embracing a better life

GPS driven lap comparison

Ir. Jelle De Bock
prof. Steven Verstockt

imec IDLab UNIVERSITEIT GENT Universiteit Antwerpen

Problem description

- Cyclocross races are held on **technical tracks** under **varying weather circumstances**
- Most riders wear **GPS watches to record their races and/or training rides**
- The GPS files will contain **laps of the same circuit**
- GPS data might provide **interesting timing and sensor data insights**
  - On a per lap level
  - On sectors within a lap level
- **BUT:**
  - GPS data is known to be rather messy when external circumstances are not that great (forest, bad weather, ...)
  - A naive approach to process the GPS files based on distance / point matching might not be sufficient for these reason

PRINCIPLE OF LAP SEGMENTATION AND COMPARISON

Athletes' Workout files

GPS file processing + pre-processing

Detected laps in workout files

Hand drawn GPS base course

LAP & SEGMENT extraction / matching

N sectors of length *l* in each lap and course

Match sector start and end with sectors in hand-drawn course

Derive sector elapsed time of uploaded laps of our athletes

Sector number	Elapsed time *lap i* of *athlete j* (s)
1	3
2	5
3	10
4	5
...	.

COMPARE and ANALYZE

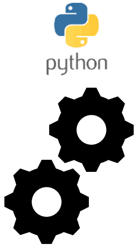
imec IDLab UNIVERSITEIT GENT Universiteit Antwerpen

GPS data extraction & pre-processing

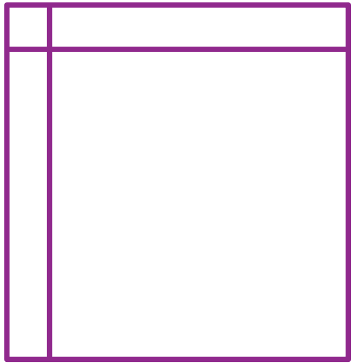
Extraction of GPS data

Schematic overview

GPS recording device produces a workout activity file



Conversion
.fit → timestamped csv file



Tabular GPS data

Extraction of GPS data

Conversion result

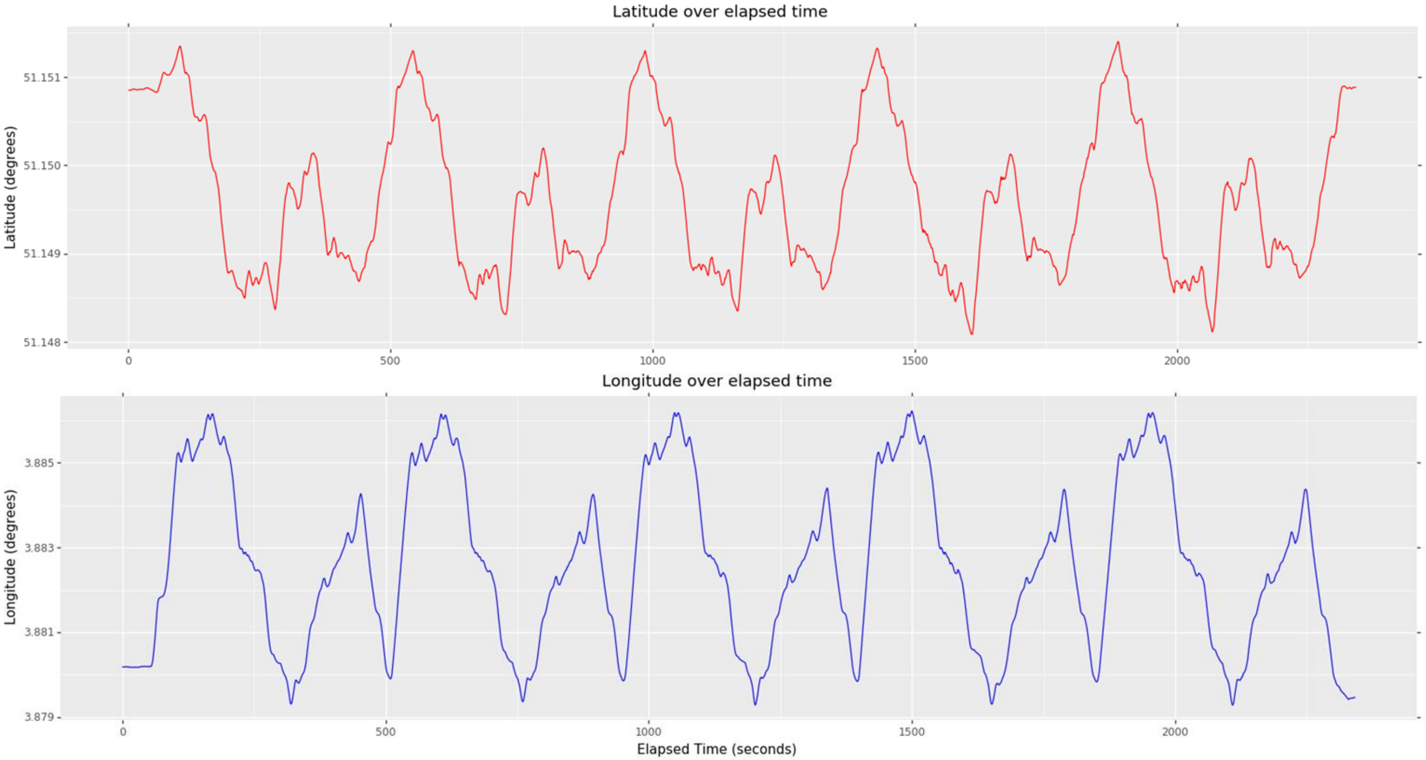
For lap extraction and the later matching process we create a Python Pandas Dataframe:

	latitude	longitude	Datetime	Elapsed_time
2019-11-23 14:00:12	51.150850	3.880187	2019-11-23 14:00:12	0.0
2019-11-23 14:00:13	51.150850	3.880183	2019-11-23 14:00:13	1.0
2019-11-23 14:00:14	51.150850	3.880180	2019-11-23 14:00:14	2.0
2019-11-23 14:00:15	51.150853	3.880182	2019-11-23 14:00:15	3.0
2019-11-23 14:00:16	51.150853	3.880183	2019-11-23 14:00:16	4.0
2019-11-23 14:00:17	51.150853	3.880183	2019-11-23 14:00:17	5.0
2019-11-23 14:00:18	51.150853	3.880187	2019-11-23 14:00:18	6.0
2019-11-23 14:00:19	51.150858	3.880188	2019-11-23 14:00:19	7.0
2019-11-23 14:00:20	51.150861	3.880188	2019-11-23 14:00:20	8.0
2019-11-23 14:00:21	51.150865	3.880185	2019-11-23 14:00:21	9.0

LAP / SECTOR EXTRACTION

Lap extraction from workout files

Investigating the repetitive patterns in GPS traces (i.e. laps in a GPS file)



Latitude
Longitude
Timestamp

3D

Latitude
Timestamp

2D

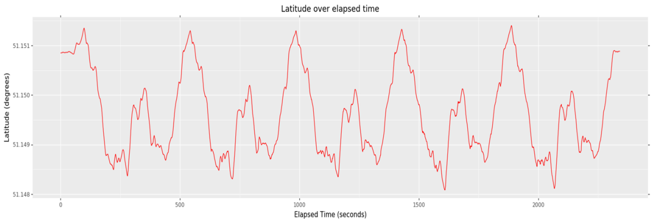
OR

Longitude
Timestamp

2D

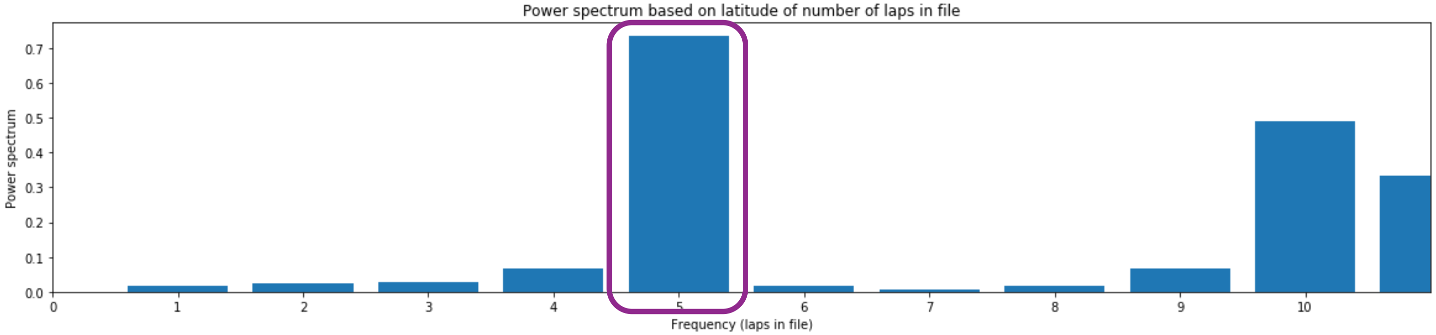
Fourier frequency analysis on GPS data

Fourier analysis on the 2D representation (latitude or longitude over time) of the GPS track



Discrete Fourier Analysis (DFT) is a technique to detect the harmonic content of synthesized (sound) waves

number of laps: *the frequency with the highest power within the spectrum*



Finding lap split points

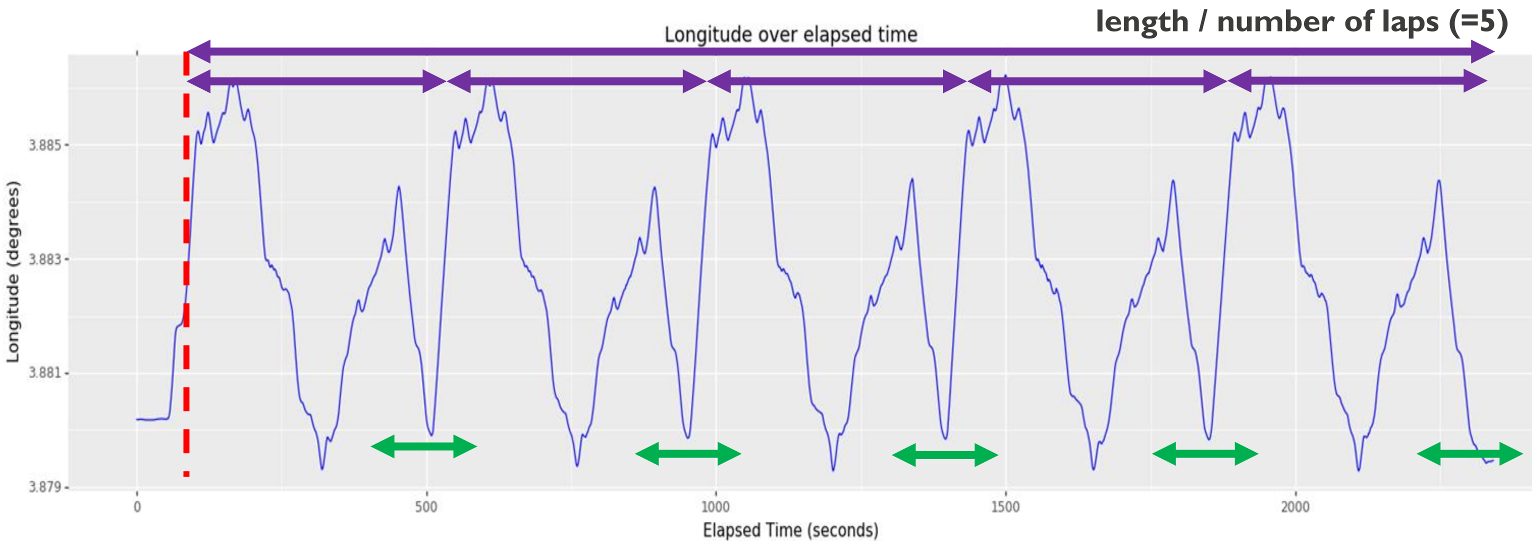
Find the lap intervals based on the reported total completed laps (Fourier analysis results)

Set search start

I. Manual (our approach)
II. Based on speed threshold

Define rough boundaries of the laps

Refinement of rough boundaries by neighbourhood search



Align the sectors within detected laps with a base course

Using geospatial distance driven coordinate matching heuristic

Principle

- Hand drawn course (very accurate) (=base course)
- Workout file of athletes (can be very inaccurate based on external circumstances)

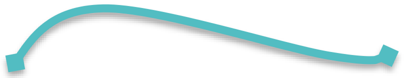
Step 1

Find the start candidate point in the workout file with a similar elapsed distance as in the hand-drawn course.

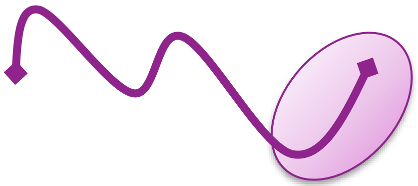
Step 2:

Search the optimal matching point based in a buffer around the start candidate point;

- Elapsed distance in both the lap and the course
- Haversine distance between lap on point and lap on the course



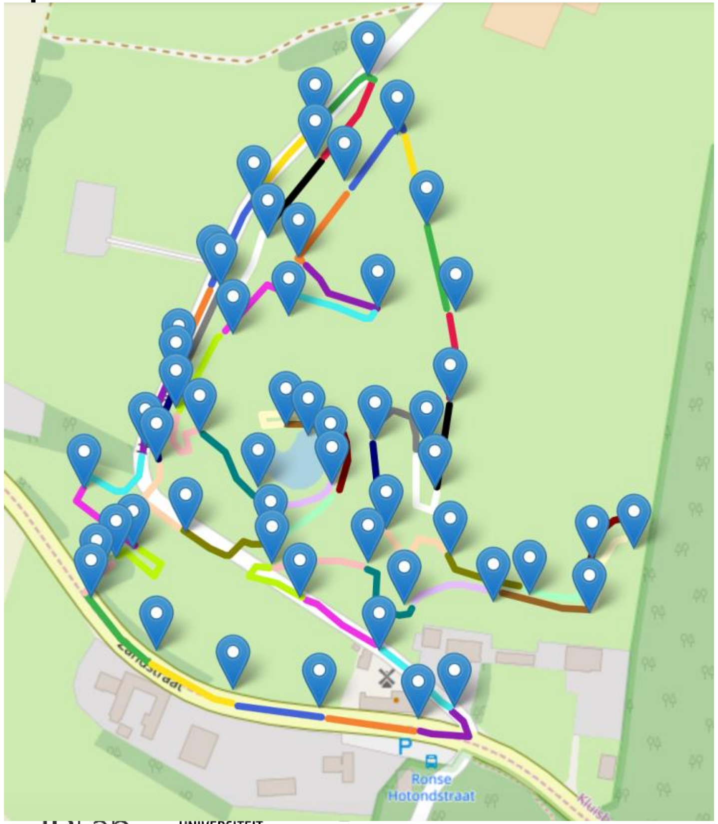
Course sector



Workout file sector

Convert *aligned laps* into *timetables*

Principle



Procedure

- Divide the lap in sectors of a certain length *l*
- Calculate the time spent in each sector for each lap

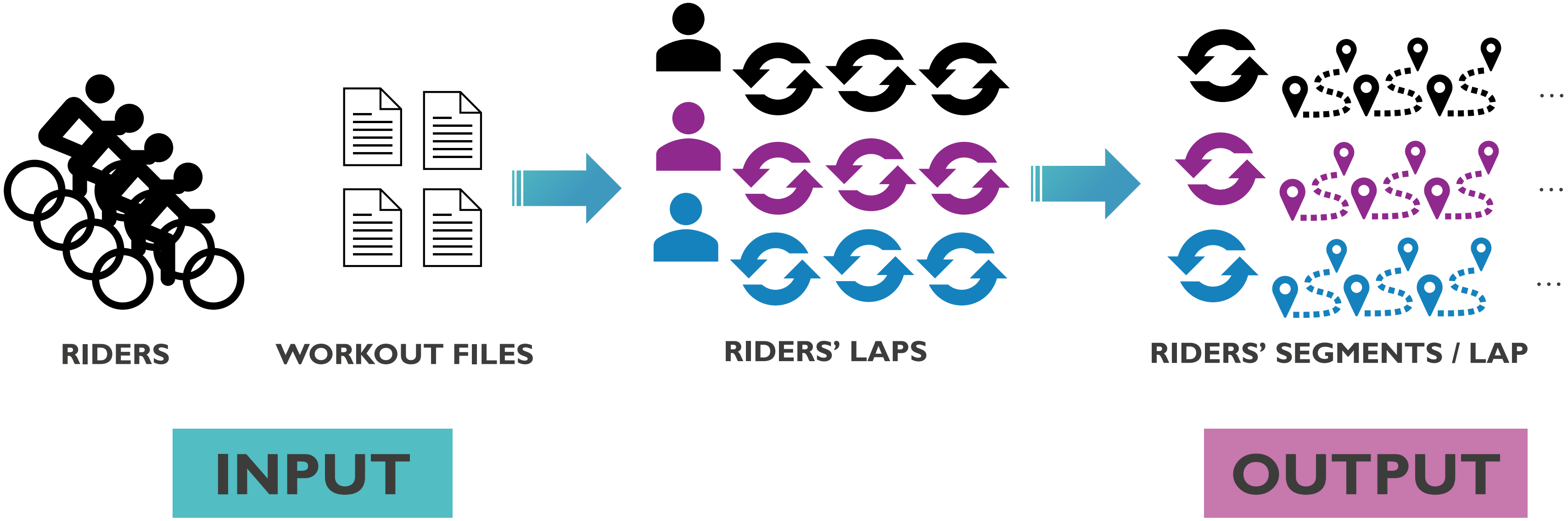
Resulting in

Differently coloured line segments are the sectors of a course with ***length l***

The **markers on the image are the matched sector split points** using the alignment method of the previous slide.

SECTOR TIME ANALYSIS

Output data of lap and sector segmentation



Sector engagement score

Principle

- Per sector we have data of multiple riders who completed the sector during multiple laps
- It is also known that due to various circumstances the segment matching in the workout files will not always be perfect (i.e., outliers might exist)
- The sector engagement of sector *i* is represented by the following formula

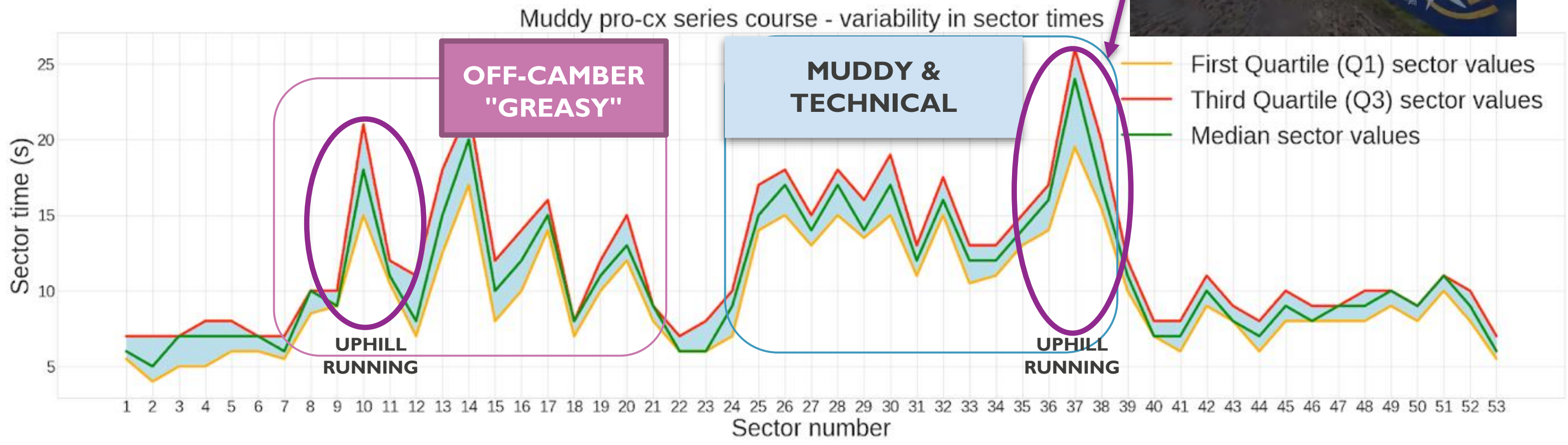
$$score_i = ln\left(\frac{med_i}{max-min}\right) \bullet med_i + \left(1 - \frac{med_i}{max-min}\right) \bullet \overline{med}$$

$$\begin{aligned} with: \quad & med_i \leftarrow median\ of\ the\ athletes'\ sector\ times\ for\ sector\ i \\ & max \leftarrow maximum\ of\ all\ the\ median\ values\ of\ the\ sector\ times \\ & min \leftarrow maximum\ of\ all\ the\ median\ values\ of\ the\ sector\ times \\ & \overline{med} \leftarrow average\ of\ all\ the\ median\ values\ of\ the\ sector\ times \end{aligned}$$

- The scores for the different sectors of a lap are scaled from 0 to 10 in the following case study example

Sector engagement score

Case study – DVV Trophy Ronse 2019



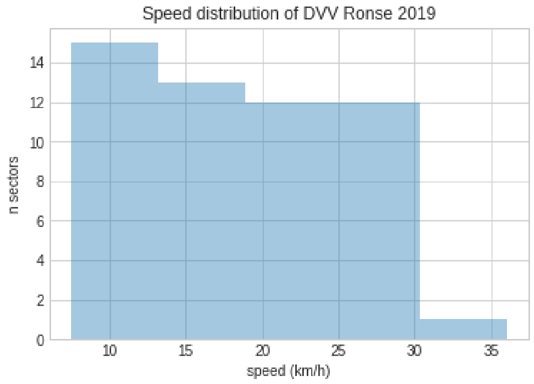
Observation for this race

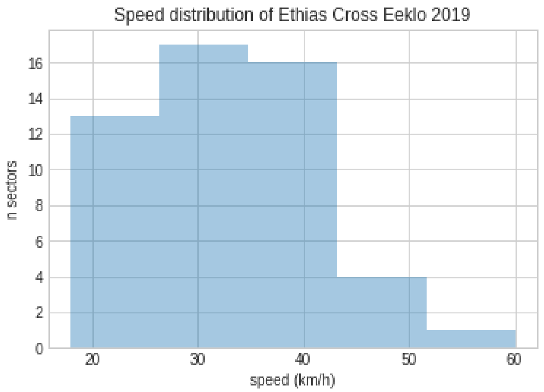
- Tougher sectors seem to have larger IQR (interquartile distance)
- Peaks are running sectors or off-camber difficult sectors
- Paved start- and finish-area are faster sectors, but also less variation in sector times

Course speed distribution

Principle and example

- We have for each course sector the athletes sector times and its (fixed) sector length
- With the mean sector time and its length, we can get an overall idea of the ***speed distribution*** of the course
- This can be very usefull to ***help characterize the type of track and match it with tracks that have a similar distribution***

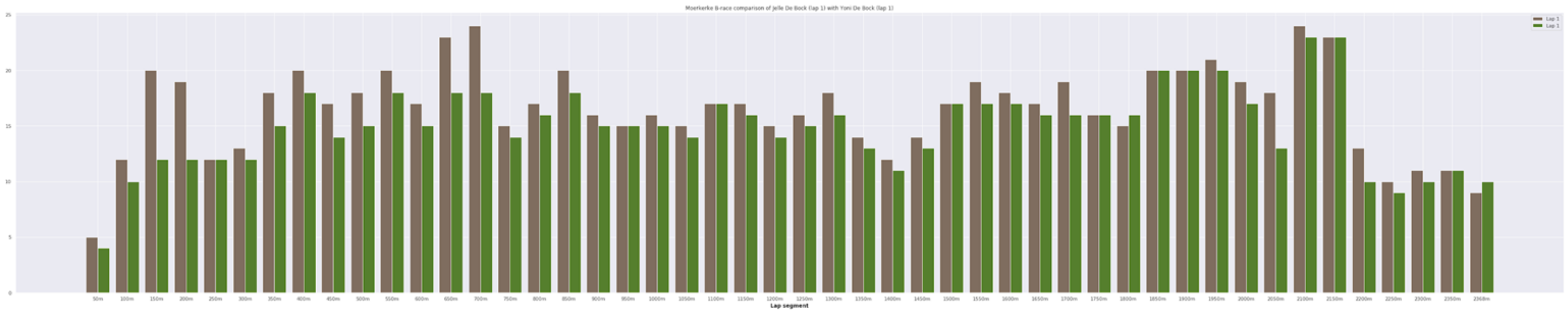




Head to head comparison

Case study

Smaller elite series cyclocross race - comparison between GPS files Jelle (34th) & Yoni (16th) De Bock



Conclusion:

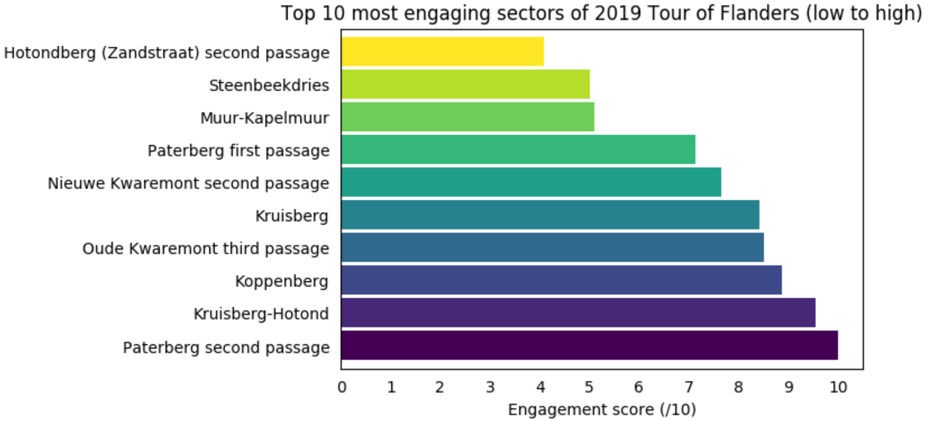
- Significantly slower near beginning (bad start position) and end (stretch of gravelish-road to power on) of lap
- Also slower on the parts with slippery, grassy turns followed by straights which require powerful accelerations

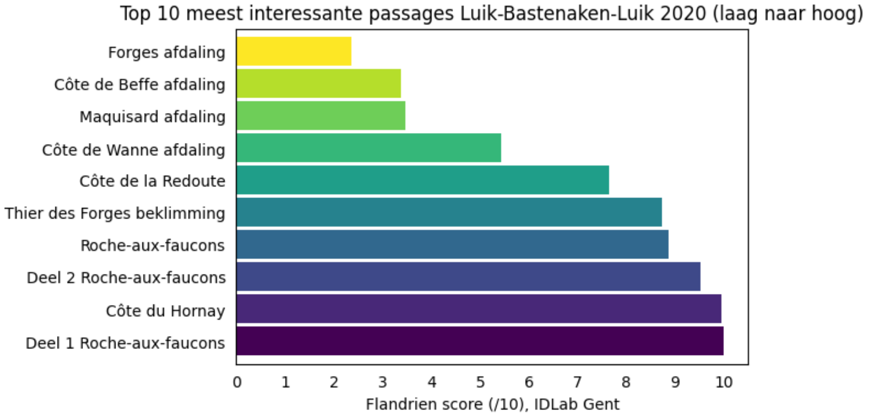
[EXTRA] SECTOR ENGAGEMENT IN ROAD RACES

Which were the most engaging segments of road races

Principle and results

- Rather than circular lap GPS files, we search for interesting segments within a larger trajectory (e.g., cobbled climbs, descents or technical zone)
- **Principle:** extract the timings of these segments from the riders' GPS files (c.f.r., Strava KQOM segments) and calculate the engagement score of the different segments of a race





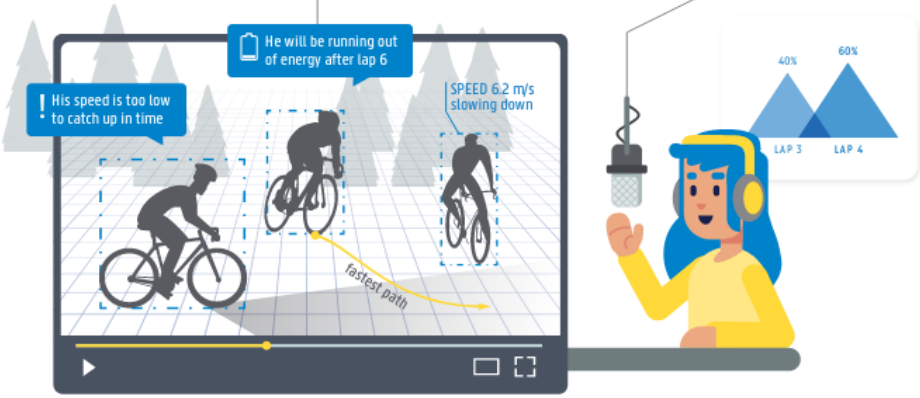
imec IDLab UNIVERSITEIT GENT Universiteit Antwerpen

DAIQUIRI, Data & Artificial Intelligence for QUantified Reporting In sports, profoundly enhances the way in which editors, directors and content creators conduct their job.

DAIQUIRI is an imec.icon research project funded by imec and VLAIO

AI algorithms will avoid data overload, enable sensor-video matching, provide dynamic captioning and create multi-modal stories.

The outcome will be **a scalable data workflow** supporting media channels in translating sports sensor data into engaging, real-time messages and visualizations.



vrt
UNIVERSITEIT GENT
imec
Universiteit Antwerpen
arinti
IDLab INTERNET & DATA LAB
In The Race
NEP
AGENTSCHAP INNOVEREN & ONDERNEMEN Vlaanderen is ondernemen
VIDEOHOUSE media facilities

IMEC-ICON DAIQUIRI PROJECT

imec & Vlaams agentschap innoveren en ondernemen

UNIVERSITEIT GENT

Thank you for your attention

Q&A

Jelle De Bock
Researcher at IDLab
(IMEC - Ghent University)
jelle.debock@ugent.be
Jelle De Bock
@jelledebock

Steven Verstockt
Professor and Researcher at IDLab
(IMEC - Ghent University)
steven.verstockt@ugent.be
Steven Verstockt