também tem influência nesse comportamento, como por exemplo o jogador Russel Westbrook com uma média baixa de 16.6% de tentativas fora do perímetro, convertendo desses apenas 25.8% em pontos<sup>8</sup>.

Além de um estilo de jogo dentro do “garrafão”, o time do *Houston Rockets* na temporada 2019/20, dominou uma zona mais abrangente, a chamada “zona morta”, evidenciado pela partes inferiores da Figura 4. Como grandes contribuidores em acertos de quadra nos *corners* (cantos), podemos citar Danuel House com 36.3% de acertos de 3 pontos<sup>8</sup>, e o jogador Ben Mclemore com 40% de acertos de fora do perímetro<sup>8</sup>.

Com relação aos acertos de quadra fora do perímetro nas zonas chamadas “wings”(diagonais frente à cesta), temos leves densidades, mostrando que o real estilo de jogo do *Houston Rockets*, como mandante, foi concentrado na zona morta.

A Figura 5 mostra a verossimilhança do modelo de misturas gaussianas, parametrizado conforme descrito na Subseção 3.2.1, para o time do *Houston Rockets* fora de casa, i.e, como visitante.

Ao analisarmos a Figura 5 notamos um comportamento em quadra muito parecido com o time estando na situação de mandante. Entretanto, as regiões dos cantos aparentam ter uma maior variabilidade, e as diagonais fora do perímetro visualmente são mais densas. A maior variabilidade em locais de acertos de quadra e uma maior distribuição entre a localização dos pontuadores sugere a idéia de que o time do *Houston Rockets*, aparentemente, não tem um estilo de jogo dominante quando visitante. Esse padrão pode ser explicado, entre outros fatores, pela pressão psicológica exercida por diversas variáveis no basquete, entre eles: apoio da torcida, efeitos visuais, tempo de viagem, distanciamento da família, etc. Legaz-Arrese et al. (2013), mostra como diversos fatores somados ao “efeito casa” podem afetar o desempenho em um jogo. Harris & Roebber (2019), analisaram todas as temporadas da NBA anteriores à avaliada nesse

<sup>6</sup><https://scikit-learn.org>

<sup>7</sup><https://sheffieldml.github.io/GPyOpt>

<sup>8</sup><https://www.basketball-reference.com>



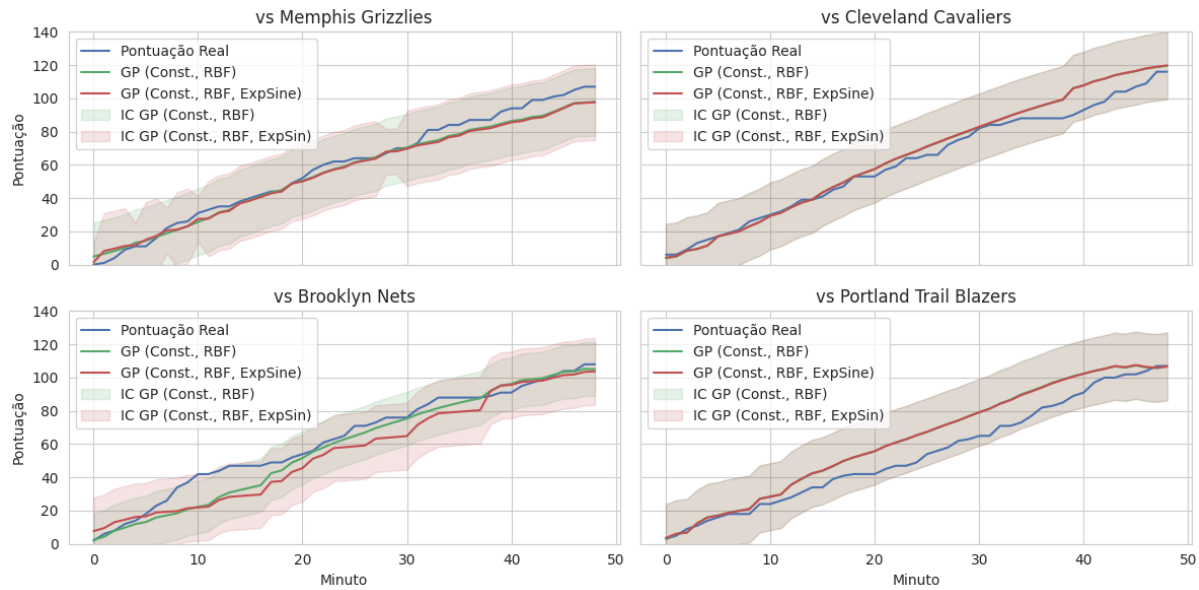


Figura 3. Distribuições preditivas dos modelos de processos gaussianos *RBF* e *RBF+ExpSin*, para quatro jogos arbitrários do *Houston Rockets* na temporada 2019/20

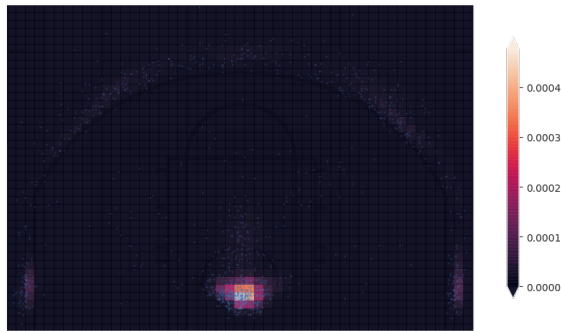


Figura 4. Verossimilhança do modelo de misturas gaussianas para acertos de quadra do *Houston Rockets* como Mandante do jogo.

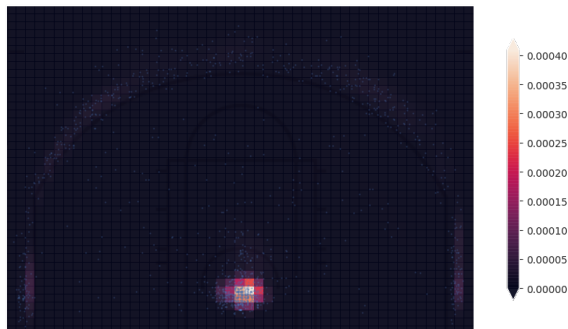


Figura 5. Verossimilhança do modelo de misturas gaussianas para acertos de quadra do *Houston Rockets* como Visitante.

artigo, e concluíram que a vantagem do time mandante é maior para franquias que fazem mais acertos de quadra dentro do perímetro. Tendo em vista o estilo do time texano, o comportamento dos *Rockets* em quadra tende a mudar, levemente, quando há confrontos fora de casa.

## 5. Conclusão

Ajuste de Modelos utilizando dados de pontuações em esportes é um desafio, devido à diversas variações e fatores latentes que podem contaminar os dados e reduzir a informação resultante. Nesse artigo, um modelo de processos gaussianos foi avaliado em um conjunto de séries temporais de pontuações acumuladas do time da NBA *Houston Rockets* na temporada regular 2019/20, para previsão da pontuação minuto a minuto de jogos futuros. Ademais, um modelo de misturas gaussianas foi ajustado aos dados para compreensão dos padrões de acertos de quadra do time em estudo. Nossas conclusões foram:

1. Para a problemática desse artigo, **modelos de processos gaussianos com kernel *RBF* tem desempenhos estatisticamente iguais do que os processos gaussianos com kernel *RBF+ExpSin***;
2. Seguindo a filosofia da *Navalha de Occam*<sup>9</sup>, **podemos utilizar um modelo de processos gaussianos com kernel *RBF* para modelar, de forma satisfatória**

<sup>9</sup>A navalha de Occam (também conhecida como a “lei da parcimônia”) é um princípio lógico onde a melhor solução é aquela que apresenta a menor quantidade de premissas possíveis

( $\bar{R}^2 = 0.9113$ ,  $\overline{MAE} = 8.1478$ ), **pontuações minuto a minuto de jogos futuros, dado jogos passados**;

- Através da distribuição das localizações acertos de quadra, inferimos que o comportamento em quadra do time do *Houston Rockets* mandante é mais "comportado" do que quando o time está como visitante, i.e, **os Rockets não tem um estilo de jogo bem definido e dominante quando estão fora de casa**.

Devido à simplicidade e flexibilidade dos processos gaussianos, o modelo com kernel *RBF* se torna uma ótima escolha por apresentar resultados equivalentes ao método mais complexo (*RBF+ExpSin*). Ademais, o modelo escolhido pode ser uma ferramenta aos amantes do esporte para ter uma razoável estimativa de como será o progresso de pontos ao decorrer da partida de forma prévia.

## 6. Trabalhos Futuros

Esse artigo é a primeira parte de um estudo do time do *Houston Rockets* nas temporadas 2019/20 e 2020/21. Devido ao péssimo rendimento na temporada 2020/21, o time texano quebrou recordes de derrotas seguidas da franquia e da própria NBA. Desse modo, se torna importante o estudo das mudanças que levaram um time que brigou pelo título na temporada 2019/20 a estar nas últimas colocações do campeonato na temporada seguinte (2020/21).

Apesar dos modelos ajustados terem apresentados resultados satisfatórios, podemos melhorar esses resultados, além de incluir novas informações que possam dar maior credibilidade ao estudo, como:

- Utilização dos dados da temporada 2020/21 (no momento da escrita desse artigo, a temporada ainda estava ativa);
- Utilização de Inferência Variacional Estocástica para Processos Gaussianos, de forma a podermos iterativamente melhorar a distribuição preditiva de um jogo futuro minuto a minuto;
- Adição de mais variáveis aos Modelos de Processos Gaussianos, como posição do time rival na tabela de classificação no momento do jogo, distância da sede do time (em casos de jogos fora de casa), bilheteria, W/L (Vitórias/Derrotas) nos últimos jogos;
- Criar um tipo de *score* das diferentes *lineups* (escalações de jogadores), possivelmente tendo relação direta com a pontuação, assistências, rebotes, etc.
- Aplicação de Modelos de Misturas Gaussianas para acertos de quadra de jogadores específicos, mostrando as "reais" posições onde os mesmos são eficazes;

Todos os códigos para essa análise e para futuras adições, podem ser encontrados no repositório: <https://github.com/damarals/RocketsAnalysis>

## Referências

- Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387310738.
- Harris, A. R. and Roebber, P. J. Nba team home advantage: Identifying key factors using an artificial neural network. *PLOS ONE*, 14(7):1–9, 07 2019. doi: 10.1371/journal.pone.0220630. URL <https://doi.org/10.1371/journal.pone.0220630>.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014. ISBN 1461471370.
- Legaz-Arrese, A., Moliner-Urdiales, D., and Munguía-Izquierdo, D. Home advantage and sports performance: Evidence, causes and psychological implications. *Universitas Psychologica*, 12, 08 2013. doi: 10.11144/Javeriana.UPSY12-3.hasp.
- Murphy, K. P. *Machine learning : a probabilistic perspective*. MIT Press, Cambridge, Mass. [u.a.], 2013. ISBN 9780262018029 0262018020.
- Silva, S. S. d. Segmentação de imagens utilizando combinação de modelos de misturas gaussianas, 2014.
- Zhang, H., Wen, T., Zheng, Y., Xu, D., Wang, D., Nguyen, T. M., and Wu, Q. M. Two fast and robust modified gaussian mixture models incorporating local spatial information for image segmentation. *J. Signal Process. Syst.*, 81(1):45–58, October 2015. ISSN 1939-8018. doi: 10.1007/s11265-014-0898-8. URL <https://doi.org/10.1007/s11265-014-0898-8>.