

ENSAE-ENSAI
&
OpenClassrooms

Certificat « Data Analyst »

**Analyse en composante principale
et caractérisation de la charge
d'entraînement au rugby en
comparaison avec la compétition**

Présenté par : Xavier Barbier

Mentor : Sarah Ghidalia, Data Scientist

RÉSUMÉ

Objectif : l'analyse visait utiliser une analyse en composante principale (ACP) pour identifier quelle combinaison de variables représentant les charges interne et externe permettant de réduire le nombre de dimensions caractérisant la charge d'entraînement (CE) au rugby. Ceci afin de proposer un seul indicateur décrivant la séance en comparaison avec le pire scénario possible. **Méthode** : Les données GPS et accéléromètres de 22 pratiquants masculins ont été recueillis pendant une saison sportive. 65 variables initiales ont été incluses dans un processus de réduction basé sur la qualité de représentation ($\cos^2 > 0.5$), la corrélation avec des composantes principales ($\text{corr} > 0.7$), la corrélation entre variables et leur contribution aux composantes principales. L'ACP a été conduite sur l'ensemble des sujets et des observations. **Résultats** : l'ACP finale explique 82.7% de la variance. Soit près de 30% de plus que l'ACP initiale. Les Dim.1 et Dim.2 comptent respectivement pour 68.2% et 14.5% de la variance expliquée. Les observations de compétition, "Forwards" et "8" se sont montrées les plus proches du pire scénario possible en comparaison des observations de leurs catégories respectives. **Conclusion** : Les résultats suggèrent que les praticiens pourraient caractériser la spécificité de la CE au rugby avec 1 indicateur de performance spécifique unique tout en limitant la perte d'informations.

Mots-clés: Rugby; GPS analysis; periodisation; training load; principal component analysis; Monitoring

INTRODUCTION

Dans le sport de compétition, le rôle des entraîneurs est de planifier et de proposer des séances d'entraînement pour améliorer la performance physique, et le développement des qualités techniques et tactiques des joueurs afin de permettre la réussite de ces derniers lors des compétitions [1].

Historiquement cette préparation à la performance scindait les contenus des séances en plusieurs composantes : préparation physique (PP), technique et tactique.

Vigarelo définissait la technique en 1983 comme " l'ensemble des moyens transmissibles à mettre en œuvre par l'homme, pour effectuer le plus efficacement possible une tâche motrice donnée" [43].

La tactique correspond à "l'activité décisionnelle réalisée en cours d'action" [44].

Pradet définissait en 1996 la PP comme "l'ensemble organisé et hiérarchisé des procédures d'entraînement qui visent au développement et à l'utilisation des qualités physiques du sportif" [31].

La PP est alors qualifiée selon la relation entre PP et technico-tactique. Classant la PP comme étant dissociée, associée ou intégrée [2] du contenu technico-tactique. Cette classification définit la PP dissociée en une séance planifiée dans une autre unité de temps. Puis la PP associée étant dans une même séance dans une même unité de temps, mais successivement. Enfin la PP intégrée comme étant le sport en lui-même. Les efforts étant ici modulés par modifications des espaces de jeu, règles, nombres de joueurs, ou bien encore droits et interdits pour partenaires et adversaires.

Les efforts réalisés sont plus communément appelés la charge d'entraînement (CE). La surveillance quotidienne de la CE dans le but de prendre des décisions fondées sur des preuves des effets de celle-ci est une stratégie très répandue dans le sport de compétition [3,4]. Répondant à des principes fondamentaux de planification (continuité, progressivité, réversibilité, alternance, spécificité), elle est traditionnellement quantifiée par une relation entre volume et intensité.

La CE est généralement représentée comme une CE externe et une CE interne. Ces dernières se définissent respectivement comme le travail effectué par l'athlète (par exemple, distance parcourue, nombre de sprints) et la réponse physiologique associée (par exemple, la fréquence cardiaque, la perception de l'effort) à l'activité [5].

Différentes publications ont également mis en lumière une relation entre la CE et la performance et entre la CE et les risques de blessure [6,7].

La densification du calendrier des compétitions et le développement d'outils de mesures de l'activité physique, ont eu pour conséquence une utilisation grandissante de séances de préparation physique intégrée. Les entraîneurs y voyant l'intérêt d'un gain de temps face à des calendriers hebdomadaires denses et également un moyen de préparer les athlètes aux demandes de la compétition. Enfin, une adhésion accrue des pratiquants ne saurait être négligée. Par exemple, les données d'entraînement collectées par les entraîneurs grâce aux

GPS ont déjà été identifiées pour aider à détecter les risques de blessures [8] et les changements dans les qualités physiques [9]. C'est donc pourquoi les GPS (Géo-localisation par satellite) font actuellement partie des outils les plus communément utilisés par les équipes de sports collectifs professionnels [3].

La technologie GPS a été créée en 1973 par l'armée des Etats-Unis. L'ouverture de son utilisation au public en 1996 a permis le développement des premiers appareils en Australie dès le début des années 2000 [19].

Les capteurs dits GPS se composent, comme leur nom l'indique, d'un dispositif GPS, mais également le plus souvent de capteurs inertiels.

Même si la qualité des données recueillies dépend de la qualité de la technologie utilisée [10], et que des réserves subsistent pour les données d'impacts et collisions [18], les données liées à la locomotion sont actuellement considérées comme fiables [11,12,13,14,15,16,17]. Avec plus de 200 indicateurs proposés [19,20,21], la problématique d'identification des indicateurs clés de performance et de leur utilisation, a mené, entre autres, à la naissance de postes dédiés au suivi scientifique au sein des structures sportives de haut-niveau. Avec pour mission de collecter, analyser et proposer un retour d'informations. Ces informations ayant pour objectifs d'être utilisées pour évaluer et améliorer les pratiques et la prise de décision des entraîneurs [24,25,26]. Actuellement, les accélérations (à différents seuils), la distance totale, la distance à haute vitesse et les variables issues de l'évaluation métabolique (Annexe 1) comptent parmi les variables les plus utilisées dans le suivi de l'entraînement pour décrire les séances et la CE induite [3].

Considérant la prise de décision et la planification comme un aspect très important du rôle de l'entraîneur [22,23], la communication des 7-8 indicateurs clés les plus utilisés est ensuite devenue une nouvelle problématique. De précédents travaux ont montré que bien que les entraîneurs aient des connaissances en sciences du sport, ils perçoivent ces données comme seulement quatrièmes dans leurs intérêts pour les prises de décisions [27]. De même ils estiment que "trop d'informations", "mauvaise communication du département de sciences du sport" et "l'absence de but commun" sont des obstacles à l'utilisation des données de la science du sport pour leur prise de décision [28]. Ces travaux mettant également en lumière un décalage entre les entraîneurs et le "pôle data". Cherchant à résoudre cette problématique la communauté scientifique s'attache depuis peu à définir des bonnes pratiques dans l'analyse ou la visualisation des données sportives [29]. Par ailleurs, certaines études visent à réduire le nombre de dimensions/variables présentées. Mais également retrouver une description de la CE sous une forme plus traditionnelle, de volume et intensité, exprimée en pourcentage du pire scénario de compétition [30].

Malgré cela, la sélection des variables reste basée sur une littérature décrivant des relations bivariées et de variables isolées [32, 33, 34, 35, 36, 37]. Ceci s'ajoutant à l'absence de pondération de celles-ci dans le calcul, laisse un doute sur la qualité de la réduction de dimension.

C'est sur la base de ces éléments que l'intérêt de la méthode de l'analyse en composante principale (ACP) a été envisagée.

L'objet de cette analyse est d'utiliser une ACP pour de réduire le nombre de dimensions utilisées afin de caractériser les séances en comparaison des demandes de la compétition, et ainsi aider les entraîneurs dans leur planification des séances.

MÉTHODE

Participants

L'analyse porte sur 22 pratiquants de rugby. Plus d'informations ne peuvent être indiquées en raison d'un accord de confidentialité.

Procédure

Les données ont été recueillies sur 42 semaines et fournies sur 2 tables : entraînements et compétitions. Les données d'entraînement concernent 155 séances et les données de compétitions concernent 30 rencontres.

Données brutes

La table entraînement contient 2243 observations et 81 variables.

La table compétitions contient 279 observations et 90 variables.

Chaque observation décrit une séance, entraînement ou compétition. Celles-ci sont caractérisées par des variables qualitatives et quantitatives. Les variables qualitatives indiquent les informations relatives à l'un des sujets (Athlète, Squad, Position, Zone, Participation), et à la séance (Session, Drill, Date, Week, Day Code, Weather, Grass).

Les variables quantitatives désignent les indicateurs recueillis par les unités GPS individuelles.

Les seuils de vitesse étaient individualisés conformément à de précédentes recherches [50].

Préparation

Seules les variables disponibles dans les 2 tables ont été conservées. Les doublons ont été supprimés. Les valeurs manquantes ont été traitées par une imputation par la moyenne en associant des variables qualitatives (Athlete, Session, Drill, Squad, Date).

L'ensemble des variables sont présentées dans le tableau 1 en annexe. Les variables quantitatives sont considérées comme des indicateurs de volume ou d'intensité.

Analyse statistique

Une analyse en composante principale (APC) a été utilisée. Cette méthode de réduction de dimension s'inscrit dans notre objectif et a déjà fait l'objet de travaux précédemment [38,39,40,41,42]. Les observations présentant des données aberrantes n'ont pas été retenues pour l'analyse. La matrice de corrélations (Annexe, figure 1) indiquent des variables majoritairement corrélées en elles.

Toutes les données ont été centrées et réduites avant la réalisation de l'ACP.

62 variables sont utilisées lors de cette APC initiale mais seulement les mieux représentées et les plus corrélées avec les composantes peuvent la variabilité des observations.

Une sélection des variables a ensuite été réalisée à partir d'un seuil de corrélation ($\text{corr} > 0.7$) et d'un seuil de qualité de représentation ($\text{cos}^2 > 0.5$) conformément avec des travaux précédents [39,40,41].

Les variables similaires ou fortement corrélées ont été réduites au profit de celles présentant les contributions les plus importantes aux composantes principales.

Cette sélection de variables a été utilisée afin de projeter les observations sur le 1er plan factoriel. Pour chaque sujet, des observations théoriques "Game max" correspondant aux valeurs maximales pour chaque variable ont été calculées, dans une démarche semblable à de précédents travaux [30].

Pour chaque sujet, cette distance sur le plan factoriel entre "game max" et l'observation (DGM) caractérise la différence entre l'effort de la séance et le pire scénario possible en compétition.

Cette distance a été mesurée en distance Manhattan, appelée aussi taxi-distance, qui est définie par la distance entre deux points parcourue par un taxi lorsqu'il se déplace dans une ville où les rues sont agencées selon un réseau ou quadrillage et en utilisant les déplacements horizontaux et verticaux du réseau.

Les moyennes et écarts-types pour la DGM par groupe "Type", "squad" et "position" ont été calculées.

Les données ont été préparées et analysées en utilisant le logiciel RStudio Cloud (Version 1.2.5042-1)

RÉSULTATS

53,5% de la variance totale est expliquée grâce aux 2 dimensions obtenus lors de l'ACP initiale regroupant 62 variables. Les Dim.1 et Dim.2 comptent respectivement pour 38.4% et 15.2% de la variance expliquée. L'éboulis des valeurs propres et les cercles corrélations de l'ACP initiale sont présentés dans les figures 2 et 3.

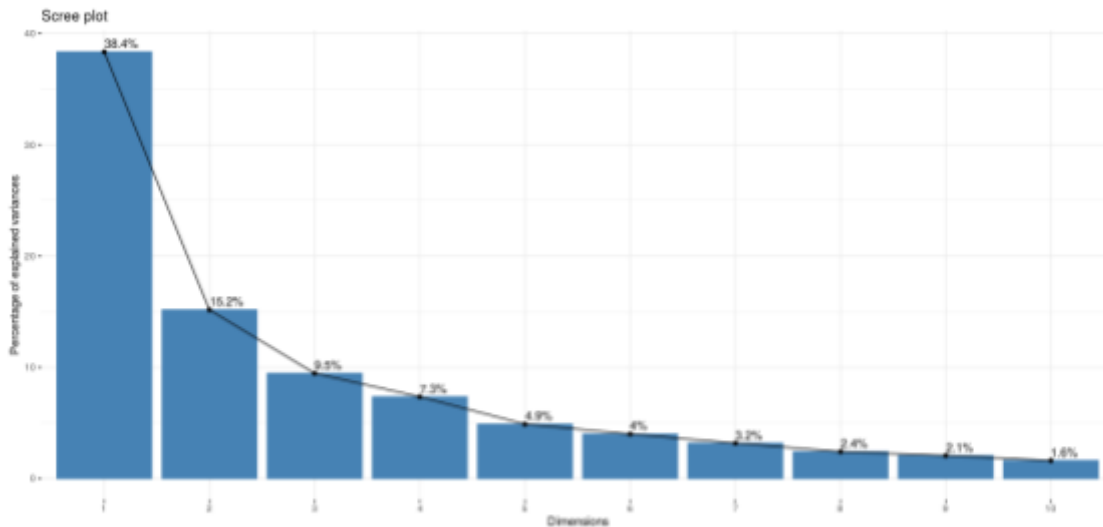


Figure 2 : Eboulis des valeurs propres en %

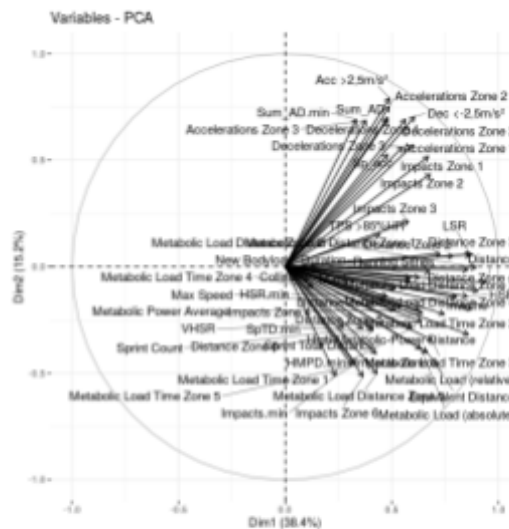


Figure 3 : cercle des corrélations

La sélection des variables pour l'ACP finale a permis de retenir 6 variables : "Acc >2.5m/s²", "Distance", "High metabolic power distance", "HSR", "LSR" et "Metabolic load distance Zone 1".

6 variables ont été conservées pour l'ACP finale. Les 6 variables retenues présentent encore une corrélation globale significative (Fig.4).

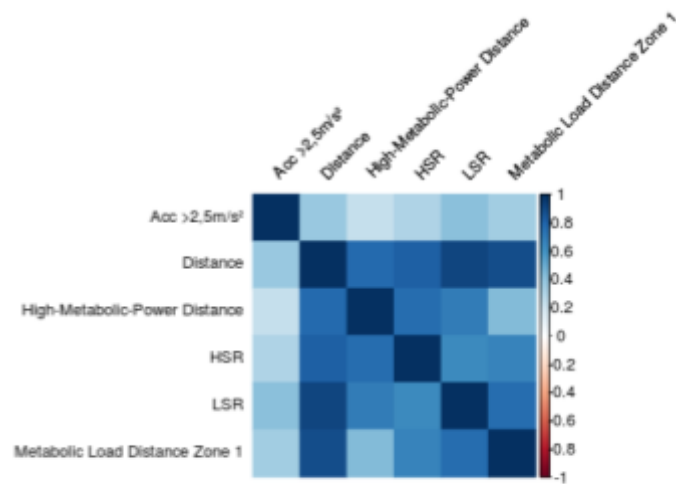


Figure 4 : matrice des corrélations 6 variables finales

L'ACP finale permet d'expliquer 82,7% de la variance totale (Fig.3) sur le 1er plan factoriel. Soit près de 30% de plus que l'ACP initiale. Les Dim.1 et Dim.2 comptent respectivement pour 68.2% et 14.5% de la variance expliquée. L'éboulis des valeurs propres et les cercles corrélations de l'ACP finale sont présentés dans les figures 5 et 6.

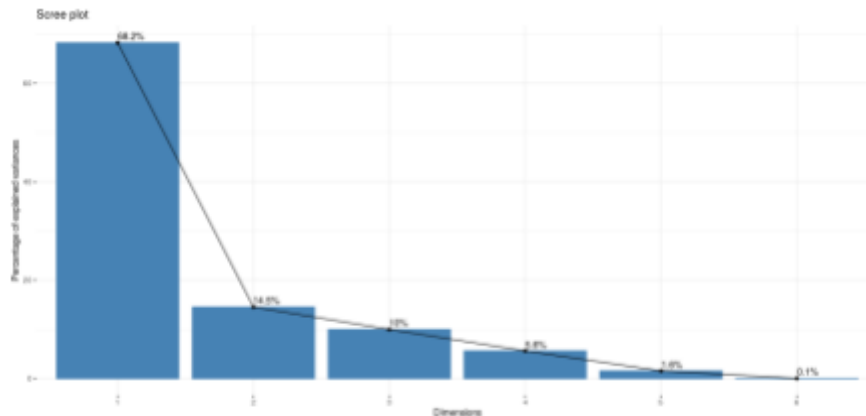


Figure 5 : éboulis des valeurs propres ACP finale

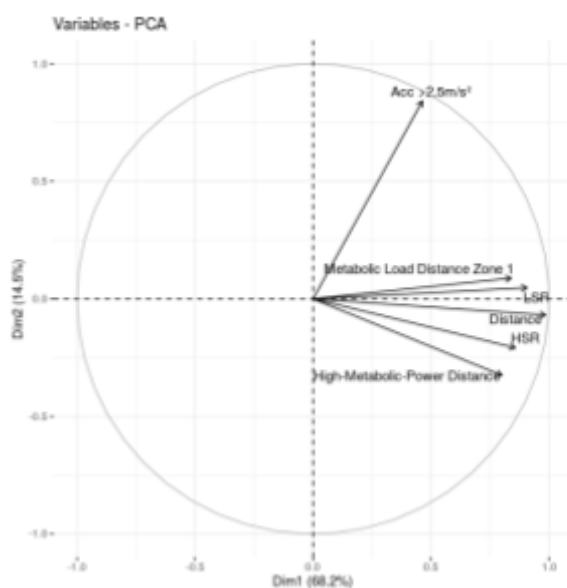


Figure 6 : cercle des corrélations ACP finale

Le tableau 2 propose les contributions des variables aux différentes dimensions du 1er plan factoriel. La figure 7 expose l'évolution des contributions des variables sur l'ensemble des dimensions de l'ACP.

Tableau 2 : contribution des variables aux composantes principales		
Variables	Contribution Dim.1	Contribution Dim.2
Distance	23.74327	0.5434410
LSR	20.11687	0.2555846
HSR	17.96003	4.9749076
Acc >2,5m/s ²	5.25443	81.1827895
High-Metabolic-Power Distance	15.67443	12.1506777
Metabolic Load Distance Zone 1	17.25097	0.8925996

Contribution des variables aux dimensions

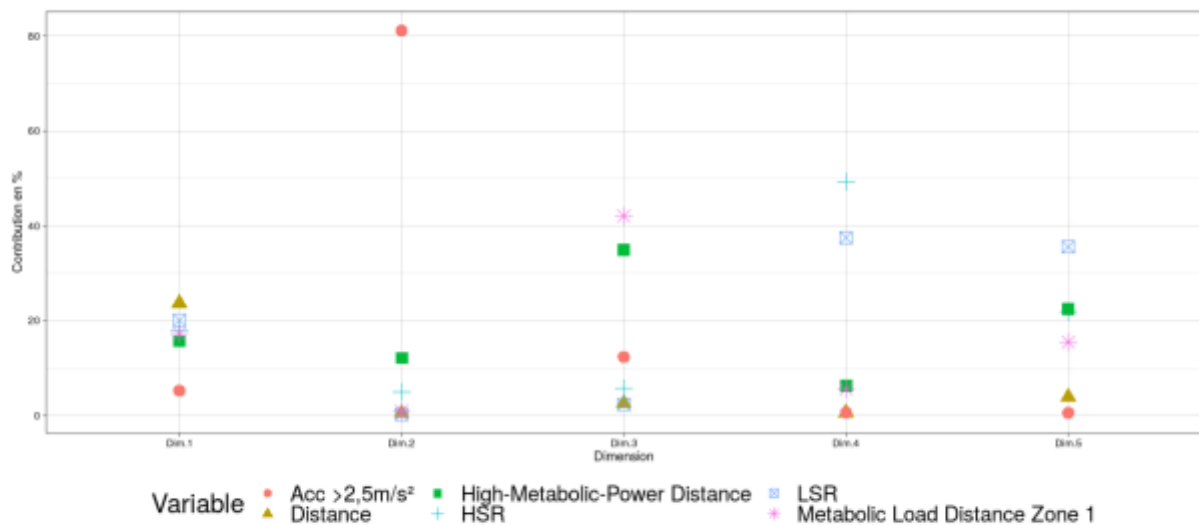


Figure 7 : contribution des variables selon les dimensions (%)

Les variables contribuent entre 5 et 23% pour la Dim.1 et 0.2 et 81% pour la Dim.2.

Les variables les plus importantes sont "Distance" et "Acc >2.5m/s²", respectivement pour les Dim.1 et Dim.2.

La figure 8 présente les projections sur le 1er plan factoriel des observations correspondant aux valeurs maximales pour chaque sujet ("Game max").

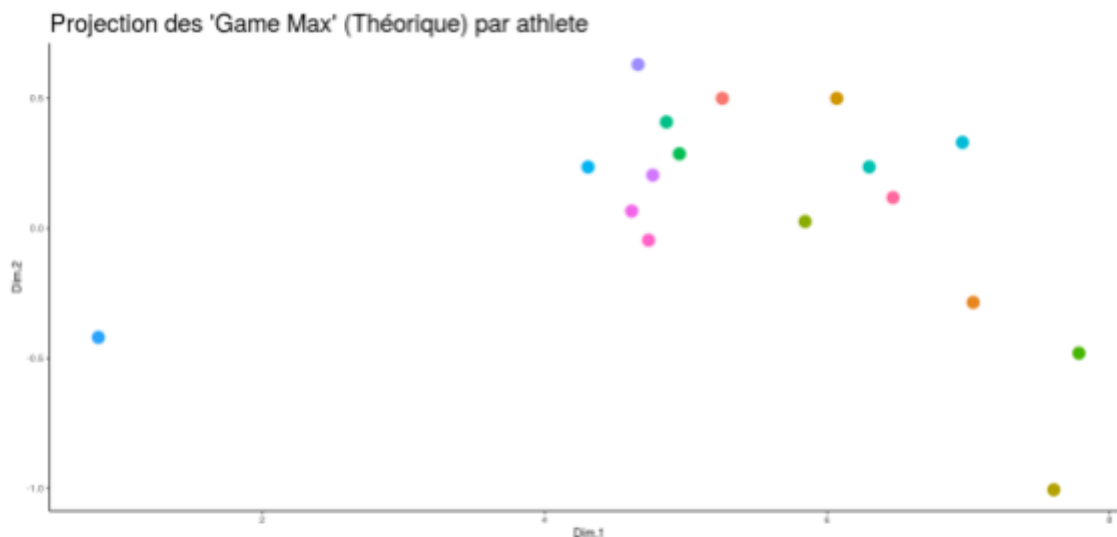


Figure 8 : Projection des observations maximales par athlète sur 1 plan

La figure 9 illustre la projection sur le 1er plan factoriel de l'ensemble des observations.



Figure 9 : projection des observations selon type

La figure 10 expose les valeurs selon le type de séance.

La moyenne pour "distance game max" (DGM) obtenue est de 6.42 ± 3.39 .

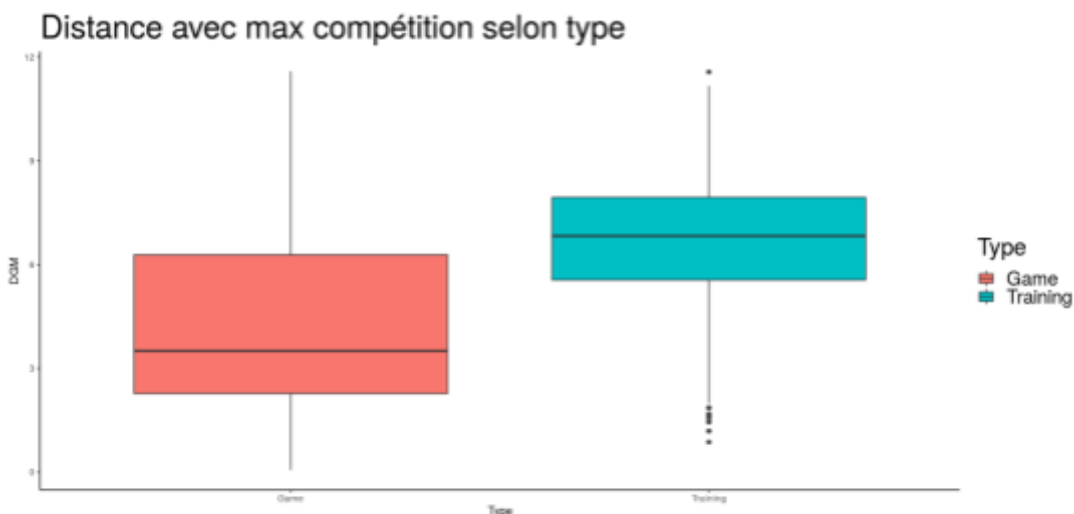


Figure 10 : DGM selon type de séance

Les figures 11 et 12 présentent les différentes valeurs selon les groupes "Squad" et "Position". Le tableau 3 offre les valeurs moyennes de DGM pour les différents groupes.

La moyenne pour DGM obtenue est de 3.37 ± 2.83 pour le groupe "Game" et 6.69 ± 1.74 pour le groupe "Training".

La moyenne pour DGM obtenue est de 6.73 ± 2.1 pour le groupe "backs" et 5.91 ± 1.99 pour le groupe "Forwards".

La moyenne pour DGM obtenue est 6.01 ± 2.0 pour le groupe "2eme ligne", 2.25 ± 1.64 pour le groupe "Prop", 2.65 ± 0.8 pour le groupe "8", 6.78 ± 2.0 pour le groupe "Fly half", 7.55 ± 1.1 pour le groupe "Scrum half", 6.40 ± 2.4 pour le groupe "Center", 6.95 ± 1.94 pour le groupe "Wings", 5.55 ± 1.9 pour le groupe "Fullback" et 5.95 ± 2.1 pour le groupe "Flanker".

Tableau 3 : distance avec pire scénario possible (DGM) selon groupes (moyenne ± ET)

Groupes	DGM
Game	3.37±2.83
Training	6.69±1.74
Backs	6.73±2.1
Forwards	5.91±1.99
2eme ligne	6.01±2.0
Prop	2.25±1.64
8	2.65±0.8
Fly half	6.78±2.0
Scrum half	7.55±1.1
Center	6.40±2.4
Wings	6.95±1.94
Fullback	5.55±1.9
Flanker	5.95±2.1

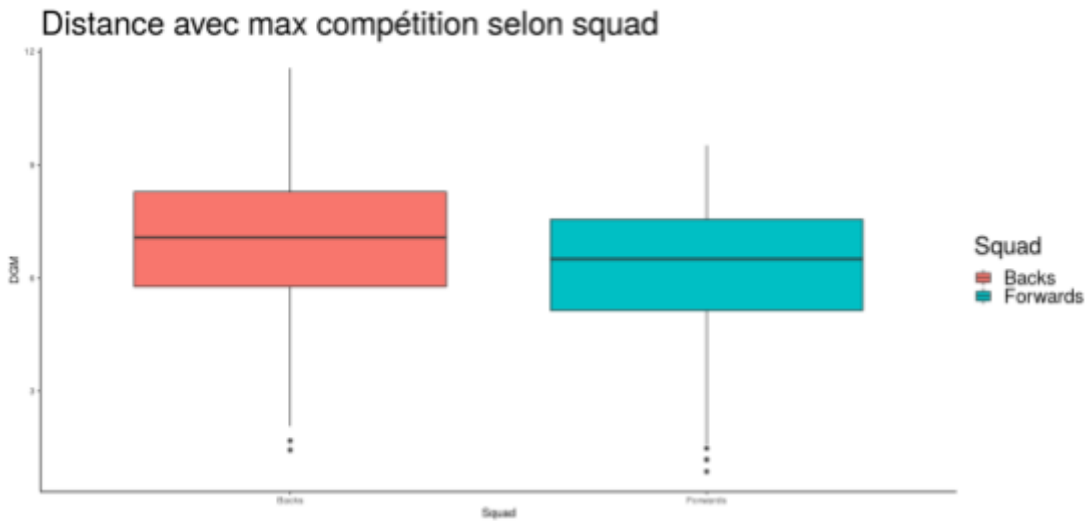


Figure 11 : DGM selon squad

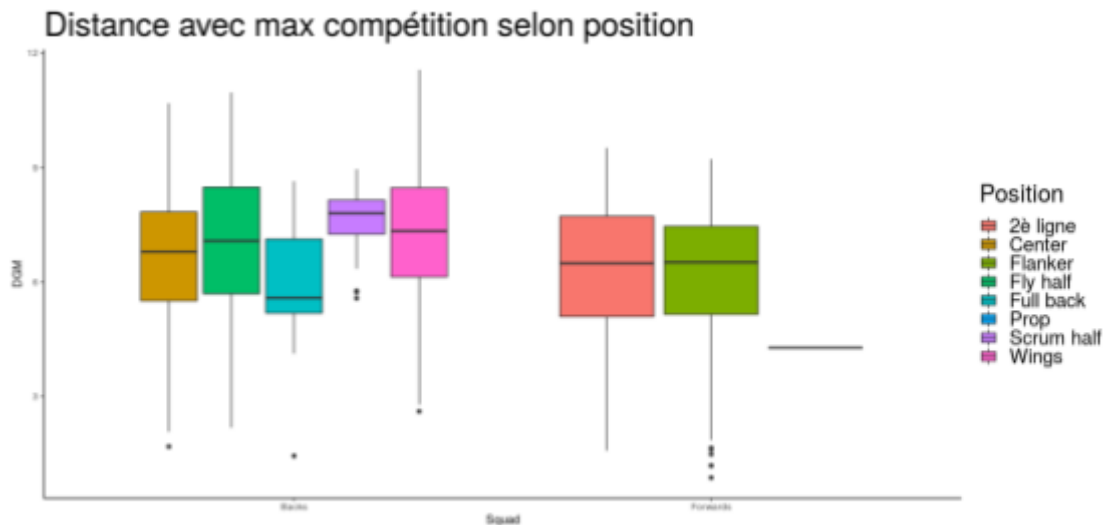


Figure 12 : DGM selon position

DISCUSSION

L'analyse avait pour objectif dans un premier temps d'utiliser une ACP afin de réduire les dimensions caractérisant la charge d'entraînement de séance de rugby. Puis d'en un second temps d'utiliser la distance Manhattan afin d'apprécier la différence entre chaque séance et le pire scénario possible en compétition.

Les critères de sélection ont permis de réduire à 6 variables finales sur les 65 initiales. Des seuils similaires à de précédents travaux ont été utilisés et d'autres valeurs seuils auraient certainement amenées à des résultats différents. Cependant, une variance expliquée à hauteur de 82.7% sur le 1er plan factoriel indique une réduction de dimensions sans perte d'information efficace.

Aucun indicateur caractérisant l'intensité n'a été retenu. La proportion importante d'observations d'entraînement en comparaison des observations. L'intensité des séances d'entraînement pouvant être plus homogène en elle, et moins avec les compétitions. Il serait nécessaire de procéder à des ACP en séparant les deux situations afin d'observer quels indicateurs sont conservés dans chacun des contextes.

Il est à noter l'absence dans les variables retenues de variables concernant exclusivement les impacts. Une hypothèse serait une mauvaise mesure des impacts. Ce point a d'ailleurs été mis en lumière à travers différentes études [45,46,47,48]. Une seconde hypothèse serait la nature même de l'ACP. Les variables liées aux impacts sont fortement asymétriques (oblique positive et tendance centrale proche de 0) et les valeurs extrêmes sont regroupées dans un faible nombre d'observations. L'ACP retient fort logiquement les variables permettant d'expliquer un maximum de variance pour l'ensemble des observations.

Parmi les 6 variables finales, 3 variables ("Acc >2.5m/s²", "Distance", "HSR") sont classiquement utilisées [3].

Les 3 autres (“High metabolic power distance”, “LSR”, “Metabolic load distance Zone 1”) le sont moins systématiquement.

Au regard de la nature des variables et de leurs contributions aux composantes du 1er plan, la Dim.1 pourrait caractériser le “Volume”, tandis que la Dim.2 pourrait caractériser “l’intensité”.

La projections des observations sur le 1er plan factoriel permet, comme attendu, de mettre en avant une zone différente entre entraînement, compétition et “game max”. On remarque que les observations de compétitions sont principalement caractérisées par la Dim.1. Tandis que les observations d'entraînement semblent plus étalées sur l'ensemble sur 1er plan factoriel.

Les valeurs de DGM présentées mettent en lumière une différence entre “game” et “training”. Le match étant fort logiquement plus proche des demandes maximales.

On observe que les distributions de GDM selon “squad” sont globalement similaires, mais que l’hypothèse d’une différence inter-position pour le groupe “forwards” demeure. Il serait intéressant de vérifier les liens entre “position” et DGM afin de déterminer les différences d’exposition à une CE spécifique. Une hypothèse pour expliquer ces différences entre groupe de “position” serait le choix de réaliser une ACP sur l’ensemble du groupe. Il pourrait être logique d’attendre des résultats différents d’une sélection de variables, puis d’une ACP, par groupe de “position”. Une tendance qui pourrait être similaire à de précédents travaux semblent indiquer, a minima, une variation de la variances expliquée et des différentes contributions aux composantes principales selon le groupe “position” [38].

Enfin, il serait nécessaire de reproduire cette approche dans différents contextes (niveau de compétition, style de jeu, classement, etc) afin d’en apprécier les nuances.

CONCLUSION

L’analyse a présenté une méthode contemporaine et claire pour réduire le nombre d’indicateurs de performance spécifique au rugby. Les résultats suggèrent que les praticiens pourraient caractériser la spécificité de la CE au rugby avec 1 variable unique tout en limitant la perte d’informations.

L’utilisation de cette approche peut permettre de combiner des indicateurs clés de charges interne et externe utilisés dans le cadre du suivi de l’athlète. Puis de proposer un indicateur clé spécifique à leur contexte particulier.

Remerciements

Sarah Ghidalia pour sa curiosité sur le sujet et ses remarques avisées.

BIBLIOGRAPHIE

- 1) Williams, A & Reilly, T.P.. (2000). Talent identification and development in soccer. *Journal of sports sciences*. 18. 657-67. 10.1080/02640410050120041.
- 2) Aubert, F (2002). *Approches athlétiques de la Préparation Physique*. Colloque formation continue. Tours.
- 3) Akenhead, Richard & Nassis, George. (2015). Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *International Journal of Sports Physiology and Performance*. 10.1123/ijsp.2015-0331.
- 4) Mccall, Alan & Davison, Michael & Andersen, Thor & Beasley, Ian & Bizzini, Mario & Dupont, Gregory & Duffield, Rob & Carling, Christopher & Dvorak, Jiri. (2015). Injury prevention strategies at the FIFA 2014 World Cup: Perceptions and practices of the physicians from the 32 participating national teams. *British journal of sports medicine*. 49. 603-8. 10.1136/bjsports-2015-094747.
- 5) Wallace, L & Slattery, Katie & Coutts, Aaron. (2013). A comparison of methods for quantifying training load: Relationships between modelled and actual training responses. *European journal of applied physiology*. 114. 10.1007/s00421-013-2745-1.
- 6) Eirale, Cristiano & Tol, Johannes & Farooq, Abdulaziz & Smiley, Faten & Chalabi, Hakim. (2012). Low injury rate strongly correlates with team success in Qatari professional football. *British journal of sports medicine*. 47. 10.1136/bjsports-2012-091040.
- 7) Häggglund, Martin & Waldén, Markus & Magnusson, Henrik & Kristenson, Karolina & Bengtsson, Håkan & Ekstrand, Jan. (2013). Injuries affect team performance negatively in professional football: An 11-year follow-up of the UEFA Champions League injury study. *British journal of sports medicine*. 47. 10.1136/bjsports-2013-092215.
- 8) Pappalardo, Luca & Rossi, Alessio & Cintia, Paolo. (2017). Who is going to get hurt? Predicting injuries in professional soccer. 10.13140/RG.2.2.19505.28006.
- 9) Clemente, Filipe & Nikolaidis, Pantelis & Rosemann, Thomas & Knechtle, Beat. (2019). Dose-Response Relationship Between External Load Variables, Body Composition, and Fitness Variables in Professional Soccer Players. *Frontiers in Physiology*. 10. 1-9. 10.3389/fphys.2019.00443.
- 10) Coutts, Aaron & Duffield, Rob. (2009). Validity and reliability of GPS devices for measuring movement demands of team sports. *Journal of science and medicine in sport / Sports Medicine Australia*. 13. 133-5. 10.1016/j.jsams.2008.09.015.
- 11) Jennings, Denise & Cormack, Stuart & Coutts, Aaron & Boyd, Luke & Aughey, Robert. (2010). The Validity and Reliability of GPS Units for Measuring Distance in Team Sport Specific Running Patterns. *International journal of sports physiology and performance*. 5. 328-41. 10.1123/ijsp.5.3.328.

- 12) Aughey, Robert. (2011). Applications of GPS Technologies to Field Sports. International journal of sports physiology and performance. 6. 295-310. 10.1123/ijsp.6.3.295.
- 13) Waldron, Mark & Worsfold, Paul & Twist, Craig & Lamb, Kevin. (2011). Concurrent validity and test-retest reliability of a global positioning system (GPS) and timing gates to assess sprint performance variables. Journal of sports sciences. 29. 1613-9. 10.1080/02640414.2011.608703.
- 14) Varley, Matthew & Fairweather, Ian & Aughey, Robert. (2011). Validity and reliability of GPS for measuring instantaneous velocity during acceleration, deceleration, and constant motion. Journal of sports sciences. 30. 121-7. 10.1080/02640414.2011.627941.
- 15) Johnston, Richard & Watsford, Mark & Pine, Matthew & Spurr, Robert & Sporre, Daniel. (2013). Assessment of 5 Hz and 10 Hz GPS units for measuring athlete movement demands. International Journal of Performance Analysis in Sport. 13. 10.1080/24748668.2013.11868646.
- 16) Vickery, Will & Dascombe, Ben & Baker, John & Higham, Dean & Spratford, Wayne & Duffield, Rob. (2013). Accuracy and Reliability of GPS Devices for Measurement of Sports-Specific Movement Patterns Related to Cricket, Tennis, and Field-Based Team Sports. Journal of strength and conditioning research / National Strength & Conditioning Association. 28. 10.1519/JSC.0000000000000285.
- 17) Roe, Gregory & Darrall-Jones, Josh & Shaw, Will & Black, Christopher & Till, Kevin & Jones, Ben. (2016). Validity of 10 HZ GPS and Timing Gates for Assessing Maximum Velocity in Professional Rugby Union Players. International Journal of Sports Physiology and Performance. 12. 10.1123/ijsp.2016-0256.
- 18) Gabbett, Tim. (2012). Quantifying the Physical Demands of Collision Sports: Does Microsensor Technology Measure What It Claims to Measure?. Journal of strength and conditioning research / National Strength & Conditioning Association. 27. 10.1519/JSC.0b013e318277fd21.
- 19) www.catapultsports.com/about
- 20) mclloyd.com/
- 21) statsports.com/pro-series-software/
- 22) Cushion, Christopher & Nelson, Lee & Armour, Kathy & Lyle, John & Jones, Robyn & Sandford, Rachel & Callaghan, Claire. (2015). Coach Learning and Development: A Review of Literature.
- 23) Mata, Rui & Gomes, Rui. (2013). Winning or not winning: The influence on coach-athlete relationships and goal achievement. Journal of Human Sport and Exercise. 8. 986-995. 10.4100/jhse.2013.84.09.
- 24) Buchheit M. 2017. Want to see my report coach. Aspetar Sports Med J. 6:36–43.

- 25) Robertson, Samuel & Bartlett, Jon & Gatin, Paul. (2016). Red, Amber or Green? Athlete Monitoring in Team Sport: The Need for Decision Support Systems. *International journal of sports physiology and performance*. 12. 10.1123/ijsp.2016-0541.
- 26) Ward, Patrick & Windt, Johann & Kempton, Thomas. (2019). Business Intelligence: How Sport Scientists Can Support Organisation Decision Making in Professional Sport. *International Journal of Sports Physiology and Performance*. 14. 1-10. 10.1123/ijsp.2018-0903.
- 27) Brink, M.S. & Kuyvenhoven, Jurian & Toering, Tynke & Jordet, Geir & Frencken, Wouter. (2018). What do football coaches want from sport science?. *Kinesiology*. 50.
- 28) Nosek, Perry & Brownlee, Thomas & Drust, Barry & Andrew, Matthew. (2020). Feedback of GPS training data within professional English soccer: A comparison of decision making and perceptions between coaches, players and performance staff. *Science and Medicine in Football*. 10.1080/24733938.2020.1770320.
- 29) Thornton, Heidi & Delaney, Jace & Duthie, Grant & Dascombe, Ben. (2019). Developing Athlete Monitoring Systems in Team Sports: Data Analysis and Visualization. *International Journal of Sports Physiology and Performance*. 10.1123/ijsp.2018-0169.
- 30) Owen, Adam & Djaoui, Léo & Newton, Matt & Malone, Shane & Mendes, Bruno. (2017). A contemporary multi-modal mechanical approach to training monitoring in elite professional soccer. *Science and Medicine in Football*. 1-6. 10.1080/24733938.2017.1334958.
- 31) *La Préparation physique*, Michel Pradet, (1996)
- 32) Hill-Haas, Stephen & Dawson, Brian & Impellizzeri, Franco & Coutts, Aaron. (2011). Physiology of Small-Sided Games Training in Football. *Sports medicine (Auckland, N.Z.)*. 41. 199-220. 10.2165/11539740-000000000-00000.
- 33) Owen, Adam & Wong, Del P. & Paul, Darren & A, Dellal. (2014). Physical and Technical Comparisons between Various-Sided Games within Professional Soccer. *International journal of sports medicine*. 35. 286-282. 10.1055/s-0033-1351333.
- 34) Malone, James & Di Michele, Rocco & Morgans, Ryland & Burgess, Darren & Morton, James & Drust, Barry. (2014). Seasonal Training-Load Quantification in Elite English Premier League Soccer Players. *International journal of sports physiology and performance*. 10. 10.1123/ijsp.2014-0352.
- 35) Raastad, Truls & Owe, Simen & Paulsen, Gøran & Enns, Deborah & Overgaard, Kristian & Crameri, Regina & Kiiil, Steinar & Belcastro, Angelo & Bergersen, Linda & Hallén, Jostein. (2010). Changes in Calpain Activity, Muscle Structure, and Function after Eccentric Exercise. *Medicine and science in sports and exercise*. 42. 86-95. 10.1249/MSS.0b013e3181ac7afa.
- 36) Young, Warren & Hepner, Jamie & Robbins, Daniel. (2012). Movement Demands in Australian Rules Football as Indicators of Muscle Damage. *Journal of strength and conditioning research / National Strength & Conditioning Association*. 26. 492-6. 10.1519/JSC.0b013e318225a1c4.

- 37) Silva, João & Rebelo, António & Marques, Franklim & Pereira, Laura & Seabra, André & Ascensão, António & Magalhães, José. (2014). Biochemical impact of soccer: An analysis of hormonal, muscle damage, and redox markers during the season. *Applied physiology, nutrition, and metabolism = Physiologie appliquée, nutrition et métabolisme*. 39. 432-8. 10.1139/apnm-2013-0180.
- 38) Weaving D, Dalton NE, Black C, Darrall-Jones J, Phibbs PJ, Gray M, et al. The Same Story or a Unique Novel? Within-Participant Principal-Component Analysis of Measures of Training Load in Professional Rugby Union Skills Training. *International Journal of Sports Physiology and Performance*. 2018 Oct; 13 (9):1175–81.
- 39) Dubois R, Bru N, Paillard T, Le Cunuder A, Lyons M, Maurelli O, et al. (2020) Rugby game performances and weekly workload: Using of data mining process to enter in the complexity. *PLoS ONE* 15(1): e0228107.
- 40) Weaving, Dan & Jones, Ben & Till, Kevin & Marshall, Phil & Abt, Grant. (2017). Multiple Measures are Needed to Quantify Training Loads in Professional Rugby League. *International Journal of Sports Medicine*. 38. 10.1055/s-0043-114007.
- 41) Williams, Sean & Trewartha, Grant & Cross, Matthew & Kemp, Simon & Stokes, Keith. (2016). Monitoring What Matters: A Systematic Process for Selecting Training Load Measures. *International Journal of Sports Physiology and Performance*. 12. 1-20. 10.1123/ijsp.2016-0337.
- 42) Parmar, Nimai & James, Nic & Hearne, Gary & Jones, Ben. (2018). Using principal component analysis to develop performance indicators in professional rugby league. *International Journal of Performance Analysis in Sport*. 18. 1-12. 10.1080/24748668.2018.1528525.
- 43) Vigarello, G., Vives, J., (1983). Technique corporelle et discours technique. *Revue EPS*, 184, 40-47.
- 44) Mouchet, Alain. (2014). Intelligence tactique en sports collectifs.
- 45) Gabbett, Tim & Jenkins, David & Abernethy, Bruce. (2010). Physical collisions and injury during professional rugby league skills training. *Journal of science and medicine in sport / Sports Medicine Australia*. 13. 578-83. 10.1016/j.jsams.2010.03.007.
- 46) Kelly, Daniel & Coughlan, Garrett & Green, Brian & Caulfield, Brian. (2012). Automatic detection of collisions in elite level rugby union using a wearable sensing device. *Sports Engineering*. 15. 10.1007/s12283-012-0088-5.
- 47) Clarke, Anthea & Anson, Judith & Pyne, David. (2016). Proof of Concept of Automated Collision Detection Technology in Rugby Sevens. *Journal of Strength and Conditioning Research*. 31. 1. 10.1519/JSC.0000000000001576.
- 48) Brick, Maxime. (2018). Caractérisation des actions de combats en Rugby à 7 féminin par mesure accélérométriques. *Science & Rugby. FFR. Etude préliminaire*.
- 50) A, Dellal & Chamari, Karim & Wong, Del P. & Ahmaidi, Said & Keller, Dominique & Barros, Ricardo & Bisciotti, Gian Nicola & Carling, Christopher. (2011). Comparison of

physical and technical performance in European professional soccer match-play: The FA Premier League and La LIGA. European Journal of Sport Science. 11. 51-59. 10.1080/17461391.2010.481334.

ANNEXE

Tableau 1 : présentation des variables				
VARIABLE	TYPE	NIVEAUX	SEUILS	DESCRIPTION
Athlete	Qualitative	22		ID
Position	Qualitative	9		Poste joué
Squad	Qualitative	2		Groupe de position
Day code	Qualitative	12		Type de jour dans la semaine
Weeks	Qualitative	42		Semaine de l'année
Drill	Qualitative	95		Type d'exercices réalisés
Type	Qualitative	2		Identifie entraînements et compétitions
Date	Qualitative	178		Date calendaire
Duration	Quantitative			Durée de l'effort
Distance	Quantitative			Distance totale parcourue
Distance Zone X	Quantitative		6	Distance parcourue à zone de vitesse spécifique
LSR	Quantitative			Distance parcourue à vitesse faible
Marche	Quantitative			Distance parcourue à la marche
HSR	Quantitative			Distance parcourue au dessus seuil haute vitesse
VHSR	Quantitative			Distance parcourue au dessus seuil très haute vitesse
Max speed	Quantitative			Vitesse maximale atteinte
M.min	Quantitative			Vitesse relative
Sprint count	Quantitative			Nb de sprint réalisés
Accélération > 2m/s ²	Quantitative			Nb accélération au dessus seuil
Accélération Zone X	Quantitative		3	Nb d'accélération dans zones définies
Décélération > 2m/s ²	Quantitative			Nb de décélération au dessus seuil
Décélération Zone X	Quantitative		3	Nb de décélération dans zones définies

