Time-aware analysis of overlapping cycling segments from activity tracker data

1st Tilen Hliš
Faculty of Electrical Engineering and Computer Science, University of Maribor

2nd Iztok Fister
Faculty of Electrical Engineering and Computer Science, University of Maribor

3rd Patrik Rek
Faculty of Electrical Engineering and Computer Science, University of Maribor

4th Iztok Fister Jr.
Faculty of Electrical Engineering and Computer Science, University of Maribor

***Abstract*—Artificial Intelligence is used increasingly in sports to extract actionable insights and enhance athlete performance. The widespread adoption of sports activity trackers has resulted in large volumes of detailed training data. This paper presents an improved method for analyzing cycling data from TCX files. Overlapping path segments are detected across sessions using spatial filtering and Fréchet distance, and data fusion with OpenStreetMap provides contextual information. An experiment explores how a cyclist's performance varies on shared segments, depending on the time of day, revealing temporal patterns in training behavior.**

***Index Terms*—Cycling data, Data enrichment, GPS tracking, Path overlap, Time-aware analysis.**

I. Introduction

In recent years, Global Positioning System (GPS) enabled devices and wearables have become standard tools among cyclists for tracking performance, monitoring physiological metrics, and analyzing training patterns. These devices generate substantial amounts of structured data, commonly exported in formats such as Training Center XML (TCX) [1], [2]. TCX files contain detailed records of a rider's location, speed, heart rate, cadence, and power output regularly throughout each activity session.

Analyzing cycling data captured by GPS presents notable challenges. GPS inaccuracies and noise frequently complicate identifying repeated cycling segments accurately across multiple sessions [3]. In this context, a segment refers to a defined portion of a route—typically between two geographical points—over which athletes' performances (such as time or speed) are measured and compared, often using GPS data and digital platforms [4]. Moreover, variability in cycling paths and environmental factors can affect the reliability of segment-based performance analyses significantly. Addressing these challenges requires advanced techniques for detecting and evaluating overlapping segments accurately and robustly.

Recent research by Hliš et al. [5] has demonstrated the value of digital twin concepts for cycling, using graph-based representations to enable property and event analytics on fitness tracker data. Building on these foundations, this work focuses on the temporal dimension of performance analysis across recurring cycling segments.

This paper presents an improved version of the Julia-based library `TCX2Graph.jl` [6], [7], designed to analyze cycling data from TCX files systematically. The new version extends our previous work by introducing a more robust and accurate segment detection pipeline that combines advanced spatial filtering techniques and discrete Fréchet distance calculations. This enables precise detection of overlapping cycling segments, even in GPS inaccuracies. Additionally, each identified segment is enriched with detailed contextual metadata from the OpenStreetMap (OSM) [8], providing insights into the road conditions and environmental context. This enrichment is accomplished through data fusion, which integrates information from multiple data sources to produce more consistent, accurate, and useful information than any individual data source [9].

The primary research objective of the study is to investigate how performance metrics, such as speed, heart rate, and segment duration, vary when the same cycling segment is ridden under different temporal conditions. We examine variations based on the time of day (morning vs. afternoon), day of the week (weekday vs. weekend), and longitudinal trends over multiple rides. Such temporal differences may result from circadian physiological effects [10], varying traffic conditions, environmental influences, or fatigue accumulation.

The dataset analyzed in this study comprises 462 TCX files collected from repeated cycling sessions by a single athlete [11]. The overlapping segments were detected automatically using the proposed algorithm, and candidate segments were subsequently filtered based on predefined criteria, including segment length, flatness, and frequency of use. From these candidates, a single segment was selected manually for detailed temporal analysis, due to its consistent characteristics and frequent usage.

The main contributions of this work include:

- an updated and extended version of the Julia-based `TCX2Graph.jl` library [7], featuring robust overlapping segment detection,
- implementation of an efficient pipeline for segment detection, combining candidate generation, KD-tree spatial filtering, and discrete Fréchet distance validation,
- enrichment of the detected segments with detailed contextual metadata from the OpenStreetMap (OSM),
- empirical case studies investigating temporal effects (time-of-day, weekday/weekend, longitudinal progression) on cycling performance metrics within a frequently used segment.

The paper proceeds as follows: Section II presents the data preprocessing steps and the construction of the property graph. Section III details the implementation of the `TCX2Graph.jl` library, including algorithms for segment detection and data enrichment. Section IV describes the experimental design, presents and analyzes the results. Section V concludes with a discussion of the findings and suggestions for future work.

II. Materials and methods

This section describes the TCX data used in this study, outlines the structure and key attributes of TCX files, and explains how the cycling data are organized into a property graph for further analysis.

A. Structure of the TCX files

The data for this research comprises TCX files, a Garmin-developed XML format used commonly by GPS-enabled fitness tracking devices to record detailed activity data. These files include comprehensive metrics, such as GPS coordinates (latitude and longitude), altitude, heart rate, cadence, speed, and power, structured into activity sessions composed of laps and individual trackpoints.

The detailed structure of TCX files includes:

- **Activities**: Containing overall session metadata, such as the activity type, date, distance, etc.
- **Laps**: Subdivisions of activities, encapsulating metrics such as duration, distance, and physiological parameters.
- **Trackpoints**: Individual records capture precise data points at regular intervals, including spatial coordinates and performance metrics.

B. Property graph representation

The property graph model is defined formally as:

$$\langle V, E, P, Q, \alpha, \beta, \gamma \rangle, \tag{1}$$

where V are the vertices (trackpoints), E the directed edges, P the property keys, and Q the property values. The mappings α, β, and γ define connectivity and property assignments.

III. The proposed method

Detecting overlapping cycling path segments improves upon our previously published method [7], which performs basic overlap detection. In this work, we introduce a more robust pipeline that adds KD-tree spatial prefiltering and discrete Fréchet distance, enabling more accurate identification of shared segments despite the GPS inaccuracies. The core pipeline for robust overlapping segment detection is summarized in Algorithm 1, which employs the following functions:

Algorithm 1 Detection of overlapping cycling segments (`find_overlapping_segments`)

Require: GPS data $\mathcal{D}$, paths $\mathcal{P}$, ref ride r_{ref}, min len L_{min}, tol tol, step $step$, min rides min_{rides}, margin $margin$, dedup frac $dedup\_frac$
Ensure: List of detected overlapping segments

kdtrees $\leftarrow$ **build_kdtree**($\mathcal{P}$)
for all (s, e) $\in$ **gen_candidates**(r_{ref}, L_{min}, $step$) **do**
  poly $\leftarrow$ **make_polyline**($\mathcal{D}$, r_{ref}, s, e)
  circle $\leftarrow$ **bounding_circle**(poly, tol, $margin$)
  matches $\leftarrow$ [ ]
  for all ride $\in \mathcal{P}$ **do**
    for all win $\in$ **windows_in_circle**(ride, circle) **do**
      if frechet(poly, win) $\leq tol$ **then**
        Add (ride, win) to matches
      end if
    end for
  end for
  if unique(matches) $\geq min_{rides}$ **then**
    if not **dup**(s, e, results, $dedup\_frac$) **then**
      Add **format**(s, e, matches) to results
    end if
  end if
end for
return results

- **build_kdtree**: Builds KD-trees for all rides.
- **gen_candidates**: Generates all (start, end) segment index pairs in the reference ride.
- **make_polyline**: Extracts the GPS polyline for a segment.
- **bounding_circle**: Computes the spatial search region for a candidate segment.
- **windows_in_circle**: Returns sliding windows in a ride within the bounding circle.
- **frechet**: Calculates the discrete Fréchet distance between two polylines.
- **unique**: Counts unique rides in the matches.
- **dup**: Checks if a candidate is a duplicate of any accepted result.
- **format**: Formats and stores detected segments and matches.

In our study, Algorithm 1 is implemented as an upgraded version of the Julia-based library `TCX2Graph.jl` [6], [7] with an additional option of detecting overlapped segments. In line with this, it can explore not only TCX files, but also data produced by OSM. As a result, the upgraded `TCX2Graph.jl` is structured into several key components, each performing specialized tasks, as follows:

- data fusion with OSM,
- advanced overlapping segment detection.

The tasks mentioned in detail are discussed in detail in the remainder of the paper.

A. Data fusion with the OpenStreetMap

Data fusion in `TCX2Graph.jl` enhances raw GPS data by integrating additional road features from OSM. Each trackpoint extracted from the TCX files initially contains basic attributes such as latitude, longitude, altitude, and other cycling metrics (e.g., heart rate, cadence, speed, power). The library enriches this dataset by querying OSM's Overpass API [12] for extensive contextual data, including road attributes [13]. The fusion process occurs in batches to optimize API usage, improving the granularity and utility of the GPS data for subsequent analyses significantly.

After the data fusion step, each trackpoint is structured with the attributes presented in Table I.

TABLE I
ATTRIBUTES OF A TRACKPOINT AFTER OSM DATA FUSION.

Attribute			
TCX		OSM	
altitude	cadence	lane_markings	smoothness
longitude	speed	lit	surface
latitude	maxspeed	crossing	landuse
heart_rate	time	width	
power	distance	incline	
file_name		barrier	

B. Advanced overlapping segment detection

To identify overlapping segments across multiple cycling sessions reliably, `TCX2Graph.jl` employs a robust spatial indexing and similarity assessment approach with detailed algorithmic steps:

1) **Reference ride selection**: A reference ride is selected from the available TCX files (Figure 1). Each segment must meet a minimum length specified by the parameter L_{min} in Algorithm 1. This ensures that only meaningful segments are evaluated.
2) **Spatial prefiltering using KD-tree**: A KD-tree [14]–[16] spatial index is built for each other ride. For every candidate segment, a bounding circle is defined using the segment's extent and the margin parameter $margin$ in Algorithm 1 (Figure 2). The KD-tree is used to identify rides falling within this bounding circle, reducing unnecessary distance calculations.
3) **Candidate segment evaluation**: The filtered trackpoints from other rides are compared to the candidate segment using the discrete Fréchet distance [17] (Figure 3). This accounts for the shape and sequencing of the points, handling GPS noise, and temporal misalignments.
4) **Distance computation and validation**: If the computed Fréchet distance between the candidate and matching segments is below the defined threshold tol in Algorithm 1, the match is considered a potential overlap.

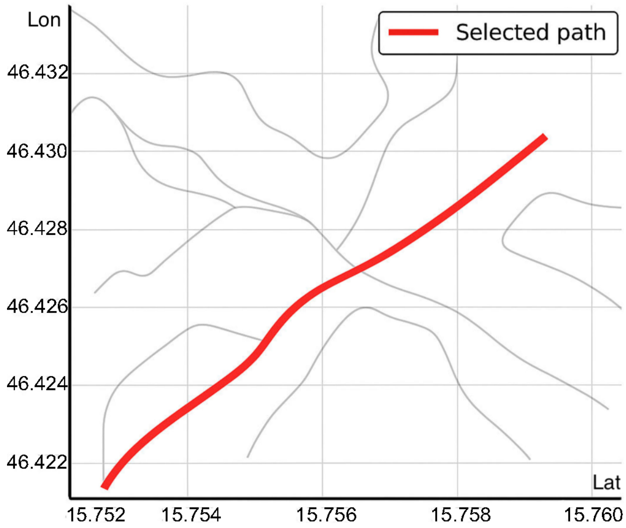

Fig. 1. GPS paths from TCX files; red line shows reference path.

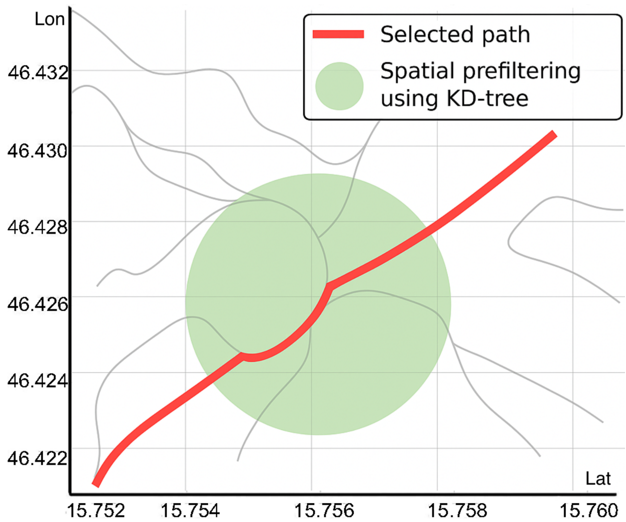

Fig. 2. Representation of KD-tree based prefiltering.

5) **Threshold-based filtering**: The segment is retained only if it is traversed by at least min_{rides} in Algorithm 1 different rides, ensuring the robustness of overlap detection.
6) **Deduplication procedure**: Before adding a new segment to the result set, it is checked against already accepted segments using the overlap fraction threshold $dedup\_frac$ in Algorithm 1. Segments that overlap existing ones significantly are discarded to preserve uniqueness.

IV. EXPERIMENTS AND RESULTS

Our experimental work aimed to show that fusion of data from two sources, i.e., TCX activity files and OSM geospatial data and information, improves the time-aware analysis of overlapping segments substantially. All analyses are conducted using the updated detection pipeline implemented in the improved version of `TCX2Graph.jl` [7], which provides

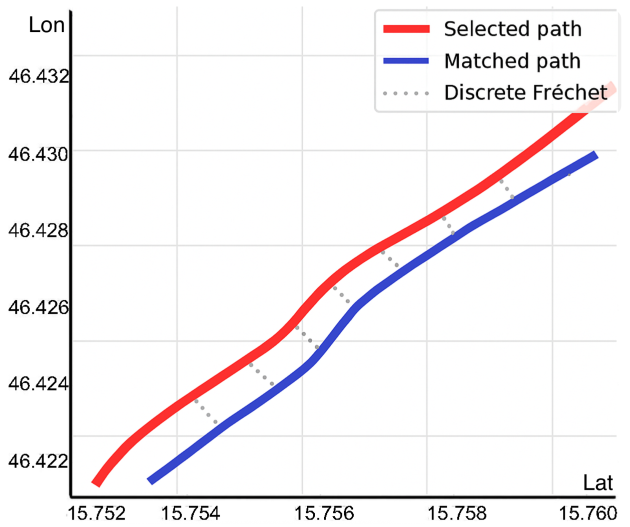

Fig. 3. Evaluation of candidate segment against other rides using discrete Fréchet distance.

higher accuracy and new analytical capabilities compared to our earlier work.

The underlying dataset consists of 462 TCX files collected from repeated cycling sessions of a single cyclist. This collection enables visualization and further analysis. Figure 4 is a graphical representation of a property graph with raw data from the TCX files.

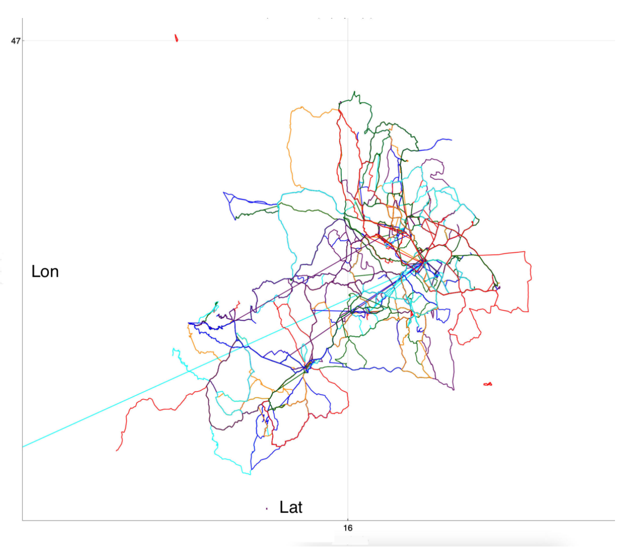

Fig. 4. Representation of a property graph (462 tcx files).

The overlapping segments were identified using the detection pipeline, and filtered based on segment length, elevation stability, and frequency of traversal. The main parameter settings for the overlapping segment detection algorithm are summarized in Table II.

A single segment—1,027 meters long and used in 52 different rides—was selected for detailed temporal analysis, due to its consistency and rich coverage. The characteristics

TABLE II
PARAMETER SETTINGS FOR THE OVERLAPPING SEGMENT DETECTION ALGORITHM

Parameter	Symbol	Value
Reference ride index	r_{ref}	50
Minimum segment length	L_{min}	1000.0 m
Tolerance	tol	75.0 m
Window step	$step$	1
Minimum number of rides	min_{rides}	52
Prefilter margin	$margin$	150.0 m
Deduplication threshold	$dedup\_frac$	0.5

of the selected segment are summarized in Table III, and its

TABLE III
CHARACTERISTICS OF THE SELECTED SEGMENT.

Attribute	Value
Length	1027 meters
Surface type	Asphalt
Smoothed elevation gain (per ride)	1.85 m (min: 0.00 m, max: 4.12 m)
Number of rides covering segment	52
Road type	Rural

usage across 52 rides is visualized in Figure 5, where start and end points of overlapping segments differ slightly across rides due to the tolerance parameter used in the algorithm.

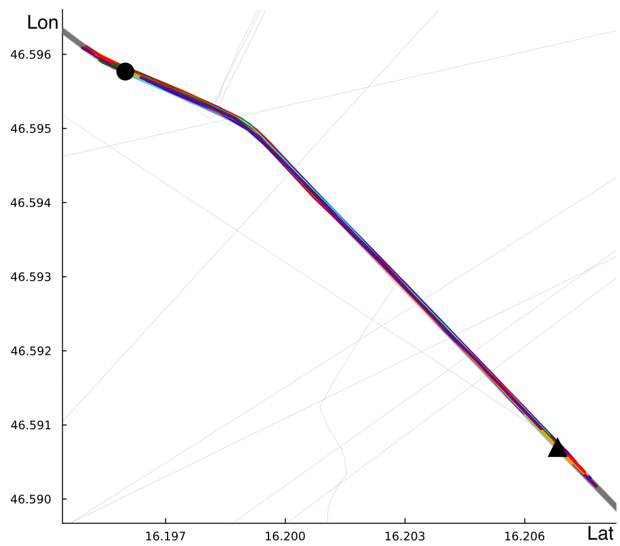

Fig. 5. Selected segment (visualization of 52 rides).

Importantly, the preliminary analysis revealed that the selected segment appears at a consistent relative position in each ride, typically, within the first 40% of the ride distance. This uniform positioning minimizes potential confounding from accumulated fatigue, making the segment well-suited for fair temporal comparisons.

The selected segment satisfies the key criteria for temporal performance comparisons:

- **Sufficient duration:** The segment spans approximately 1 km, providing enough time (typically 1–2 minutes) for heart rate, speed, and cadence to stabilize [18], [19].
- **Consistent profile:** The flat terrain is essential for isolating temporal factors without confounding the effects from elevation [20].

All the comparisons were conducted using Welch's two-sample *t*-test for unequal variances [21]. Linear regression was used to evaluate temporal performance trends. Three case studies were performed, as follows:

- case study 1: morning vs. afternoon performance,
- case study 2: weekday vs. weekend performance,
- case study 3: performance progression over time.

A. Case study 1: Morning vs. Afternoon performance

Research question 1 (RQ1): Does the cyclist's performance on the selected segment differ between morning and afternoon rides?

The rides were grouped by start time into **morning** (before 12:00, $n = 21$) and **afternoon** (12:00 or later, $n = 31$). Each ride was evaluated based on average speed, heart rate, and segment duration. Performance differences between the morning and afternoon sessions are shown in Table IV, with the distribution of speeds visualized in Figure 6.

TABLE IV
PERFORMANCE BY TIME OF DAY

Metric	AM (n=21)	PM (n=31)	*p*-value
Avg. Speed (km/h)	28.29 ± 2.82	31.36 ± 4.49	**0.004**
Avg. Heart Rate (bpm)	136.21 ± 11.35	141.13 ± 11.80	0.139
Duration (s)	129.33 ± 13.81	116.87 ± 17.89	**0.007**

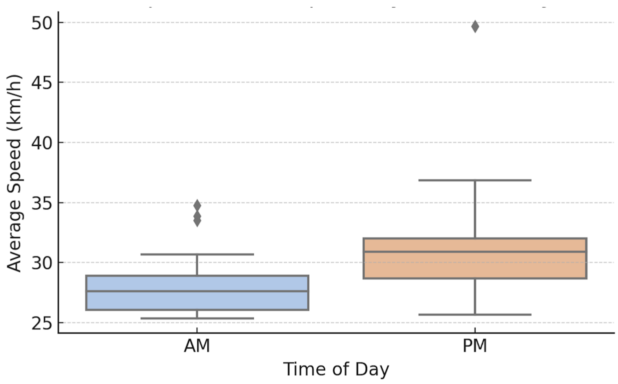

Fig. 6. Case study 1: Average speed by time of day (AM vs PM).

The afternoon rides were significantly faster and shorter in duration. The heart rate was also higher in the afternoon, although not significantly. These results suggest a possible physiological or environmental advantage later in the day [10].

The cadence data were sparse (available for only seven rides), with an average PM cadence of 95.96 rpm (SD = 4.07) and a single AM value of 104.85 rpm. Cadence was excluded from statistical testing, but is noted descriptively.

B. Case study 2: Weekday vs. Weekend performance

Research question 2 (RQ2): Does cycling performance on the selected segment differ between weekdays and weekends?

The rides were categorized by calendar day into **weekday** (Monday–Friday, $n = 38$) and **weekend** (Saturday–Sunday, $n = 14$). Detailed results comparing weekday and weekend sessions are presented in Table V, with average speed distributions shown in Figure 7.

TABLE V
PERFORMANCE BY DAY TYPE.

Metric	Weekday (n=38)	Weekend (n=14)	*p*-value
Avg. Speed (km/h)	30.58 ± 4.40	28.87 ± 3.23	0.135
Avg. Heart Rate (bpm)	140.36 ± 11.86	135.85 ± 11.24	0.218
Duration (s)	120.39 ± 18.34	126.00 ± 14.12	0.253

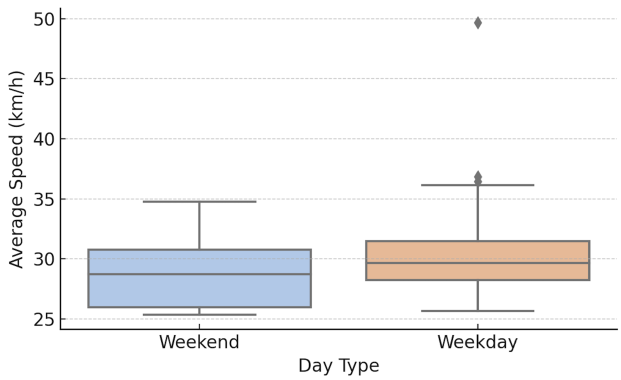

Fig. 7. Case study 2: Average speed by day type (weekday vs weekend).

While no metrics reached a statistical significance, a trend was observed toward higher performance on weekdays.

C. Case study 3: Performance progression over time

Research question 3 (RQ3): Does the cyclist's performance on the segment improve, decline, or remain stable over multiple rides?

All 52 rides were ordered chronologically. Linear regression was applied, to evaluate the trends in average speed, heart rate, and duration across sessions. The regression-based trend analysis is shown in Table VI, and Figure 8 provides a visual representation of speed progression across sessions.

TABLE VI
PERFORMANCE TRENDS OVER TIME

Metric	Slope (per ride)	*p*-value
Avg. Speed (km/h)	−0.005	0.891
Avg. Heart Rate (bpm)	+0.051	0.641
Duration (s)	+0.077	0.636

No significant trends were observed. This suggests stable performance over time, with no measurable training-related improvement on the selected segment.

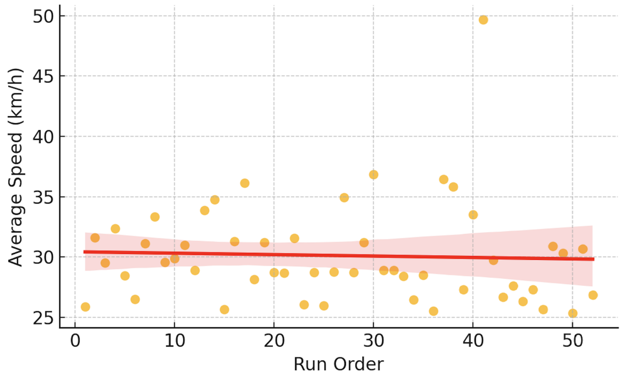

Fig. 8. Case study 3: Speed trend over time.

D. Discussion

RQ1 confirmed significant temporal differences in performance, with rides in the afternoon being faster and shorter in duration compared to sessions in the morning, which aligns with previous studies suggesting that athletic performance can improve later in the day due to circadian rhythms [10].

In contrast, **RQ2** and **RQ3** showed no statistically significant performance differences based on day type (weekday vs. weekend) or over time. These findings suggest that short segment performance is not strongly affected by scheduling patterns or accumulated training effects, at least for the selected segment and this particular athlete.

One possible explanation is that the segment's characteristics, moderate length, flat terrain, and consistent location early in the ride minimize the influence of fatigue, motivation, or route variation, which makes time of day a more dominant factor in performance variability than the day of the week or longitudinal progression.

Although the results offer useful information, the study has some limitations. Focuses on a single cyclist, which limits generalizability. Moreover, the analysis does not control for external contextual factors such as weather, wind, or traffic, which could influence performance metrics. Incorporating such data in future work may help disentangle physiological effects from environmental conditions.

V. Conclusion

This paper presented an improved approach for detecting and analyzing overlapping cycling segments from GPS-based TCX data, leveraging the updated `TCX2Graph.jl` library. The method combines candidate generation, KD-tree spatial filtering, and discrete Fréchet distance validation, and data fusion enriches the segment-level data with contextual metadata from OpenStreetMap, resulting in higher detection accuracy and interoperability.

The presented method enables practitioners to better understand how time-related factors influence performance, supporting more tailored training and analysis.

A notable strength is the consistent relative position of the selected segment across rides, minimizing the impact of fatigue and enabling fair temporal comparison.

Future research will address additional contextual factors, such as weather and traffic, and validate the methodology on larger and more diverse datasets to support broader applications in sports analytics.

Acknowledgment

The authors acknowledge the financial support from the Slovenian Research Agency (Research Core Funding No. P2-0057).

References

[1] K. Iqbal, "TCX File Format- Training Center XML File — docs.fileformat.com," https://docs.fileformat.com/gis/tcx/, [Accessed 23-05-2025].
[2] "Garmin developer program," https://developer.garmin.com/gc-developer-program, [Accessed 23-05-2025].
[3] P. A. Zandbergen and S. J. Barbeau, "Positional accuracy of assisted gps data from high-sensitivity gps-enabled mobile phones," *The Journal of Navigation*, vol. 64, no. 3, pp. 381–399, 2011.
[4] F. De Cock, N. Dardenne, F. Jockin, and B. Jidovtseff, "Validity and reliability of strava segments: Influence of running distance and velocity," *Journal of Human Sport and Exercise*, vol. 18, pp. S41–S56, 2023.
[5] T. Hliš, I. Fister, and I. Fister Jr., "Digital twins in sport: Concepts, taxonomies, challenges and practical potentials," *Expert Syst. Appl.*, vol. 258, no. C, Dec. 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2024.125104
[6] T. Hliš and I. Fister, "Tcx2graph.jl," https://juliahub.com/ui/Packages/General/TCX2Graph, 2024, [Accessed 23-05-2025].
[7] T. Hliš, I. Fister, and I. Fister, "Path overlap detection and property graph construction from tcx data for advanced analysis," in *2025 IEEE 23rd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2025, pp. 000 489–000 494.
[8] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," https://www.openstreetmap.org, 2017.
[9] J. Bleiholder and F. Naumann, "Data fusion," *ACM computing surveys (CSUR)*, vol. 41, no. 1, pp. 1–41, 2009.
[10] G. Atkinson and T. Reilly, "Circadian variation in sports performance," *Sports medicine*, vol. 21, pp. 292–312, 1996.
[11] I. Fister, S. Rauter, D. Fister, and I. Fister, *A collection of sport activity datasets with an emphasis on powermeter data*. University of Ljubljana, 2017.
[12] R. Olbricht, "Overpass api documentation," https://wiki.openstreetmap.org/wiki/Overpass_API, 2024, [Accessed 23-05-2025].
[13] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive computing*, vol. 7, no. 4, pp. 12–18, 2008.
[14] K. Zhou, Q. Hou, R. Wang, and B. Guo, "Real-time kd-tree construction on graphics hardware," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 5, pp. 1–11, 2008.
[15] Y. Chen, L. Zhou, Y. Tang, J. P. Singh, N. Bouguila, C. Wang, H. Wang, and J. Du, "Fast neighbor search by using revised kd tree," *Information Sciences*, vol. 472, pp. 145–162, 2019.
[16] P. Ram and K. Sinha, "Revisiting kd-tree for nearest neighbor search," in *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining*, 2019, pp. 1378–1388.
[17] T. Eiter and H. Mannila, "Computing discrete fréchet distance," Christian Doppler Laboratory for Expert Systems, Tech. Rep., 1994, technical Report CD-TR 94/64.
[18] L. Passfield and J. H. Doust, "Changes in cycling efficiency and performance after endurance exercise," *Medicine & Science in Sports & Exercise*, vol. 32, no. 11, pp. 1935–1941, 2000.
[19] S. Vogt, Y. Schumacher, K. Roecker, H.-H. Dickhuth, U. Schoberer, A. Schmid, and L. Heinrich, "Power output during the tour de france," *International journal of sports medicine*, vol. 28, no. 09, pp. 756–761, 2007.
[20] S. Padilla, I. Mujika, J. Orbananos, and F. Angulo, "Exercise intensity during competition time trials in professional road cycling," *Medicine and science in sports and exercise*, vol. 32, no. 4, pp. 850–856, 2000.
[21] G. D. Ruxton, "The unequal variance t-test is an underused alternative to student's t-test and the mann–whitney u test," *Behavioral Ecology*, vol. 17, no. 4, pp. 688–690, 2006.