

# Diferenciando Perfis de Corredores por Meio de Pontos de Mudança nos Treinos

Nathália Tito<sup>1</sup>, Balthazar Paixão<sup>1</sup>, Lucas G. Tavares<sup>1</sup>,  
Eduardo Ogasawara<sup>1</sup>, Glauco F. Amorim<sup>1</sup>

<sup>1</sup>Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

{nathalia.tito, balthazar.paixao, lucas.tavares}@aluno.cefet-rj.br

eogasawara@ieee.org, glauco.amorim@cefet-rj.br

**Abstract.** *Street running has been attracting more practitioners worldwide. Concurrently, there is a noticeable spread of approaches based on time series and event detection in enhancing sports performance. In this context, this article analyzes and compares the points of change in the time series of individual training for more and less experienced runners. The results indicate a significant difference ( $p < 0.05$ ) between the proportions of change points for the two experience levels, providing an alternative indicator capable of differentiating athlete profiles, which can support increasingly specialized recommendation models by offering personalized feedback for each goal.*

**Resumo.** *A corrida de rua vem atraindo cada vez mais praticantes em todo o mundo. Em paralelo, nota-se uma disseminação de abordagens baseadas em séries temporais e detecção de eventos na potencialização de desempenho no âmbito do esporte. Nesse contexto, esse artigo analisa e compara os pontos de mudança nas séries temporais de treinos individuais de corredores mais e menos experientes. Os resultados encontrados indicam uma diferença significativa ( $p < 0,05$ ) entre as proporções de pontos de mudança para os dois níveis de experiência, fornecendo uma alternativa de indicador capaz de diferenciar perfis de atletas, que pode apoiar modelos de recomendação cada vez mais especialistas ao fornecer comentários personalizados para cada objetivo.*

## 1. Introdução

Os corredores urbanos participam de corridas populares por diversos motivos, incluindo ocupação do tempo, razões psicológicas, sociais, pessoais e competitivas [Parra-Camacho et al., 2019]. O aumento da popularidade dessas corridas contribui para a melhoria da qualidade de vida, ao ter um efeito positivo nos índices biomédicos de saúde [Hespanhol Junior et al., 2015]. Com o crescente número de participantes a cada ano, esse esporte está se tornando cada vez mais difundido. As tradicionais corridas de rua no Brasil refletem essa tendência, mostrando mudanças no desempenho e crescimento no número de competidores [Thuany et al., 2023].

Embora popular, a prática de correr pode ser desafiadora, e diversos fatores podem impedir que um atleta complete uma prova. A frequência da passada, por exemplo, pode afetar negativamente o desempenho na corrida [Mo and Chow, 2019]. Além disso, a escolha de uma estratégia inadequada pode levar à falha em alcançar objetivos, já que

diferentes perfis de corredores requerem diferentes abordagens para atingir suas metas [Ely et al., 2008].

Estudos indicam que o desempenho dos atletas pode ser previsto por variáveis exógenas [Gómez-Molina et al., 2017], permitindo ajustes estratégicos antecipados com recomendações personalizadas. Smyth et al. [2022] ressaltam a importância de sistemas de recomendação específicos, dado que atletas possuem perfis variados e exigem orientações distintas. Modelos de aprendizado de máquina integram essas recomendações personalizadas, como demonstrado por Berndsen et al. [2017], que associam o sucesso dos treinos a recomendações adaptáveis durante a corrida.

Neste contexto, este trabalho propõe a criação de indicadores para diferenciar atletas menos experientes dos mais experientes, utilizando a detecção de pontos de mudança nas séries temporais de treinos. Além de servir como um indicador de classificação de perfil, a implementação de uma análise *on-line* desses pontos de mudança poderia fornecer comentários em tempo real durante uma atividade, identificando momentos de mudanças comportamentais, e orientando ajustes nas ações dos atletas [Teune et al., 2022]. Este processo tem potencial de impacto, especialmente para corredores iniciantes que não têm acesso à orientação profissional.

## 2. Trabalhos Relacionados

Diversos estudos investigam os fatores que impactam o desempenho de um atleta na corrida. Com o uso de modelos de aprendizado de máquina, fatores antropométricos, fisiológicos e biomecânicos podem ser avaliados na predição de desempenho [Gómez-Molina et al., 2017]. A mudança no ritmo durante uma prova também pode estar relacionada a fatores climáticos [Ely et al., 2008]. O histórico de treinamento é um forte preditor de performance, com modelos de previsão de séries temporais sendo aplicados para prever o desempenho dos corredores com base em treinos passados [El-Kassabi et al., 2020].

O uso de métodos aplicados a séries temporais está crescendo no âmbito esportivo. Komitova et al. [2022] realizou uma revisão completa sobre o uso de séries temporais, sobretudo no esporte, exemplificando casos de uso e abordando as oportunidades na área. Por exemplo, Yong et al. [2020] utilizou um modelo de média móvel autorregressiva (ARMA) para prever o resultado médio de futuras maratonas, a partir de séries temporais dos resultados passados da Maratona de Boston.

Na detecção de eventos em séries temporais no esporte, Corbett et al. [2019] analisou a produção física dos jogadores de futebol como uma série temporal, avaliando mudanças estratégicas durante uma partida e identificando picos de intensidade. A abordagem de pontos de mudança também foi explorada por Teune et al. [2022], que analisou o comportamento de estatísticas descritivas antes e após pontos de mudança no desempenho de atletas em partidas de futebol. Na corrida, Van den Berghe et al. [2020] estudou a detecção de alterações no pico de aceleração tibial dos atletas, utilizando *feedback* interativo musical em tempo real, com base no paradigma de aprendizagem por reforço.

Apesar da disseminação dessas abordagens, a aplicação do conceito de pontos de mudança na corrida como indicador de classificação de perfil ainda não foi amplamente explorada na literatura. Portanto, o trabalho aqui proposto visa diferenciar atletas mais experientes dos menos experientes através da detecção de pontos de mudança, baseando-se

na hipótese de que corredores mais experientes mantêm velocidades constantes, enquanto os menos experientes apresentam maiores oscilações [Ely et al., 2008].

### 3. Método

Nesta seção, é apresentada uma visão geral das etapas principais do método proposto. O objetivo é descrever como os dados foram obtidos, tratados e transformados, elencando os instrumentos e técnicas utilizadas para converter as informações em séries temporais, e posteriormente analisá-las sob a perspectiva de detecção de eventos.

#### 3.1. Coleta de Dados

Todos os dados utilizados nas análises foram cedidos por voluntários praticantes de corrida de rua e usuários do aplicativo Strava<sup>1</sup>. O dispositivo registra uma ampla gama de informações durante as corridas, incluindo dados como velocidade, distância, altitude, ritmo cardíaco e localização geográfica. Essas informações são registradas de forma contínua, a cada segundo, proporcionando um conjunto detalhado e granular de dados para análise.

#### 3.2. Conversão para Séries Temporais

Os dados brutos foram fornecidos no formato GPX (GPS Exchange Format), exigindo uma etapa de pré-processamento para estruturá-los de forma a torná-los mais acessíveis para análise. Devido à captura detalhada das informações, segundo a segundo, foi viável transformar a variável de interesse — neste caso, a velocidade — em séries temporais.

#### 3.3. Limpeza de Dados

Após a conversão, os dados passaram por um processo de limpeza, excluindo-se registros com velocidades superiores a 35 km/h, considerando improvável que corridas ocorram a essa velocidade, evitando assim a inclusão de dados de outros esportes, como ciclismo. Treinos fora da frequência de segundos foram descartados. Além disso, para evitar incluir treinos iniciais de atletas experientes, o primeiro ano do histórico desses atletas foi excluído. Em seguida, foi selecionada uma amostra aleatória de vinte corridas para cada perfil, respeitando os critérios de descarte mencionados.

#### 3.4. Agregação Temporal

Devido à alta volatilidade das séries temporais, determinada por sua frequência, foi necessário aplicar uma agregação temporal. Utilizamos, portanto, uma janela de 30 segundos. No trabalho de Teune et al. [2022], uma janela de 60 segundos foi sugerida e também testada nesse trabalho, mas não utilizada devido à avaliação de perda de informação.

#### 3.5. Detecção de Ponto de Mudança

No contexto da detecção de eventos, para este estudo, focou-se especificamente na detecção de pontos de mudança. Conceitualmente, pontos de mudança são intervalos de tempo em uma série temporal nos quais as propriedades dos dados mudam repentinamente [Takeuchi and Yamanishi, 2006].

---

<sup>1</sup>Disponível em: <https://www.strava.com/>. Último acesso em: 02 de julho de 2024.

Todos os métodos<sup>2</sup> do Harbinger para detecção de pontos de mudança foram testados. No entanto, optou-se pelo método de Suavização Exponencial (ETS) devido à sua capacidade de lidar de maneira eficaz com a volatilidade e as flutuações inerentes aos dados temporais, pois considera não apenas a tendência e sazonalidade, mas também os componentes de erro, o que permite uma modelagem mais precisa e adaptativa das variações dos dados ao longo do tempo.

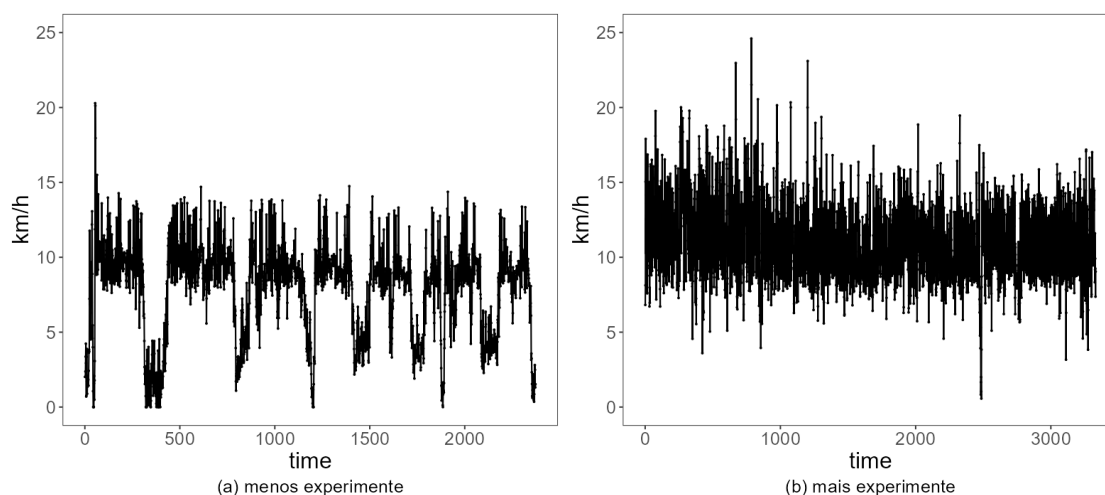
### 3.6. Implementação

Para limpeza e extração das séries temporais a partir dos dados de GPS, utilizou-se a linguagem de programação Python em conjunto com a biblioteca Gpxpy<sup>3</sup>, que permite o acesso e a manipulação dos dados no formato GPX. As análises das séries temporais e a detecção de pontos de mudança foram realizadas utilizando a linguagem de programação R e o pacote Harbinger, um *framework* para detecção de eventos em séries temporais [Salles et al., 2020].

## 4. Resultados

Como uma prova de conceito, seis voluntários, sendo três homens e três mulheres, forneceram consentimento para participar deste estudo. Os atletas foram previamente classificados com base na experiência e familiaridade com a corrida, conforme informado no momento da coleta dos dados. Três voluntários foram classificados como experientes, e para isso, precisavam ter praticado corrida regularmente por um ano ou mais.

Antes de realizar a agregação temporal nas séries e a detecção de eventos, foi conduzida uma análise exploratória preliminar para compreender o comportamento da velocidade nas duas amostras. A Figura 1 apresenta exemplos de treinos dos perfis, ilustrando suas velocidades ao longo do tempo.

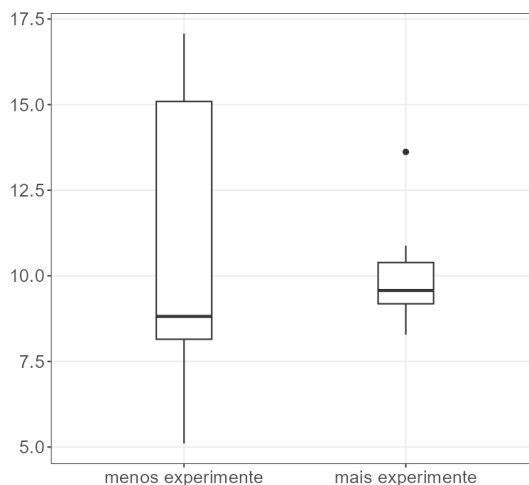


**Figura 1. Velocidade de um treino de um atleta menos e mais experiente**

<sup>2</sup>Disponível em: [https://nbviewer.org/github/cefet-rj-dal/harbinger/tree/master/change\\_point](https://nbviewer.org/github/cefet-rj-dal/harbinger/tree/master/change_point). Último acesso em: 19 de agosto de 2024.

<sup>3</sup>Disponível em: <https://pypi.org/project/gpxpy/>. Último acesso em: 19 de agosto de 2024.

Para avaliar se a média da velocidade poderia ser um indicador de distinção entre os perfis, foi calculada a média da velocidade para cada corrida em ambas as amostras. O *boxplot* na Figura 2 mostra a distribuição dessas médias, destacando uma maior dispersão no perfil dos menos experientes.



**Figura 2. Velocidade média das amostras de treinos dos atletas menos e mais experientes**

Um teste de *Levene*<sup>4</sup> foi realizado para avaliar a homogeneidade das variâncias nas duas amostras de médias. Os resultados rejeitaram a hipótese nula de homogeneidade de variâncias ( $p < 0,05$ ), indicando que as variâncias dos grupos são significativamente diferentes. Em seguida, um teste não paramétrico de *Wilcoxon*<sup>5</sup> foi aplicado para testar a similaridade entre as médias das velocidades, resultando na aceitação da hipótese de similaridade ( $p > 0,05$ ). Isso sugere que as médias das velocidades não são estatisticamente diferentes entre os perfis, inviabilizando o uso dessa métrica como indicador de classificação.

A seguir, as séries temporais agregadas foram analisadas, e a proporção de pontos de mudança em relação ao tamanho de cada série foi calculada para ambos os perfis. A média total do percentual de pontos de mudança foi de 8,6% nas séries dos menos experientes e 6,7% nas séries dos mais experientes. Um teste de *Wilcoxon* foi novamente utilizado para testar se essas proporções diferiam significativamente entre os perfis. O resultado rejeitou a hipótese de similaridade ( $p < 0,05$ ), validando o uso deste indicador para diferenciação de perfis. A Tabela 1 expõe o resultado desses testes.

A detecção de pontos de mudança nas séries temporais permite uma melhor interpretação do *antes* e *depois* dos eventos, o que pode ser útil tanto para os corredores quanto para os treinadores [Teune et al., 2022]. Adicionalmente, foi analisado o comportamento da velocidade após o primeiro ponto de mudança, partindo da hipótese de que corredores menos experientes podem ter dificuldade em controlar a velocidade após uma mudança brusca. Para os menos experientes, a média da velocidade variou de 11,34 km/h para 10,45 km/h, enquanto para os mais experientes variou de 10,05 km/h para 9,59 km/h. Em

<sup>4</sup>Disponível em: <https://www.rdocumentation.org/packages/lawstat/versions/3.2/topics/levene.test>.

<sup>5</sup>Disponível em: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/wilcox.test>.

**Tabela 1. Resultados dos testes para os indicadores**

| Indicadores            | Menos experientes | Mais experientes | <i>p</i> -value |
|------------------------|-------------------|------------------|-----------------|
| Desvios das médias     | 3,75              | 1,15             | 0,001**         |
| Médias das velocidades | 10,65 km/h        | 9,85 km/h        | 0,77            |
| Pontos de mudança (%)  | 8,6%              | 6,7%             | 0,01*           |

\*\* $p < 0.001$ ; \* $p < 0,01$ .

ambos os casos, os testes de *Wilcoxon* confirmaram a igualdade das médias de velocidade ( $p > 0,05$ ) antes e depois do ponto de mudança, como mostra a Tabela 2. Esse fato sugere que ambos os perfis de corredores mantêm um padrão de velocidade, tanto antes quanto depois da primeira mudança brusca.

**Tabela 2. Resultados dos testes por perfil das médias das velocidades antes e depois do primeiro ponto de mudança**

| Perfil            | Antes      | Depois     | <i>p</i> -value |
|-------------------|------------|------------|-----------------|
| Menos experientes | 11,34 km/h | 10,45 km/h | 0,1             |
| Mais experientes  | 10,05 km/h | 9,59 km/h  | 0,06            |

## 5. Limitações

Esse estudo não está isento de limitações. A amostra de participantes foi relativamente pequena e um número maior poderia fornecer resultados mais robustos e generalizáveis. Outra limitação está na seleção das amostras. Não foram estabelecidos critérios para a amostragem dos treinos, o que pode ter levado à comparação de tipos de treinos diferentes, como treinos em ladeira com treinos em terreno plano. Adicionalmente, a análise concentrou-se apenas na métrica velocidade, mas sabe-se que variáveis como distância percorrida, o terreno e as condições climáticas poderiam influenciar o desempenho dos corredores. Por fim, a detecção de pontos de mudança foi realizada utilizando Suavização Exponencial (ETS). Embora essa técnica tenha se mostrado eficaz, a exploração de outras metodologias de detecção de eventos poderia conferir resultados adicionais.

## 6. Considerações Finais

O objetivo do presente estudo foi identificar indicadores capazes de classificar atletas em mais e menos experientes na corrida, por meio da detecção de eventos em suas séries temporais de treino. A abordagem de detecção de pontos de mudança na corrida, em tempo real, pode apoiar modelos especializados de recomendação, além de fornecer alerta e sugestões para tomadas de decisão.

A análise revelou que atletas menos experientes apresentam mais eventos em suas séries, e esse indicador demonstrou-se promissor na classificação dos perfis dos atletas, considerando que a média da velocidade não apresentou relevância estatística.

Para futuros estudos, pretende-se ampliar a amostra de participantes, a inclusão de outras variáveis relevantes, a aplicação de diferentes metodologias de detecção de eventos para uma análise mais abrangente e detalhada do desempenho dos corredores, e com isso testes que comprovem a eficácia da classificação proposta.

## Referências

- Berndsen, J., Lawlor, A., and Smyth, B. (2017). Running with recommendation. In *CEUR Workshop Proceedings*, volume 1953, pages 18 – 21.
- Corbett, D., Sweeting, A., and Robertson, S. (2019). A change point approach to analysing the match activity profiles of team-sport athletes. *Journal of Sports Sciences*, 37.
- El-Kassabi, H. T., Khalil, K., and Serhani, M. A. (2020). Deep learning approach for forecasting athletes' performance in sports tournaments. In *ACM International Conference Proceeding Series*, pages 203 – 208.
- Ely, M. R., Martin, D. E., Chevront, S. N., and Montain, S. J. (2008). Effect of ambient temperature on marathon pacing is dependent on runner ability. *MSSE*, 40(9):1675 – 1680.
- Gómez-Molina, J., Ogueta-Alday, A., Camara, J., Stickley, C., Rodríguez-Marroyo, J. A., and García-López, J. (2017). Predictive variables of half-marathon performance for male runners. *Journal of Sports Science and Medicine*, 16(2):187 – 194.
- Hespanhol Junior, L., Pillay, J., and van Mechelen, W. (2015). Meta-analyses of the effects of habitual running on indices of health in physically inactive adults. *Sports Medicine*, 45:1455–1468.
- Komitova, R., Raabe, D., Rein, R., and Memmert, D. (2022). Time Series Data Mining for Sport Data: a Review. *IJCSS*, 21(2):17 – 31.
- Mo, S. and Chow, D. H. (2019). Reliability of the fluctuations within the stride time series measured in runners during treadmill running to exhaustion. *Gait and P.*, 74:1 – 6.
- Parra-Camacho, D., González-Serrano, M. H., González-García, R. J., and Calabuig Moreno, F. (2019). Sporting habits of urban runners: Classification according to their motivation. *International Journal of Environmental Research and Public Health*, 16(24).
- Salles, R., Escobar, L., Baroni, L., Zorrilla, R., Ziviani, A., Kreischer, V., Delicato, F., Pires, P. F., Maia, L., Coutinho, R., Assis, L., and Ogasawara, E. (2020). Harbinger: Um framework para integração e análise de métodos de detecção de eventos em séries temporais. In *Anais do Simpósio Brasileiro de Banco de Dados*, pages 73–84. SBC.
- Smyth, B., Lawlor, A., Berndsen, J., and Feely, C. (2022). Recommendations for marathon runners: on the application of recommender systems and machine learning to support recreational marathon runners. *User Modeling and User-Adapted Interaction*, 32(5):787 – 838.
- Takeuchi, J. and Yamanishi, K. (2006). A unifying framework for detecting outliers and change points from time series. *IEEE Transactions on Knowledge and Data Engineering*, 18(4):482–492.
- Teune, B., Woods, C., Sweeting, A., Inness, M., and Robertson, S. (2022). A method to inform team sport training activity duration with change point analysis. *PLOS ONE*, 17(3):1–11.
- Thuany, M., Vieira, D., Villiger, E., Gomes, T. N., Weiss, K., Nikolaidis, P. T., Sousa, C. V., Scheer, V., and Knechtle, B. (2023). An analysis of the São Silvestre race between 2007–2021: An increase in participation but a decrease in performance. *SMHS*, 5(4):277 – 282.
- Van den Berghe, P., Gosseries, M., Gerlo, J., Lenoir, M., Leman, M., and De Clercq, D. (2020). Change-point detection of peak tibial acceleration in overground running retraining. *Sensors*, 20(6).
- Yong, W., Lingyun, P., and Jia, W. (2020). Statistical analysis and ARMA modeling for the big data of marathon score. *Science and Sports*, 35(6):375 – 385.