# Evolution et analyse des GP duathlons sur la période 2018-2023

Une image contenant arbre, personne, équipement sportif, plein air

Description générée automatiquement

[Pierre Violeau](http://www.linkedin.com/in/violeaupierre)

## Introduction

Le Duathlon est l’enchainement d’une portion de course à pied avec une portion de vélo avant de terminer par une nouvelle portion de course à pied. Le format le plus courant sur les Grand Prix de duathlons est le suivant :

5km course à pied / 20km vélo / 2.5km course à pied

Le Championnat de France des clubs de Duathlon se compose de 3 divisions, que ce soit pour le championnat féminin ou masculin. Les équipes des divisions 1 et 2 s’affrontent sur un championnat composé de 5 étapes en D1 et 4 étapes en D2. La D3 se déroule sur un format ¼ final, ½ finale et finale ou seules les meilleures équipes de chaque manche se qualifie pour la suivante. A la fin de la saison les 3 premières équipes de la finale D3 accèdent à la D2, les 3 premières équipes de la D2 accèdent à la D1 et enfin les 3 dernières équipes en D1 et D2 sont relégués dans la division inférieure.

Chaque club peut engager jusqu’à 5 athlètes par étape. Le classement est réalisé au cumul des places des 3 premiers athlètes de chaque équipe. A l’issue de chaque étape, un classement club est établi. A l’issue des 5 étapes, Les clubs féminins et masculins en tête du classement général de la D1 sont déclarés Champions de France.

Le Championnat de France et particulièrement la D1 est l'endroit où évolue les meilleurs duathlètes du monde, et il ne faut pas s'étonner que chaque année le podium des championnats d'Europe et des Mondiaux, soit garni d'athlètes issus de ses rangs.

Comme tout championnat, les grands prix de duathlon ont beaucoup évolué avec le temps, le but de cette étude est de déterminé quelles sont les évolutions sur la période 2018 - 2023 et les marqueurs de performance en fonction du championnat (D1, D2, Féminin, Masculin).

## Méthodologie

Les données des années 2023 et 2022 ont été récupéré par webscraping sur Prolivesport, pour les autres antérieur à 2022, les données proviennent directement de l’onglet résultat sur le site de la FFTRI. Les données venant du site de la FFTRI, on subit un traitement manuel pour homogénéiser le format, les colonnes et conserver uniquement les informations pertinentes.

La suite de la préparation des jeux de données a été réalisé dans un notebook en python. Pour chaque étape on retrouve un premier dataframe dit « scratch » qui regroupe le classement et les temps intermédiaire pour chaque course de l’étape (D1F, D1H, D2F, D2H) et un second dataframe dit « class\_etape » qui regroupe le classement par équipe et le temps des 3 athlètes classant chaque équipe. Plusieurs fonctions ont été développé spécialement pour l’analyse, tout d’abord les données de chaque étape sont nettoyées et enrichies avec l’ajout d’un delta temps entre le meilleur temps pour chaque partie de la course et chaque athlète. Le delta est disponible en (s) et en (s/km) pour comparer différent format de course. Un score est calculé pour chaque athlète, la formule du calcul est la suivante :

Score = Meilleur temps portion de course / Temps de l’athlète sur la portion \* 10000

L’étape suivante consiste à regrouper les différentes étapes d’une année au sein d’un même dataframe avant ensuite de regrouper toutes les années ensemble et de retirer les colonnes sans informations importantes ou alors celle n’étant disponible que pour certaine année. Les deux dataframes (« scratch » et « class\_etape ») sont ensuite sauvegardés au format csv pour pouvoir être utiliser ultérieurement lors de l’analyse et la visualisation des données.

Le notebook de la création des deux dataframes est disponible [ici](https://github.com/VioleauPierre/Grand_Prix_Duathlon_analysis/blob/main/GP_Duathlon_create_dataframe.ipynb)

Pour l’analyse des données, le choix a été fait de représenter les deltas de temps à l’aide de diagramme en boîte pour faciliter l’observation et une l’interprétation des données. Il y a donc une figure par partie de course et sur chaque figure un diagramme par année (ex :cf figure 1). Il y a aussi une figure pour représenter le delta de temps du 3ème athlète de chaque équipe par rapport au meilleur temps (ex : cf figure 3)

Grâce au score précédemment calculé, il est possible d’étudier la corrélation entre les différentes parties de la course entre elle et avec le score total. Cette corrélation est représentée via une matrice de corrélation (ex :cf figure 2), les valeurs sont obtenues par calcul du coefficient de corrélation de Pearson.

Enfin, la dernière visualisation proposée est la dynamique de course (ex : cf figure 4 et 5). Cette figure représente l’évolution des écarts entre tous les membres du top 30 final et le meilleur temps à plusieurs stades de la course : fin de cap 1, fin du vélo et à l’arrivé. Cela permet d’étudier comment chaque athlète du top 30 a évolué sur une course donnée. Il y a une figure par course. Ce rapport contiendra uniquement quelques exemples de dynamique de course, l’intégralité des graphiques est disponible dans le notebook [ici](https://github.com/VioleauPierre/Grand_Prix_Duathlon_analysis/blob/main/Data_Visualization.ipynb)

## Résultats

La figure 1 représente les écarts en s/km pour chaque année pour le top 30. Il y a un graphique par portion de course, plus un graphique sur la course entière.

* Cap 1 : 2019 est sans contestation possible l’année ou la cap 1 a été la plus resserré sur la période étudiée. On retrouve 15 athlètes à moins de 1,4 s/km du meilleur temps sur l’ensemble de la saison ! Le championnat 2020 perturbé par la COVID et condensé sur la fin de saison est celui pour lequel les écarts sont les plus important. 2023 marques un resserrement des écarts sur la cap 1 par rapport aux 3 années précédentes.
* Vélo : 2020 présente les plus grands écarts comme sur la cap 1. Les différences de préparations et d’investissement des athlètes lors de cette année particulière sont probablement la cause de ces écarts. 2021 est aussi une saison bousculée par la COVID mais à l’inverse de 2020 on retrouve des écarts assez faibles sur le vélo. On remarque aussi que les valeurs sont beaucoup moins dispersées autour de la médiane que sur la cap 1, les athlètes roulant bien souvent en pack, il est logique de retrouver plus d’athlète autour d’une même valeur d’écart en s/km.
* Cap 2 : 2022 marque un changement sur la cap 2. Les écarts sont plus importants au sein du top 30 à partir de cette année-là. La portion vélo ne créer certes pas plus d’écart sur ces années-là mais elle cause une fatigue importante aux organismes qui se répercutent ensuite sur la cap 2. La cause est probablement à chercher du côté des très bon cyclistes présents en D1H. En remontant vers l’avant de la course après une cap 1 dans le ventre mou, ils limitent les écarts à vélo et créer une dynamique de regroupement. En revanche par la même occasion, si plus d’athlètes arrivent ensemble à T2, beaucoup sont entamés par la portion vélo et l’écart entre les meilleurs duathlètes et les autres est donc exacerbé sur la cap 2.
* Total : On peut pratiquement reporter les tendances de la cap 2 sur le total, les années 2022 et 2023 présentent une plus forte dispersion de valeurs, il est intéressant de noter à quel point les écarts sont plutôt resserrés pour le top 15 (médiane) mais supérieur ensuite jusqu’au top 30.

Cette analyse permet uniquement de comparer la densité des performances par rapport au meilleur temps de chaque course pour chaque année. Elle ne permet pas de comparer les performances d’une année avec une autre, les performances étant dépendantes des parcours et conditions de course On ne peut pas affirmer à partir de ces boites à moustaches qu’en 2023 le niveau est plus élevé qu’en 2018.

### D1 Homme

Une image contenant texte, capture d’écran, diagramme, Caractère coloré

Description générée automatiquement

Une image contenant texte, capture d’écran, carré, Rectangle

Description générée automatiquementUne image contenant capture d’écran, texte, carré, Rectangle

Description générée automatiquement

Etudions maintenant les matrices de corrélations pour les années 2023 et 2019. (Les autres matrices pour la D1H sont disponibles directement dans le notebook [ici](https://github.com/VioleauPierre/Grand_Prix_Duathlon_analysis/blob/main/Data_Visualization.ipynb).

L’année 2019 est représentative de la tendance avant 2020 et 2023 de la tendance après 2020. Pour les deux cas on remarque que le score de la partie vélo est celui le plus corrélé avec le score Total. En revanche là ou la score de la cap 1 est aussi fortement corrélé avec le score total pour 2019, il n’est que moyennement corrélé pour 2023.

Une image contenant texte, diagramme, Rectangle, capture d’écran

Description générée automatiquementEn 2019, les meilleurs coureurs sur la cap 1 était fortement susceptible d’être à l’avant à l’arrivé de la course. En revanche en 2023, ce n’est plus le cas, et une bonne position à la fin de la cap 1 ne mets pas à l’abri d’un retour sur la portion vélo et la cap 2. Surtout lorsque l’on se penche sur la corrélation entre la Cap 1 et la Cap 2, celle-ci n’est vraiment pas significative en 2023 alors qu’en 2019 la corrélation est assez forte. En 2023, les meilleurs coureurs de la cap 1 ne sont pas forcément ceux de la cap 2 et inversement. Cela renforce l’idée d’une portion vélo désormais beaucoup plus sélective qu’en 2019 qui fatigue les organismes des duathlètes, ne permettant pas à ceux un peu limite sur la portion vélo de courir une cap 2 au niveau de leur performance sur la cap 1. La performance sur la portion vélo ne semble pas être lié à celle sur les cap 1 et/ou 2.

Un autre indicateur intéressant est le delta de temps entre le vainqueur et le 3ème de chaque équipe. On remarque rapidement la différence entre 2022/2023 et les autres années. C’est la conséquence d’une portion vélo plus disputé à l’avant de la course. Les écarts avec les athlètes ne figurant pas dans le/les premiers packs à vélo sont dès lors beaucoup plus important à l’arrivé.

Pour analyser chaque course en détail, il est possible de représenter la dynamique de course sur un graphique, par soucis de clarté, uniquement le top 30 est représenté (les dynamiques de course pour chaque année et chaque épreuve en format « classique » ou M sont à retrouver directement sur le notebook [ici](https://github.com/VioleauPierre/Grand_Prix_Duathlon_analysis/blob/main/Data_Visualization.ipynb).

Prenons par exemple les courses de Parthenay 2022 et 2023 qui se sont déroulées sur un parcours quasiment identique avec deux configurations de course différentes pour jouer le top 30 :

Une image contenant texte, capture d’écran, ligne, diagramme

Description générée automatiquement

Dynamique de course du TOP 30 de la D1H sur l'épreuve de Parthenay en 2023

En 2023, après une cap 1 sélective avec déjà des écart entre plusieurs duathlètes en moins de 10 secondes et la majorité du reste à entre 20 secondes et 40 secondes. Sur la partie vélo on assistera à un regroupement et tous les écarts de la cap 1 sont effacés à l’arrivé à T2, seuls quelques duathlètes pointés à plus de 40 secondes après la cap 1 parviendront à revenir sur l’avant. La cap 2 va donc être décisive pour tous les duathlètes de ce peloton. On remarque que la cap 2 a créé plus d’écart dans le top 30 pour une distance de 2,5km que la cap 1 de 5km, ce qui appui l’idée que la portion vélo a créé beaucoup de fatigue malgré la présence d’un peloton fourni.

Une image contenant texte, diagramme, capture d’écran, ligne

Description générée automatiquement

Dynamique de course du TOP 30 de la D1H sur l'épreuve de Parthenay en 2022

En 2022, la cap 1 a été encore plus sélective qu’en 2023 et on ne retrouve pas de groupe entre 20 et 40 secondes mais plutôt un échelonnement des duathlètes entre 20 et 60 secondes. Cet échelonnement va entrainer la création de deux packs sur la portion vélo et les duathlètes n’ayant pas réussi à rejoindre le premier pack sont déjà assuré d’être en dehors du Top 15. Dans l’ensemble, la cap 2 a créé les mêmes écarts que la cap 1.

### 3.2 D1 Femme