

Comment mettre les données au service du sport de haut niveau

D'après la conférence d'Adrien Sedeaud
www.mediachimie.org

Adrien Sedeaud est chercheur et adjoint au directeur de l'IRMES (Institut de Recherche Médicale et d'Épidémiologie du Sport), pôle Performance/INSEP (Institut National du Sport, de l'Expertise et de la Performance).

Cette conférence a été réalisée à partir des données de l'ouvrage du même titre, paru aux éditions INSEP en 2024.

Introduction

Les données transforment notre quotidien. Ces évolutions impliquent la mobilisation de nouvelles ressources et la construction de nouveaux

savoirs et savoir-faire. Le monde de la performance sportive en est un exemple remarquable via l'émergence de technologies embarquées, de capteurs de toutes sortes, le développement de méthodes

d'analyses et des investissements financiers conséquents sur ce champ pouvant générer un avantage concurrentiel.

La conférence aborde différents exemples concrets réalisés par le pôle Performance de l'INSEP qui apportent des données dans le sport de haut niveau en se focalisant sur l'estimation de potentiels, le monitoring, l'estimation des risques de blessures ou encore l'analyse de la concurrence.



Figure 1

Image récupérée d'une vidéo d'un essai de rugby vu par un spectateur.



Figure 2

Image récupérée d'une vidéo d'un essai de rugby vue par un analyste des données.

Les **Figures 1** et **2** montrent un exemple du support scientifique qui peut être apporté aux staffs et aux athlètes des équipes de France à travers l'utilisation des données. La **Figure 1** est une image récupérée d'une vidéo où on voit un essai de l'équipe du rugby du Stade Rochelais qui joue en noir.

La **Figure 2** est celle récupérée à partir d'une vidéo réalisée par un analyste des données.

Dans le maillot de chaque joueur, entre les omoplates, il y a des capteurs GPS de 50 Hz, qui permettent d'obtenir 50 données par seconde sur les 3 axes. On mesure ainsi les accélérations et les décélérations. Le même type de données est capté lors des phases de préparation et aux entraînements. On dispose ainsi de beaucoup de données pour assurer un suivi de l'athlète et des équipes.

Les données sont omniprésentes, comme dans notre société, mais c'est peut-être un peu plus exacerbé dans le sport de haut niveau.

1 Utilisation des données dans le haut niveau

La **Figure 3** résume les principaux objectifs de l'utilisation de ces données par les équipes de l'INSEP.

Analyse de la concurrence

Les données obtenues à partir de la vidéo de la **Figure 2** permettent d'analyser les victoires et les défaites des adversaires des Français et de fournir des informations aux staffs des équipes de France. L'analyse qui leur est transmise permet

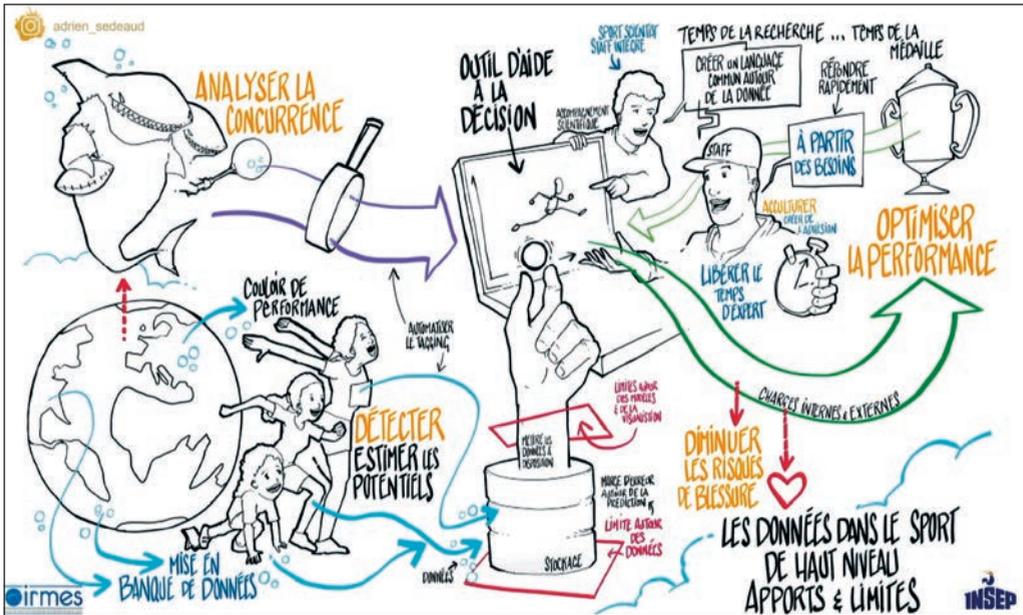


Figure 3

Schéma vulgarisant l'utilisation des données dans le sport.

d'accéder à l'ensemble des données que l'on peut capitaliser sur les adversaires, les façons dont ils perdent et gagnent, pour essayer de les mettre dans des schémas qui vont leur être défavorables, et qui vont être favorables aux athlètes français.

Détection des talents

L'IRMES crée des outils à partir des données pour aider les fédérations dans l'**estimation de potentiel**, afin de ne pas faire de la détection à un instant T du niveau de performance de jeunes athlètes qui sont assez peu prédictives d'une performance future. Sont ainsi créés des **indicateurs** en fonction de différents paramètres d'analyses de performances dépendant de l'âge biologique, de l'âge relatif, et de l'âge d'entraînement.

Optimisation des performances

Énormément d'informations sont capitalisées avec deux objectifs principaux : optimiser la préparation des athlètes pour gagner en performance et faire de l'estimation de risques de blessures pour éviter le seuil fatidique à très haut niveau, lorsque l'athlète qui atteint son optimum de performance est sur une ligne de crête où il est très proche de ce risque de blessure.

Outils d'aide à la décision

Ces données doivent pouvoir être utilisées pour répondre aux questions des entraîneurs et des athlètes. Ce sont des outils d'aide à la décision et il faut choisir en conséquence les variables à mesurer sur les capteurs, les outils pour les capter, leur restitution et la visualisation ou la modélisation

qui peut être fournie à partir de ces données et voir comment les utiliser pour améliorer la performance.

L'apport est concomitant des limites et le devoir déontologique implique d'expliquer tous les biais qui peuvent être incorporés. On ne capte qu'une partie des informations qui font la pluralité des performances. Plus le travail est fait en collaboration avec les staffs, plus les données sont des outils d'aide à la réflexion plus qu'à la décision, parce qu'elles fournissent souvent plus de questionnements.

1.1. Détections des talents : outils d'estimation de potentiel

Trois biais importants à prendre en compte sont encore aujourd'hui très peu considérés par l'ensemble des fédérations : l'âge relatif, l'âge biologique et l'âge d'entraînement (Figure 4).

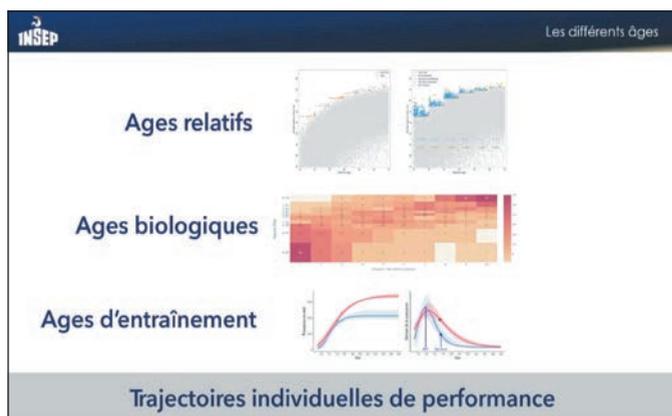


Figure 4

Présentation des 3 biais liés aux âges dans les trajectoires individuelles des performances des jeunes athlètes.

Âge relatif/âge biologique

Que ce soit au niveau départemental et régional, mais aussi – et surtout – au niveau des équipes de France jeunes, il y a une répartition non homogène des mois de naissance des enfants sélectionnés pratiquant une activité sportive, avec beaucoup plus d'individus nés en début d'année qu'en fin d'année, et ce dans tous les sports. C'est ce qu'on appelle l'« effet de l'âge relatif ».

Parce qu'ils sont relativement plus vieux en âge, ces enfants sont aussi biologiquement plus vieux. C'est le deuxième effet, celui de l'**âge biologique** : il peut ainsi y avoir un an et 364 jours de différence d'âge biologique entre deux enfants de la même année. Bien entendu, celui qui est plus vieux a des avantages considérables dans la pratique physique, mais aussi dans la coordination musculo-cérébrale. Cette maturation se fait à des timings individuels qu'il est facile aujourd'hui de mesurer et de fournir aux fédérations. Lorsque des tests ou des performances sont réalisés, les performances sont recalibrées en fonction de cet effet de l'âge relatif. Ainsi, on regarde quel est le niveau de performance pour l'individu qui a 12 ans et 1 jour, 12 ans et 2 jours, 12 ans et 3 jours, etc., et on recalibre la relation entre la performance et l'âge.

Ces premières étapes, qui ne sont pas chronophages pour les entraîneurs ou pour les staffs, puisqu'elles sont établies à partir d'une date de naissance et des performances, permettent de prendre conscience des biais de sélection. **L'athlète le plus**

âgé d'une catégorie sera alors plus performant qu'un athlète de la même catégorie, mais plus jeune.

Cela a permis notamment à la Fédération Française de Natation de réfléchir sur les individus qu'ils invitaient aux championnats jeunes, et donc de faire venir des jeunes athlètes qui n'ont pas, aujourd'hui, le niveau de performance en recalibrant leur performance. Dans la catégorie des 12-14 ans, la différence entre l'âge relatif et l'âge biologique sur le niveau de performance est très importante, pouvant atteindre deux années.

De même, la puberté agit et perturbe la performance des jeunes athlètes.

Au moment du pic pubertaire croisé entre garçons et filles, les individus qui ont biologiquement passé leur pic pubertaire sont 100 % prêts pour exprimer leur plein potentiel alors que ceux qui sont avant le pic pubertaire sont prêts à 10, 20 ou 30 %. Donc cela fait concourir entre eux des individus qui n'ont absolument pas les mêmes capacités.

Il a été observé sur les quatre derniers championnats de France des catégories juniors que les individus les plus performants sont ceux qui ont passé le pic pubertaire depuis 4 ans et demi et que les individus les moins performants sont ceux en plein pic pubertaire.

La littérature scientifique sur ce sujet est abondante mais malheureusement encore peu utilisée aujourd'hui. Concrètement, sur le terrain, ces effets se retrouvent à tous les niveaux, et les

performances dans les catégories jeunes même juniors ne présagent pas d'une réussite future car ce sont des populations disparates, et les œillères sont posées sur les individus qui sont à un instant t les plus performants et pas sur ceux qui ont le potentiel de devenir les plus performants.

L'INSEP crée des équations de calibration pour que, en fonction de l'effet de l'âge biologique, on recalibre les performances existantes pour chaque discipline, sport ou épreuve dans les catégories U9, U10, U11, U12.

Âge d'entraînement

Si deux jeunes athlètes ont le même niveau de performance mais que l'un pratique l'athlétisme depuis un an et l'autre depuis six ans, les capacités de progrès ne sont pas les mêmes, mais cela aussi est encore très peu considéré. L'INSEP modélise donc la relation, pour chaque discipline, entre la performance et l'âge d'entraînement, ce qui permet de recalibrer.

1.2. Évaluation et optimisation des performances

La Fédération Française de Ski (FFS) a étudié l'évolution des performances des skieurs professionnels (selon des critères définis par la FFS) en fonction de leur âge tout au long de leur carrière. Un grand nombre de carrières ont ainsi été étudiées. À travers toutes ces trajectoires individuelles de performance, on observe une progression au fil des années, suivie d'un pic puis d'un plateau.

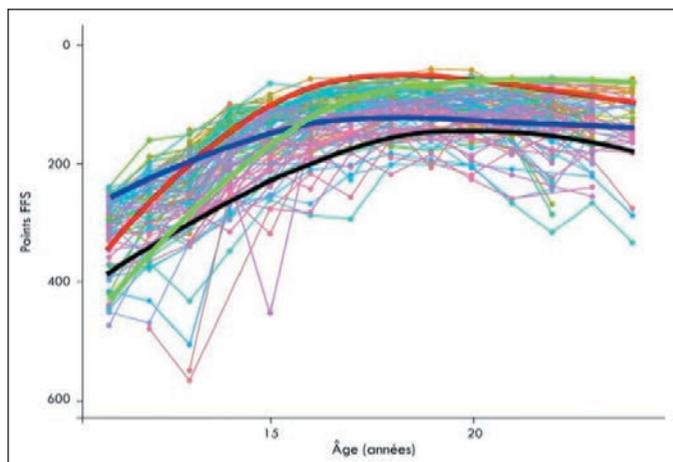


Figure 5

Les 4 trajectoires de performance les plus observées chez les sportifs tout au long de leurs carrières.

Meilleures performances annuelles (Points FFS) en fonction de l'âge des skieurs alpins français depuis la saison 2004-2005. Les points FFS correspondent à une performance relative dépendant du niveau global de la course et de l'écart au vainqueur de la course : plus les points sont faibles, plus la performance est élevée.

Que ce soit pour ces données dans le domaine du ski ou pour des millions de données en natation ou en athlétisme, on retrouve toujours quatre trajectoires lorsqu'on utilise des méthodes de classification (**Figure 5**).

Les individus avec la trajectoire noire ont des niveaux de performance assez faibles tout au long de leur vie. Ils sont bien détectés et continuent en pratiquant du sport santé. Les individus en bleu ont des niveaux de performance élevés dans les catégories jeunes ; ils sont très rapidement détectés et on leur propose en général d'intégrer, selon le sport, des centres de formation, des pôles France, et ainsi de suite, mais on s'aperçoit qu'en fait leurs niveaux de performance va stagner par la suite. Les individus en rouge ont des

niveaux de performance assez élevés quand ils sont jeunes ; ils sont donc repérés, puis ils continuent à progresser mais n'atteignent pas le niveau de champion olympique ou du monde.

Les athlètes sur lesquels il faut porter le plus d'attention sont ceux des courbes vertes. Ils ont des niveaux de performance faibles dans les catégories jeunes, mais s'ils poursuivent, ce sont ceux qui atteindront plus tard des niveaux dignes des champions du monde ou olympiques.

Il faut donc créer des outils pour tenter d'estimer ces trajectoires individuelles et construire des parcours de vie ou parcours de performance.

Sur la base des données de la FFS, l'INSEP a réalisé une modélisation mathématique de la trajectoire des skieurs pour estimer la trajectoire qu'un athlète pourrait suivre au cours de sa carrière à partir de quelques points de performance pour répondre à la demande des entraîneurs qui est toujours la même : **entre ces 50 athlètes qui ont le même niveau à 12 ans, à partir de 3 à 4 performances historiques, lequel peut potentiellement être le plus fort ?**

Cela permet aux sélectionneurs et aux entraîneurs d'avoir, dans un premier temps, des informations sur l'historique des performances des jeunes athlètes, puis d'étudier selon les différents athlètes sélectionnés les possibilités de trajectoire de chacun d'eux. Le but n'est pas de prédire une performance mais d'estimer l'ensemble des possibles, notamment pour des individus avec des niveaux de

performance plutôt faibles au début de leur carrière, mais qui pourraient atteindre des niveaux de champion du monde ou olympique (courbe verte de la **Figure 5**). Ce sont donc des pièces de puzzle qui constituent de nouveaux outils pour les entraîneurs et leur permettent de confirmer leurs intuitions.

La **Figure 6** est un exemple fictif du parcours de vie d'un jeune athlète.

Son niveau de performance est reporté en fonction de l'âge. Il débute à 3 ou 4 ans (par exemple au tennis), arrête, puis reprend et intègre une structure jeune de haut niveau. Il rentre ensuite dans un pôle France ou une académie, autour de son pic de vélocité de croissance, donc de son pic pubertaire. Il peut alors se blesser, mais reprendre et devenir professionnel, et atteindre un âge de pic de performance qui sera estimé. En plus des éléments vus précédemment, les études d'un grand nombre de cas ont permis d'estimer les facteurs qui ont le plus de poids en fonction des âges (**Figure 7**).

Pendant la jeunesse de l'athlète : le lieu de naissance et de vie

Très jeune, par exemple à 3 ans, lorsque l'athlète commence à jouer au tennis, la pratique délibérée, c'est-à-dire le jeu, est peut-être du point de vue de la coordination, ce qui permet le meilleur développement pour cette tranche d'âge. Vient ensuite le lieu d'habitation ou de résidence. De nombreuses études ont montré qu'il valait mieux être rural, ou en tout cas habiter dans des villes inférieures

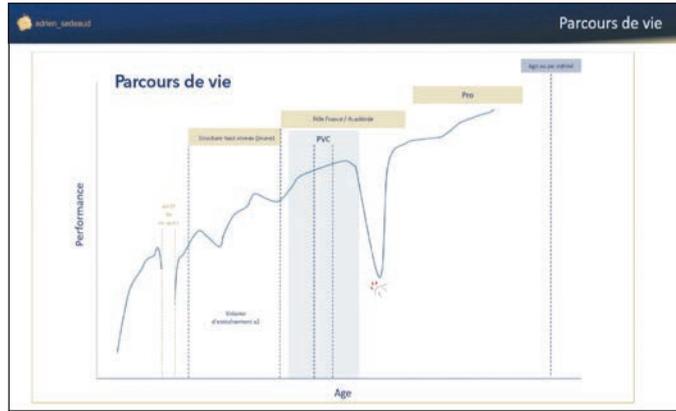


Figure 6

Parcours de vie possible pour un jeune athlète.

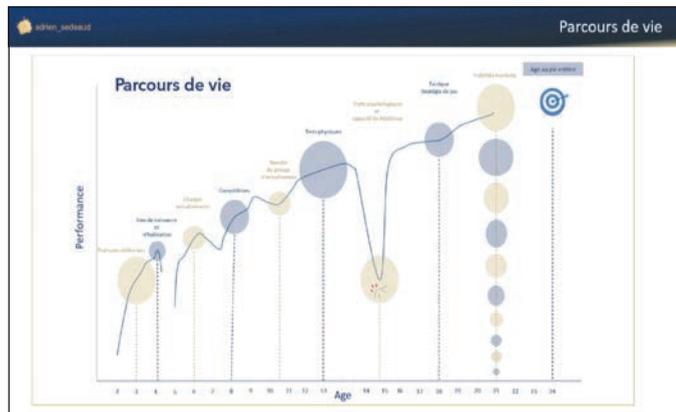


Figure 7

Parcours de vie possible pour un jeune athlète, en considérant les facteurs prépondérants pour sa carrière.

à 100 000 habitants, pour devenir champion olympique. Les explications possibles à ce constat sont que le jeune athlète a de nombreuses opportunités de se déplacer s'il habite à la campagne, où il ira à l'école à pied ou à vélo. De plus, s'il prend une licence dans le club de son village, il se verra proposer une multi-licence : il fera du tennis, mais sa licence

de tennis lui permettra de faire aussi de l'athlétisme, du judo et toutes les activités qui sont proposées.

La place dans la fratrie est aussi un élément essentiel : ce sont souvent les deuxièmes et troisièmes enfants les plus performants car ils passent beaucoup de temps à courir après le ou les grands frères et sœurs.

Au niveau de l'entraînement : l'impact des structures sportives de haut niveau et l'entrée dans le circuit professionnel

Les analyses des données montrent l'importance dans cette courbe de progression individuelle, des charges d'entraînement, des charges de compétition ainsi que la densité des groupes d'entraînement. Lorsqu'il entre dans une structure de très haut niveau, les coéquipiers des entraînements quotidiens vont fortement impacter la capacité de progression du jeune athlète.

Autour de l'âge du pic pubertaire, les capacités et les tests physiques sont nettement améliorés par le climat hormonal.

Les traits psychologiques et la capacité mentale de résilience peuvent être pertinents en cas de blessure.

Lorsqu'il entre sur les circuits professionnels, l'aspect tactique et la vision stratégique du jeu seront importants.

Il faut identifier de manière individuelle les éléments qui ont le plus d'importance dans une réussite future sur des individus très jeunes.

Puisque la performance est multifactorielle et complexe, il

faut essayer de percevoir tous ces éléments qui ont un impact sur le jeune athlète à 21 ans. Ces impacts sont récurrents, individuels et ils évoluent avec le temps. C'est ce que l'INSEP essaye de mesurer grâce à des modèles de plus en plus complexes, à partir d'informations sur des groupes d'individus pour comprendre individuellement l'impact de chacune de ces variables au cours du temps.

Sont ainsi construits des parcours de vie modélisés, afin d'estimer des possibilités pour un grand nombre d'athlètes. Cette modélisation est poursuivie notamment autour de l'âge du pic de performance estimé.

Estimer le niveau de performance des athlètes et des équipes

Ces estimations font partie des missions de l'INSEP et permettent de créer des couloirs de performance pour visualiser les trajectoires de performance des athlètes français.

Ce type de programme est réalisé avec les équipes de Florian Rousseau, Directeur de la performance du cyclisme sur piste et ancien athlète multimédaille, 11 fois champion du monde et deux fois champion olympique.

Les meilleures performances des sauteurs à la perche, médaillés olympiques lors des 4 dernières olympiades, ont été reportées en fonction de l'âge (de 16 à 37 ans), chaque point étant la meilleure performance à 16 ans, 17 ans, 18 ans, 19 ans, et ainsi de suite pour tous ces athlètes. La synthèse de ces données a permis la

construction d'un couloir de performance, en vert sur la **Figure 8**. Pour les Jeux de Londres, cela a permis de visualiser où se situaient les athlètes français par rapport aux trajectoires communes des médaillés olympiques.

L'INSEP travaille aujourd'hui sur le programme « **Médaillabilité** », qui permet d'estimer la probabilité de l'ensemble des athlètes de la planète d'être champions olympiques demain. 30 millions de données sont reçues chaque nuit et mises à jour lors de toutes les compétitions internationales pour les catégories jeunes et les catégories seniors. Ces données sont exploitées par une équipe de 4 chercheurs qui analysent comment les athlètes français se situent par rapport à la concurrence internationale, afin de fournir des visualisations pour les décideurs et pour les staffs.

Ces programmes ne sont pas destinés à la prédiction mais à l'estimation.

Modélisation de scénarios de courses

Est ainsi établi un tableau de bord récapitulant les données recueillies sur un athlète et les prédictions de performance afin d'estimer ses chances de médaille.

Par exemple, la Fédération Française d'Athlétisme a financé des travaux de recherche pour construire le profilage des athlètes français avec une batterie de tests physiques et physiologiques à réaliser et toutes les informations sur les courses de l'ensemble de la planète aussi bien en

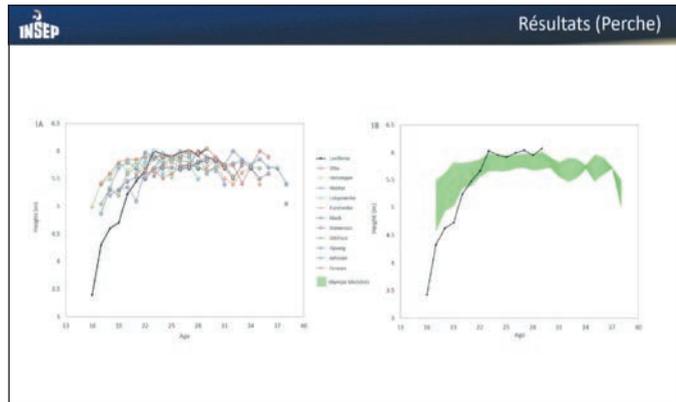


Figure 8

À gauche, les performances des sauteurs à la perche olympiques au cours des années. À droite, modélisation d'un couloir de performance issu de toutes ces performances.

termes de dénivelé, de nombre de virages, de distance que de performances des athlètes. **Le regroupement de toutes ces données permet de calculer des probabilités des scénarios de courses qui sont transmis aux coachs pour les aider dans leurs choix stratégiques et confirmer ou les rassurer dans leurs intuitions.**

Ainsi, pour l'épreuve de triathlon des Championnats d'Europe 2018, ce type d'étude avait montré que des nageurs allaient nager très vite pour essayer de créer les écarts de façon qu'il y ait, lors de la partie suivante en cyclisme, un peloton de tête et deux pelotons à la suite. Les simulations réalisées montraient qu'il fallait que l'athlète français sorte dans le premier groupe de tête en natation pour figurer sur le podium final. Les coachs savaient que les athlètes espagnols présents allaient nager très vite pour éliminer l'athlète français de la natation qui court très vite.

2 Monitoring des données

Les possibilités d'exploitation de toutes ces données sont nombreuses, comme le montre la **Figure 9** : optimisation de la performance, prévention des blessures, programmation de l'entraînement (charges d'entraînement, charges compétitives), etc.

Chaque terme de la **Figure 9** embarque derrière lui énormément de données. Par exemple, quand un staff demande d'intervenir pour avoir des informations sur le niveau de fatigue de ses athlètes, il y a de multiples façons d'étudier la fatigue : questionner pour en évaluer le niveau, effectuer un prélèvement salivaire pour avoir un ratio cortisol/testostérone, déterminer si c'est de la fatigue centrale

ou périphérique, c'est-à-dire cérébrale ou plutôt musculaire, ou de la fatigabilité, etc. Il faut donc faire des choix pour définir le type de données à capter et la façon de les capter, puis pouvoir et savoir les utiliser.

Des systèmes de gestion des athlètes centralisent les données.

Tout ce qui est capté est instantanément visible par les staffs et les athlètes, ce qui permet d'observer les cinétiques, l'évolution dans le temps et si elle sort de la variabilité individuelle.

Ensuite, le travail de recherche consiste à trouver les relations entre toutes les variables. L'hypothèse et la démarche consistent à aller chercher de l'information concernant des charges d'entraînement, des performances, des

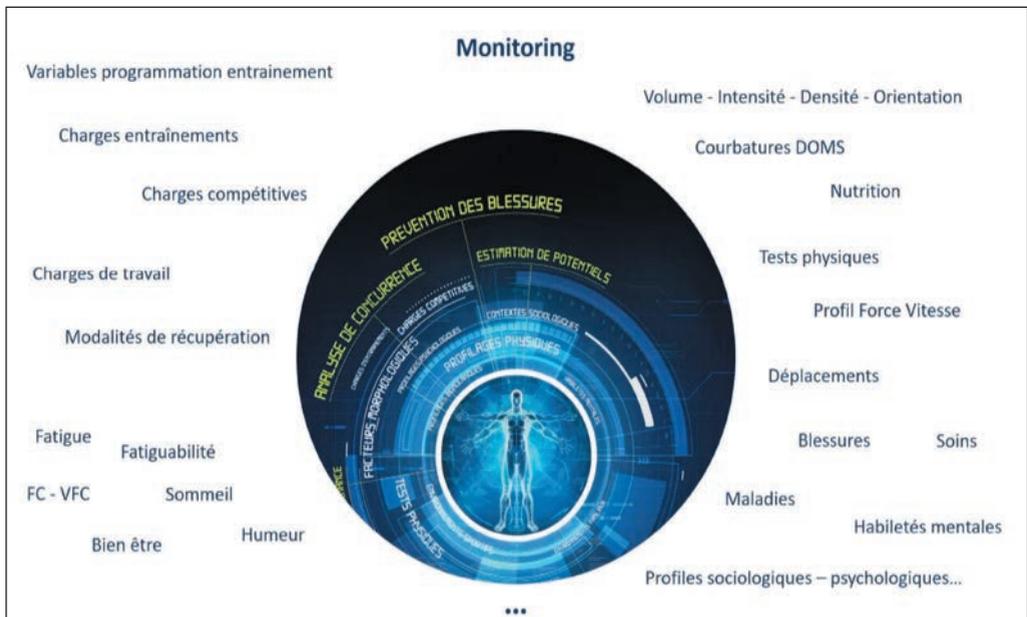


Figure 9

Applications du monitoring de données.



Figure 10

Nuage de notions pouvant affecter la performance de l'athlète.

blessures, la planification, des tests, autant de facteurs qui influencent les performances de l'athlète dans son écosystème.

La **Figure 10** présente tous les facteurs pouvant affecter la performance de l'athlète, mais les relations de ces facteurs entre eux ne sont pas identiques dans le temps. Ainsi, la fatigue dépend du type d'entraînement auquel on s'astreint selon le temps de sommeil. Pour tenter d'identifier ces relations qui sont individuelles et évolutives dans le temps, il faut capter des informations suffisamment pertinentes, mais qui peuvent se révéler intrusives, dès lors qu'on met des capteurs sur les matelas ou que l'on cherche des informations sur la situation familiale et sociale. Cela implique donc de fournir des explications aux athlètes et aux staffs.

La gestion des données sur l'athlète dans son écosystème n'est pas évidente, comme le montre l'exemple de la **Figure 11**.

Dans les deux cas, l'athlète s'est blessé, ce qui influe sur ses performances. En bas, la blessure est peut-être davantage en relation avec les charges d'entraînement ; la diminution du niveau de performance peut être liée à un sommeil impacté par des déplacements. En haut, elle est peut-être plus liée à ses capacités de récupération et à des charges d'entraînement qui étaient élevées, avec des niveaux de sommeil et de fatigue plus élevés qu'habituellement.

L'INSEP tente d'identifier les associations de facteurs qui prédisposent à des risques et d'établir des modèles de prédiction sur l'état de la forme de chaque athlète qui sont transmis aux décideurs.

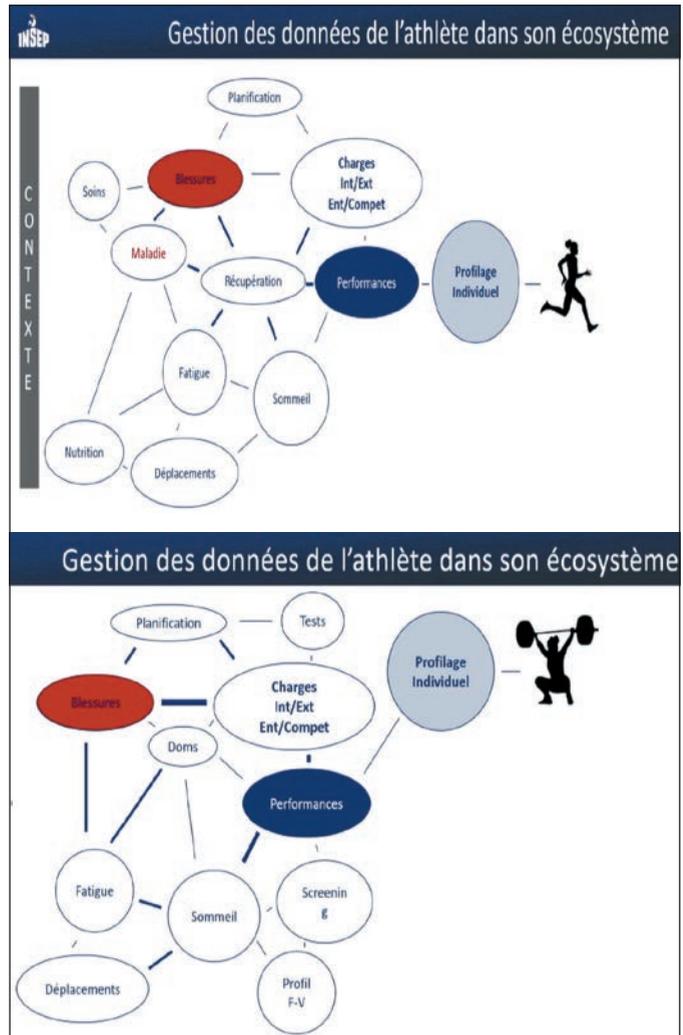


Figure 11

Les facteurs pouvant affecter la performance de l'athlète varient selon son écosystème.

Conclusion

La prédiction de performance aide les athlètes et leurs entraîneurs. Pour la mener à bien, le traitement d'une multitude de données issues de contenus déjà existants est nécessaire. Aujourd'hui, au lieu de mettre des capteurs, de faire venir les athlètes sur des plateformes de force dans les laboratoires, on capte l'information

en direct à partir d'une vidéo. L'INSEP travaille avec une dizaine de laboratoires de l'INRIA, spécialiste de l'image.

Par exemple, à partir de l'enregistrement d'une vidéo d'un match de boxe (**Figure 12**), on a la capacité de retracer et d'analyser toutes les données issues de l'ensemble des déplacements du boxeur rouge et du boxeur bleu. Cette pratique également usuelle pour d'autres sports permet de recueillir des informations sur les adversaires, mais aussi sur la quantification des charges d'entraînement, le but étant de recréer des profils force-vitesse *in situ* pour optimiser et individualiser les performances des athlètes.

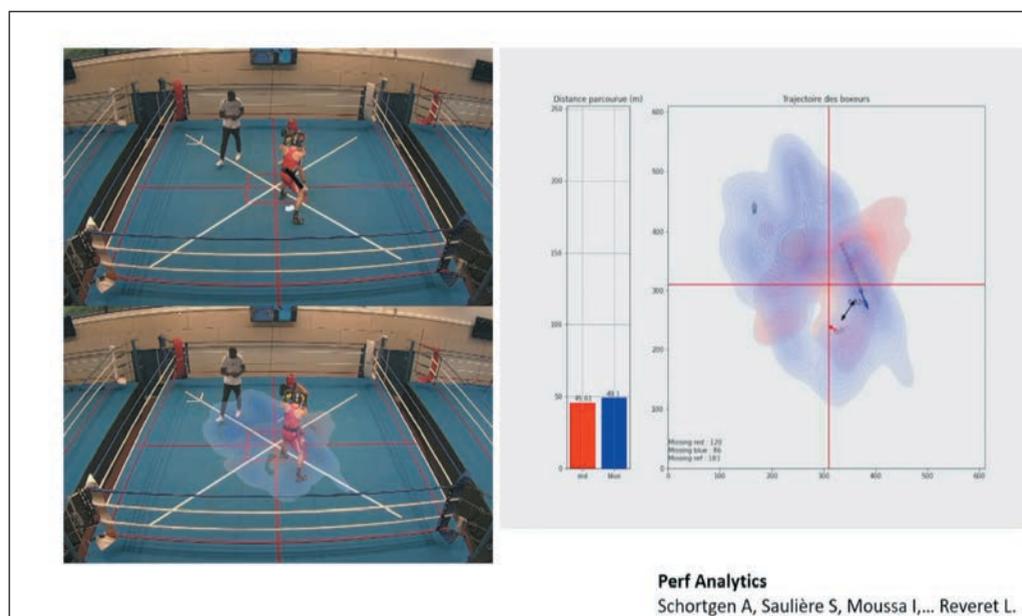


Figure 12

À gauche, image extraite d'une vidéo montrant l'acquisition de points issus du déplacement de 2 boxeurs. À droite, cartographie de leurs déplacements.