



UNIVERSITÀ DEGLI STUDI DI ROMA "FORO ITALICO"

Department of Movement, Human and Health Sciences

PhD Program in Human Movement and Sport Sciences Cycle XXXVIII – 2022-2025

Thesis Title:

"Methodological aspects of monitoring high intensity actions through global positioning systems in football"

PhD candidate: Christian Di Claudio

Supervisor: Laura Capranica

Co-Supervisors: Daniele Conte, Maurizio Fanchini

End of Thesis Submission: 9 January 2026

PhD Jury

Members

Prof. Cristina Cortis, Università Degli Studi di Cassino, Cassino, Italy.

Prof. Giancarlo Condello, Università di Parma, Parma, Italy.

Tutor

Prof. Laura Capranica, University of Rome “Foro italico”, Rome, Italy.

Co-Tutors

Prof. Daniele Conte, University of Rome “Foro italico”, Rome, Italy.

Maurizio Fanchini, PhD, “University of Verona”, Verona, Italy.

Acknowledgements

This PhD journey has been an extraordinary experience of learning, growth, and discovery, both personally and professionally. It would not have been possible without the support, guidance, and encouragement of many people to whom I would like to express my deepest gratitude.

First and foremost, I would like to express my profound appreciation to my supervisor, **Professor Laura Capranica**, for her invaluable mentorship, trust, and continuous support throughout my doctoral studies. Her expertise, insightful feedback, and encouragement have been fundamental in shaping not only this research project but also my development as a researcher and professional. Her ability to combine scientific rigor with genuine human understanding has been a constant source of inspiration.

I am equally grateful to my co-supervisor, **Professor Daniele Conte**, for his constant availability, constructive discussions, and invaluable guidance during every phase of this work. His meticulous approach to research and his passion for sport science have greatly influenced my academic journey. I truly appreciate the time and effort he has dedicated to helping me refine my ideas and improve the quality of my work.

My sincere thanks also go to **Dr. Maurizio Fanchini**, for his collaboration, scientific insight, and for providing me with the opportunity to carry out this research in such a stimulating environment. His professional expertise, openness to discussion, and trust have been essential in completing this project successfully.

Beyond academia, I would like to express my heartfelt gratitude to my **family**, for their endless love, patience, and unwavering support throughout these years. Their encouragement has been the foundation that sustained me through the most challenging moments of this journey.

Finally, my deepest and most special thanks go to my **partner, Giulia**, for her understanding, patience, and unconditional support. Her constant presence, encouragement, and belief in me have been invaluable. She has been my greatest source of motivation and serenity, and I dedicate a large part of this achievement to her.

To all of you, thank for being part of this path and for helping me reaching such an important milestone.

Table of Contents

Chapter 1: General Introduction

1.1	Introduction.....	6
1.2	Overview and Research Questions.....	8
1.3	Approaches, Subjects, and Experimental Protocols.....	8
1.3.1	Experimental Approaches.....	8
1.3.2	Brief Summary of Subjects, Study Design and Experimental Protocols.....	9
1.4	Statistical Analysis.....	10
1.5	References.....	12

Chapter 2: Systematic Review of the Literature on Football

2.1	Overview	15
2.2	Title Paper 1: “Validation process of GPS and IMU to assess high intensity accelerations and decelerations in football: a methodological systematic review”	15
2.3	Abstract	16
2.4	Introduction	17
2.5	Methods.....	19
2.6	Results.....	21
2.7	Discussion.....	24
2.8	Conclusions.....	27
2.9	References.....	29

Chapter 3: Understanding inter unit reliability of Catapult Vector S7 in controlled situations

3.1	Overview.....	35
3.2	Title Paper 2: “Inter-unit reliability of Catapult Vector S7 Device GPS and IMU Metrics for Simulated Football-Specific Drills”	35
3.3	Abstract.....	36
3.4	Introduction.....	37
3.5	Methods.....	38
3.6	Results.....	40
3.7	Discussion.....	43
3.8	Practical application	45
3.9	Conclusions.....	45
3.10	References	48

Chapter 4: Assessment inter unit reliability of Catapult Vector S7 during game situations

4.1	Overview.....	53
4.2	Title Paper 3: “Inter-unit reliability of Catapult Vector S7 Device GPS and IMU Metrics for Simulated Football-Specific Drills”	53
4.3	Abstract.....	54
4.4	Introduction.....	55
4.5	Methods.....	56
4.6	Results.....	57
4.7	Discussion.....	60
4.8	Practical application.....	62
4.9	Conclusions.....	62
4.10	References	64

Chapter 5: General Discussion and Conclusion

5.1	General Discussion.....	67
5.2	Practical implications for football performance monitoring.....	69
5.3	General limitations of the study.....	70
5.4	Conclusions and future perspective.....	71
5.5	References.....	72
5.6	List of submitted publications.....	73

Chapter 1: General Introduction

1.1 Introduction

Football is characterized by intermittent and predominantly multidirectional activity patterns (Bangsbo et al., 1994; Bloomfield et al., 2007; Shephard et al., 1999; Scott et al., 2016), requiring players to repeatedly perform high-intensity actions such as accelerations (ACC), decelerations (DEC), and changes of direction (COD). These movements impose substantial physical demands and can influence overall performance levels (Russell et al., 2016; Beato and Drust, 2021). Acceleration can be defined as the rate of change in velocity over time, where positive acceleration indicates an increase, and negative acceleration a decrease in velocity.

The ability to efficiently and rapidly modulate speed and direction through ACC and DEC is a key determinant of successful football performance (Delaney et al., 2018). Given the high mechanical and metabolic loads repeatedly imposed on players during match play (Martín-García et al., 2018; Riboli et al., 2020), it is essential to design training programs specifically aimed at preparing athletes to sustain these high-intensity efforts. Consequently, a growing body of research in team sport performance has proposed various parameters to evaluate and monitor high-intensity acceleration and deceleration (HIAD) metrics during training and competition (Delaney et al., 2018; Harper et al., 2019).

A higher frequency of HIADs has been associated with increased neuromuscular fatigue and muscle damage, as evidenced by elevated post-match creatine kinase levels (Harper et al., 2019; Gastin et al., 2019; de Hoyo et al., 2016). These effects arise because ACC and DEC impose greater metabolic and mechanical demands with respect to constant-speed running (Osgnach et al., 2010; Hader et al., 2016; Jaspers et al., 2018). Moreover, braking actions (DEC) are particularly demanding due to the greater eccentric loads applied to the musculoskeletal system, especially at high intensities. If these loads are not adequately absorbed, they may overstress soft tissues and impair neuromuscular performance (Harper et al., 2019).

Accordingly, accurate quantification and monitoring of ACC and DEC during both training and competition are fundamental for prescribing appropriate training protocols that prepare athletes to tolerate such loads (Delaney et al., 2018). Recent advances in monitoring technologies have enabled detailed assessments of external training load indicators (Malone et al., 2017). Among these, ACC and DEC are commonly measured using Global Positioning System (GPS) technology integrated within Inertial Measurement Unit (IMU) devices (Akenhead and Nassis, 2016; Cummins et al., 2013). Despite their widespread use, robust validation of these measurements remains limited. Establishing structured validation protocols is essential, as all HIAD metrics and tools must meet quality control criteria—specifically validity, reliability, and responsiveness—before being implemented in sport science or research contexts (Impellizzeri et al., 2009).

In particular, validity refers to the extent to which an instrument or test accurately measures the construct it is intended to assess (Atkinson and Nevill, 1998; Hopkins, 2000). It can be evaluated

through several dimensions—content, criterion, and construct validity—each requiring specific methodological approaches. Content validity concerns the degree to which the elements of an instrument comprehensively and appropriately represent the targeted construct in terms of clarity and relevance. Criterion validity is determined through correlation with an established “gold standard” measure. In the absence of such a reference, construct validity examines whether the instrument yields result consistent with theoretical expectations or existing empirical evidence (Mokkink et al., 2010; Koo and Li, 2016). The evaluation of validity involves a range of statistical indices, including correlation coefficients, standard error of measurement (SEM), standard error of the estimate (SEE), typical error (TE; often expressed as the coefficient of variation, CV), root mean square error (RMSE), and mean absolute error (MAE; also expressed as a percentage) (Scott et al., 2016).

Reliability, defined as “the degree to which a measurement is free from random error,” can be classified as absolute or relative (Atkinson and Nevill, 1998; Hopkins, 2000; Koo and Li, 2016). Absolute reliability, also referred to as measurement error or agreement, represents the within-subject variability observed across repeated trials and is particularly useful in longitudinal monitoring. Relative reliability, on the other hand, reflects the consistency of individuals’ rankings within a sample across repeated measurements, serving primarily a discriminative purpose (Mokkink et al., 2010). However, many reliability studies fail to clearly distinguish between these two dimensions.

When assessing the performance of GPS and IMU technologies, both intra-unit and inter-unit reliability should be considered. Intra-unit reliability provides insights into the consistency of a single device over time, which is critical for monitoring individual performance fluctuations. Inter-unit reliability becomes essential when comparing data obtained from multiple devices, as it determines whether units can be used interchangeably (Scott et al., 2016). The evaluation of reliability also relies on specific statistical tools: absolute reliability is commonly assessed using SEM or typical error of measurement (TEM), limits of agreement (Bland–Altman plots), and CV expressed as a percentage; relative reliability is typically determined using intraclass correlation coefficients (ICC) (Hopkins, 2000; Koo and Li, 2016).

Responsiveness represents another key aspect of the validation process and is defined as “the ability of an instrument to detect meaningful change over time in the construct being measured” (de Vet et al., 2011). This property is crucial for tracking variations in training and match loads over specific periods. Several methods have been proposed to quantify responsiveness, including Cohen’s effect size, standardized response mean, Guyatt’s responsiveness ratio, and the area under the receiver operating characteristic (ROC) curve (Atkinson and Nevill, 1998; Hopkins, 2000).

Recent systematic reviews have synthesized available evidence on the validation of GPS- and IMU-based metrics, examining their validity and reliability across different sport contexts (Scott et al., 2016; Crang et al., 2021; Dawson et al., 2023). The reviews by Scott and Crang focused on the capacity of wearable microtechnologies to quantify movement characteristics and sport-specific actions in intermittent team sports. Dawson and colleagues (2023) further investigated the validity and reliability of accelerometer-derived metrics from upper-back-mounted GNSS systems, with a focus on estimating external biomechanical load through instantaneous accelerations. Collectively,

these reviews assessed the psychometric properties of the parameters but did not systematically evaluate the methodological quality of the included studies.

1.2 Overview and Research Questions

Global Positioning System technology has been examined in different studies, however brand and device as well as sample characteristics, may provide differences in reliability and validity of the technology and derived metrics. The overarching objective of this research program is to enhance methodological understanding of how currently available technologies can be effectively employed to monitor and optimize athletic performance. To accurately interpret and apply the large volume of data collected during training sessions and competitive matches, it is essential to establish a sound comprehension of the correct use, strengths, and limitations of the technological systems at our disposal—particularly Global Positioning System (GPS) technology.

In light of this, the first step (Chapter 2) involves conducting a systematic review of the most recent literature—building upon and updating the last comprehensive reviews by Scott et al. (2010) and Crang et al. (2020)—to identify the current state of research and the existing methodological gaps in this field. Subsequently, the following two studies (Chapters 3 and 4) will evaluate the inter-unit reliability of HIAD metrics under different experimental protocols, including GPS- and IMU-derived variables such as Total Distance, Maximal Speed, High-Intensity Running Distance, ACC, DEC, Metabolic Power, and PlayerLoad obtained through GPS devices.

The assessment of these parameters is expected to represent an important step toward improving the interpretation of monitoring data in applied sport contexts. Moreover, it will provide a more robust foundation for data-driven decision-making aimed at manipulating training variables while minimizing potential sources of bias.

1.3 Approaches, Subjects, and Experimental Protocols

A brief description of the methods used in this research project is provided in this section.

1.3.1 Experimental Approaches

This research project aims to establish the importance of assessing the validity, reliability, and responsiveness of measures commonly used in practice to describe high-intensity acceleration and deceleration (HIAD) metrics. By providing robust evidence regarding these measurement properties, the study seeks to enhance the accuracy and interpretability of HIAD-related data in applied sport contexts. The findings are expected to support the monitoring of high-intensity demands within specific training methodologies, thereby improving the quality and precision of practitioners' day-to-day decision-making processes. Furthermore, the outcomes of this research may assist coaches in ensuring that athletes are adequately prepared to meet the physical requirements of competition. Additionally, the results could provide valuable insights into the effectiveness of specific training drills designed to develop high-intensity capacities, such as total distance, speed threshold-based

distances covered between 15–20 km/h and 20–25 km/h, maximal speed, metabolic power, metabolic power per kg, equivalent distance, equivalent distance index (calculated as the ratio between equivalent distance and total distance), acceleration ($>2.5 \text{ m/s}^2$) and deceleration ($>2.5 \text{ m/s}^2$) metrics (e.g., frequency of occurrence, distance, density) and IMU-derived metrics (e.g., PlayerLoad, 2D PlayerLoad, and PlayerLoad per meter).

Prior to data collection, a proposal has been submitted to the Institutional Review Board (IRB) of the University of Rome “Foro Italico” (CAR, 212/2024). Upon approval, potential participants have received an information letter outlining the study’s aims, data use, voluntary participation, and their right to withdraw at any time without providing a reason. Written informed consent has been obtained from all participants. For individuals under 18 years of age, parental or guardian consent has been required. In accordance with the General Data Protection Regulation (GDPR), all collected data were anonymized, and analyses were conducted solely for the purposes of the PhD project and its outcomes.

1.3.2 Brief Summary of Subjects, Study Design and Experimental Protocols

The first study, “Validation process of GPS and IMU to assess high-intensity accelerations and decelerations in football: a methodological systematic review”, submitted to “European Journal of Sport Science (EJSS)” and presented in Chapter 2, aimed to systematically review the literature on the methodological quality and statistical approaches used to evaluate the validity, reliability, and responsiveness of GPS and IMU metrics. The focus was placed on high-intensity activities that characterize the physical demands of football training and match play. This information is crucial for coaches and sport scientists to ensure that each metric has been adequately verified through a rigorous validation process, thus ensuring its trustworthiness (Impellizzeri et al., 2009).

The second study, “Inter-unit reliability of Catapult Vector S7 Device GPS and IMU Metrics for Simulated Football-Specific Drills”, submitted to “European Journal of Sport Science (EJSS)” and presented in Chapter 3, aimed to investigate the inter-unit reliability of GPS- and IMU-derived metrics under controlled locomotor conditions of continuous, accelerated, and decelerated movements simulating the typical football activity (Gualtieri et al., 2023), while considering signal quality during data collection. A total of four Catapult Vector S7 devices (Catapult Sports, Melbourne, Australia) were randomly selected from a pool of 25 units and mounted on the vertical poles of an electric golf cart. The electric golf cart was driven reproducing three different movement patterns (drills) designed to simulate football-specific locomotor demands (Dawson et al., 2024, Crang et al., 2023): (1) Drill A – continuous running (velocity range: 0–24.6 km/h); (2) Drill B – straight line-and-back running (velocity range: 0–22.5 km/h); and (3) Drill C – intermittent accelerations and decelerations (velocity range: 0–21.3 km/h). These findings may assist practitioners in improving the accuracy and reliability of performance monitoring, thereby enhancing training load management. The third and final study, “Inter-unit reliability of Catapult S7 GPS during different football training sessions”, submitted to “Journal of Sport Sciences (JSS)” and presented in Chapter 4, builds upon previous investigations on the reliability of IMU-derived metrics collected under ecological, sport-

specific conditions. The purpose of this study was to evaluate the reliability of HIAD metrics obtained from Catapult Vector S7 GPS devices during real football training sessions, introducing new considerations regarding signal quality during data collection. These findings are expected to offer valuable insights to sport scientists and practitioners when interpreting differences in external load between players during matches and training. Three elite youth players from an Italian Serie A U18 academy team volunteered to participate in this study. The research adopted an observational design in which players wore two GPS units during 33 regular football training sessions. All methodological details are provided in Chapter 4.

1.4 Statistical Analysis

The systematic review was conducted following the guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 (PRISMA 2020) and the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) (Page et al., 2020; Mokkink et al., 2010). The complete review protocol was preregistered in the Open Science Framework in May 2023 (<https://doi.org/10.17605/OSF.IO/DC6W9>).

In the second and third studies, the Shapiro–Wilk test was used to assess data normality. The Intraclass Correlation Coefficient (ICC) and Typical Error (TE), expressed as the coefficient of variation (%CV), were calculated to assess relative and absolute inter-unit reliability, respectively. Whilst ICC was used to evaluate ranking consistency between units, TE was used to determine variability between units based on the standard deviation of the differences between measurements. Uncertainty was expressed as 90% confidence limits (CL) for both ICC and TE. All reliability calculations were performed using Hopkins' reliability spreadsheet (Hopkins, 2009). ICC values were interpreted as excellent (>0.90), good (0.75–0.90), moderate (0.50–0.75), and poor (<0.50) (Koo and Li, 2016).

In the second study, a one-way analysis of variance (ANOVA) was performed with Jamovi (Jamovi 2.5.4.0) considering each GPS and IMU derived metric as the dependent variable and device position as the independent variable. When significant main effects were found, Tukey post hoc tests were conducted to identify pairwise differences between units. Statistical significance was set at $p < 0.05$. In the third study, linear mixed models (LMM) were subsequently employed to examine the significance of differences between paired units. LMM analyses were performed using Jamovi software (version 2.3.21; available at <https://www.jamovi.org>). The alpha level for statistical significance was set a priori at $p < 0.05$.

All further details regarding statistical analyses are provided in the respective chapters for each study (Chapters 2, 3, and 4).

1.5 References

1. Bangsbo J. 1994. The physiology of soccer with special reference to intense intermittent exercise. *Acta Physiol Scand.* 151(619):1–155.
2. Shephard RJ. 1999. Biology and medicine of soccer: An update. *J Sports Sci.* 17:757–786.
3. Bloomfield J, Polman R, O'Donoghue P. 2007. Physical demands of different positions in FA Premier League soccer. *J Sports Sci Med.* 6:63–70.
4. Scott A, Malone JJ, Morgans R, Burgess D, Gregson W, Morton JP, Drust B. 2016. The relationship between physical match performance and 48-h post-game creatine kinase concentrations in English Premier League soccer players. *Int J Sports Sci Coach.* 11(6):846–852.
5. Russell M, Sparkes W, Northeast J, Cook CJ, Love TD, Bracken RM, et al. 2016. Changes in acceleration and deceleration capacity throughout professional soccer match-play. *J Strength Cond Res.* 30(10):2839–2844.
6. Beato M, Drust B. 2021. Acceleration intensity is an important contributor to the external and internal training load demands of repeated sprint exercises in soccer players. *Res Sports Med.* 29(1):67–76.
7. Delaney JA, Cummins CJ, Thornton HR, Duthie GM. 2018. Importance, reliability, and usefulness of acceleration measures in team sports. *J Strength Cond Res.* 32(12):3485–93.
8. Martín-García A, Casamichana D, Gómez Díaz A, Cos F, Gabbett TJ. 2018. Positional differences in the most demanding passages of play in football competition. *J Sports Sci Med.* 17(4):563–570.
9. Harper DJ, Carling C, Kiely J. 2019. High-intensity acceleration and deceleration demands in elite team sports competitive match play: A systematic review and meta-analysis of observational studies. *Sports Med.* 49(12):1923–47.
10. Riboli A, Semeria M, Coratella G, Esposito F. 2020. Effect of formation, ball in play and ball possession on peak demands in elite soccer. *Biol Sport.* 38(2):195–205.
11. Gastin PB, Hunkin SL, Fahrner B, Robertson S. 2019. Deceleration, acceleration, and impacts are strong contributors to muscle damage in professional Australian Football. *J Strength Cond Res.* 33(12):3374–83.
12. de Hoyo M, Cohen DD, Sañudo B, Carrasco L, Álvarez-Mesa A, Del Ojo JJ, Domínguez-Cobo S, Mañas V, Otero- Esquina C. 2016. Influence of football match time-motion parameters on recovery time course of muscle damage and jump ability. *J Sports Sci.* 34(14):1363-70.
13. Hader K, Mendez-Villanueva A, Palazzi D, Ahmaidi S, Buchheit M. 2016. Metabolic power requirement of change of direction speed in young soccer players: Not all is what it seems. *PLoS ONE.* 11(3):1–21.
14. Jaspers A, Kuyvenhoven JP, Staes F, Frencken WGP, Helsen WF, Brink MS. 2018. Examination of the external and internal load indicators' association with overuse injuries in professional soccer players. *J Sci Med Sport.* 21(6):579–585.
15. Malone JJ, Lovell R, Varley MC, Coutts AJ. 2017. Unpacking the black box: Applications and considerations for using GPS devices in sport. *Int J Sports Physiol Perform.* 12 (S2):18–26.

16. Akenhead R, Nassis GP. 2016. Training load and player monitoring in high-level football: current practice and perceptions. *Int J Sports Physiol Perform.* 11(5):587–593.
17. Cummins C, Orr R, O’Connor H, West C. 2013. Global positioning systems (GPS) and microtechnology sensors in team sports: A systematic review. *Sports Med.* 43(10):1025–1042.
18. Akenhead R, Hayes PR, Thompson KG, French D. (2013). Diminutions of acceleration and deceleration output during professional football match play. *J Sci Med Sport.* 16(6): 556-561.
19. Osgnach C, Poser S, Bernardini R, Rinaldo R, di Prampero PE. 2010. Energy cost and metabolic power in elite soccer: A new match analysis approach. *Med Sci Sports Exerc.* 42(1):170–8.
20. Impellizzeri FM, Marcora SM. 2009. Test validation in sport physiology: Lessons learned from clinimetrics. *Int J Sports Physiol Perform.* 4(2):269-77.
21. Atkinson G. 2002. Sport performance: Variable or construct? *J Sports Sci.* 20:291– 292.
22. Hopkins WG, Hawley JA, Burke LM. 1999. Design and analysis of research on sport performance enhancement. *Med Sci Sports Exerc.* 31(3):472-85.
23. Mokkink LB, Terwee CB, Patrick DL, et al. 2010. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *J Clin Epidemiol.* 63:737–45.
24. Koo TK, Li MY. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 15(2):155-63. Erratum in: *J Chiropr Med.* 2017. 16(4):346.
25. Mokkink LB, Terwee CB, Patrick DL, et al. 2010. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *J Clin Epidemiol.* 63:737–45.
26. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. 2012. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 21(4):651–657.
27. Dawson L, Beato M, Devereux G, McErlain-Naylor SA. 2024. A Review of the validity and reliability of accelerometer-based metrics from upper back-mounted GNSS player tracking systems for athlete training load monitoring. *J Strength Cond Res.*
28. Crang ZL, Duthie G, Cole MH, Weakley J, Hewitt A, Johnston RD. 2021. The validity and reliability of wearable microtechnology for intermittent team sports: A systematic review. *Sports Med.* 51(3):549-565.
29. Gualtieri A, Rampinini E, Dello Iacono A, Beato M. 2023. High-speed running and sprinting in professional adult soccer: Current thresholds definition, match demands and training strategies. A systematic review. *Front Sports Act Living.* Feb 13;5:1116293.
30. Page MJ, McKenzie JE, Bossuyt PM, et al. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372:n71
31. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual Life Res.* 2010;19(4):539–549
32. Hopkins W, Marshall S, Batterham A, Hanin J. Progressive statistics for studies in sports medicine and exercise science

Chapter 2: Systematic Review of the Literature on Football

2.1 Overview

This chapter presents a systematic review examining the methodological quality and statistical approaches employed to assess the validity, reliability, and responsiveness of Global Positioning System (GPS) and Inertial Measurement Unit (IMU) metrics used to quantify high-intensity accelerations and decelerations (HIAD) in football. Given the increasing reliance on wearable tracking technologies for monitoring players' external load, understanding the robustness and methodological rigor behind these measurements is essential. Following PRISMA 2020 and COSMIN guidelines, twenty-two studies were included and evaluated through a multi-step process encompassing risk of bias, measurement quality criteria, and evidence grading using the GRADE approach.

The findings revealed a scarcity of high-quality studies, particularly concerning construct validity and responsiveness, with the majority of reliability assessments showing doubtful to moderate methodological quality. Overall, the review highlights critical methodological shortcomings in current validation research and emphasizes the need for standardized protocols and rigorous testing procedures to ensure accurate use of wearable technologies in football performance monitoring. These insights provide a foundation for the subsequent empirical investigations presented in the following chapters.

2.2 Title Paper 1

Validation process of GPS and IMU to assess high intensity accelerations and decelerations in football: a methodological systematic review"

Christian Di Claudio ^{1,2}, Maurizio Fanchini ^{2,3}, Carwyn Sharp ^{2,4}, Laura Capranica ¹, Daniele Conte ^{1,5}

1 Department of Movement, Human and Health Sciences, University of Rome "Foro Italico", Rome, Italy

2 AS Roma Football Club, Rome, Italy

3 University of Verona, Italy

4 Department of Kinesiology and Sport Studies, University of Denver, Denver, U.S.A.

5 Department of Coaching Science, Lithuanian Sports University, Kaunas, Lithuania

(submitted to "European Journal of Sport Science")

2.3 Abstract

Background: Football encompasses intermittent activities requiring high-intensity accelerations and decelerations (HIAD) and changes of direction, which may impact the players' performance.

Global Positioning System (GPS) devices packed with Inertial measurement units (IMU) and used to monitor the football activities, however little information on a sound validation of HIAD is available.

Objective: The present study aims to investigate the methodological aspects used to assess validity, reliability and responsiveness of HIAD metrics collected by GPS and IMU technology.

Methods: A systematic search of scientific articles published from January 1990 to March 2023 was conducted in three databases (PubMed, Web of Science and SPORTDiscus). Inclusion criteria were: manuscripts in English reporting information regarding the validity, reliability, and responsiveness of GPS and IMU of HIAD metrics. Following the COSMIN guidelines, the methodological quality of the HIAD metrics was assessed through a 4-point scale consisting of "very good", "adequate", "doubtful", and "inadequate".

Results: Out of 152 studies emerging from the electronic search, 15 met the inclusion criteria. An additional 6 studies were identified through a snowballing technique, resulting in a total of 21 studies. Whilst construct validity resulted in 100% "very good", criterion validity was found to be 83.3% "very good" and 16.7% "doubtful", respectively. Relative reliability resulted in 11.8% "adequate", 76.5% "doubtful", and 11.8% "inadequate", whereas absolute reliability was found to be 30% "adequate", 60% "doubtful", and 10% "inadequate", respectively. No studies investigated responsiveness.

Conclusions: In general, limitations related to methodological experimental procedures and statistical approaches adversely affected the methodological quality of the studies and the quality of evidence of the HIAD metrics. The inconsistent methodological quality concerning reliability studies warrants caution among practitioners when interpreting HIAD metrics for monitoring purposes. Therefore, further investigation is needed to ensure a high level of evidence validation process of HIAD metrics in football.

Registration: The protocol of this systematic review was registered at the Open Science Framework (<https://doi.org/10.17605/OSF.IO/DC6W9>)

Keywords:

validity, reliability, responsiveness, high intensity, external load, monitoring, soccer players

2.4 Introduction

Football is characterized by high-intensity intermittent activities encompassing sprints, accelerations, decelerations and changes of direction during matches and training sessions [1, 2, 3]. To optimize the training planning and periodization, coaching staffs are used to monitor the activity done by the players (i.e. external load) with wearable systems. The global positioning system technology (GPS) is the most used tracking technology and it assesses the player's position through connection with satellites using a sampling rate between 5-20 Hz, depending on the model [21]. The accuracy of GPS devices is highly dependent on the signal quality, specifically the horizontal dilution of precision and the number of connected satellites [13, 19, 27]. Within the plethora of variables provided by different GPS brands, football coaches and practitioners consider the following as key performance indicators: the total distance travelled, the total distance covered in relation to different speed thresholds, the maximal speed, the number of instantaneous accelerations and decelerations at different intensities, and the metabolic power [22, 23]. Additionally, GPS are usually embedded with inertial measurement units (IMUs), which can provide further information about players' physical demands (13, 18, 19, 20). Independently, from GPS signal quality, IMUs collect data using their internal accelerometer, magnetometer, and gyroscope [24, 25], thus allowing the measurement of players' locomotor activities also in indoor environments [18]. The main IMU-derived variables analyzed by football practitioners are peak inertial accelerations and decelerations, changes of directions, impacts and Playerload (i.e. a modified vector magnitude determined as the square root of the sum of the squared instantaneous rate of change in acceleration across three motion planes) [20, 26].

Considering the large number of variables provided by GPS and IMU devices, football coaches could have a comprehensive snapshot of the physical activities performed by players during matches and training. However, each metric needs to provide evidence of validity, reliability, and responsiveness before considering it in a specific sport and/or research contexts [31]. In particular, validity defines the degree to which the tool or test measures what it purposes to measure [28, 30] and it could be assessed by examining its content, criterion and/or construct validity by means of appropriate statistical approaches. In particular, content validity relates to the construct corresponding to the contents of the instrument, both in terms of compressibility and relevance; criterion validity is assessed by correlation with a gold standard for the construct to be measured; whereas in absence of a gold standard, construct validity refers to whether the instrument provides the expected scores, based on existing knowledge about a construct [31, 32]. The assessment of different types of validity require specific statistical approaches, with a wide heterogeneity of statistical parameters encompassing correlation coefficients, standard error of the measurements (SEM), standard estimate of the error (SEE), typical error (TE, also expressed as the coefficient of variation), root mean square error (RMSE), and mean absolute error (MAE, also expressed as a percentage) [34, 35]. Reliability is defined as "the degree to which the measurement is free from measurement error" and it can be classified as absolute and relative reliability [30, 36, 39]. Specifically, absolute reliability (i.e., agreement or measurement error) is the degree of variation of individuals during repeated measurements and it is useful to evaluate changes in longitudinal assessments, whereas relative reliability is the extent to which the position of the individuals in a sampling is maintained with

repeated measurements and it is useful for its discriminative property [31]. Nevertheless, reliability studies rarely differentiate between the two reliability properties [31]. In addition, GPS and IMU technologies should be evaluated considering both intra- and inter-unit reliability. Whilst intra-unit reliability could provide information on substantial daily or weekly performance variations of an athlete, it is important to consider the inter-unit reliability when the measurements of numerous devices are used to assess differences between athletes [35]. The assessment of different types of reliability requires specific statistical approaches, with absolute reliability that can be examined with standard error of measurement (SEM) or typical error of measurement (TEM), limits of agreement (i.e. Bland and Altman methods) and coefficient of variation (CV) expressed as a percentage [32], whereas relative reliability can be assessed using intraclass correlation coefficients (ICCs) [32, 37, 38].

Responsiveness is another important parameter in the validation process and has been defined as “the ability of an instrument to detect change over time in the construct to be measured” (41). Responsiveness plays a key role in monitoring the changes in match and training loads during a specific period. Over the last few decades, numerous statistical methods have been suggested for evaluating responsiveness, (42) including the Cohen’s effect size, the standardized response mean, Guyatt’s responsiveness ratio and the area under the curve examined with the receiver operating characteristics analysis (32).

Recently, systematic reviews have synthesized the results of studies focused on the validation process of GPS and IMU metrics and their validity and reliability in different sport contexts (34, 35, 43). The studies of Scott and Crang (34, 35) examined the validity and/or reliability of wearable microtechnology to quantify movement and specific actions common to intermittent team sports. The recent work of Dawson et colleagues (43) examined the validity and reliability of the accelerometer used in upper back-mounted GNSS player tracking systems, and going in deep on accelerometer-based metrics estimating external biomechanical load from the accumulation of instantaneous accelerations. All three reviews examined the of validity and reliability of the parameters without examining the methodological quality.

Nevertheless, to the best of our knowledge, there is a lack of evidence examining and synthesizing the methodological quality and the appropriateness of the statistical methods adopted in the studies focused on validity, reliability and responsiveness of GPS and IMU metrics in football

Therefore, the present study aimed to systematically review the literature on the methodological quality and statistical methods used to evaluate the validity, reliability, and responsiveness of GPS and IMU metrics focusing on high-intensity activities characterizing the physical demand of training and matches in football. This information is crucial for football coaches to provide evidence that each metric has been adequately verified through a quality validation process (31) and therefore trustable.

2.5 Methods

2.5.1 Study Design and Registration

The present systematic review has been performed following the guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-analyses 2020 (PRISMA 2020), and Consensus-based Standards for the selection of health Measurement Instruments (COSMIN: <http://www.cosmin.nl>) (44, 46). The complete protocol for this systematic review was preregistered in the Open Science Framework in May 2023 (<https://doi.org/10.17605/OSF.IO/DC6W9>).

2.5.2 Search strategy

The relevant literature was identified systematically by reviewing three electronic databases [i.e., PubMed (MEDLINE), Web of Science, and SPORTDiscus (EBSCO)] in the period between January 1990 and March 2023. Studies were identified by searching abstracts, titles, and keywords across the three databases and using the following words and their combinations: (reliability OR validity OR responsiveness) AND (soccer OR football OR football association) AND (global positioning system* OR gps OR global navigation satellite system OR microtechnology OR wearable OR neuromuscular OR acceleration* OR deceleration* OR mems OR micro-electrical mechanical system OR inertial measurement units OR imu). To ensure the inclusion of the most updated articles, alert notifications of new publications were activated until December 2024. In addition, the snowballing technique was applied to the reference lists of the included articles to identify any possible missed contribution.

2.5.3 Eligibility criteria

The following inclusion criteria were used for the selection of the studies: full-text articles written in English; the assessment of the typology of validity and/or reliability and/or responsiveness of wearable microtechnology (GPS and IMU device) to quantify high-intensity accelerations and decelerations metrics (HIAD) common to football; studies focusing on HIAD derived from GPS (i.e. instantaneous accelerations count, decelerations count, distance in meters covered during instantaneous accelerations and decelerations, metabolic power and derived metrics) and IMU (i.e. player load and derived metrics, peak acceleration and deceleration, angular and force metrics derived from inertial accelerometer) devices. Validity, reliability and responsiveness were defined accordingly to previous research (31, 34, 35). Studies that did not meet the inclusion criteria were removed from the list (supplementary material 1).

2.5.4 Study selection

Titles, abstracts and full texts were screened independently by two raters (CDC and DC) using the Endnote X9.3.3 (Clavirate, Philadelphia, USA) and Rayyan software (Online version 1.4.3) (45). Studies were included after agreement between the two raters, and in case of disagreements, a third author (MF) was involved for a final decision. Inter-rater reliability of the inclusion process between two authors (CDC and DC) was calculated as the percentage of agreement and Cohen's k (i.e., idostatistics.com) and considered as follows: 0.01 - 0.20 slight agreement, 0.21- 0.40 fair agreement,

0.41- 0.60 moderate agreement, 0.61 - 0.80 substantial agreement, and 0.81 - 1.00 almost perfect or perfect agreement.

2.5.6 Data extraction

Data extraction was performed by one author (CDC) with a second author (DC) available to support in case of doubts. A Microsoft Office Excel spreadsheet (2019 Version 16.78.3) specifically customized for measurement properties (i.e. validity, reliability, responsiveness) was used at this stage. The following data were extracted: first author, year of publication, sample size (i.e., number and level of athletes), GPS and IMU brand, signal quality during data collection, the analyzed physical load measures, type of movement and protocol, type of measurement property (i.e., validity, reliability, responsiveness) and statistical analyses performed.

2.5.7 Assessment of Methodological Quality and Data Analysis

According to the COSMIN guidelines, the included studies were examined through three steps. The first step is the evaluation of the methodological quality of the included studies by using the COSMIN Risk of Bias checklist. The second step involves applying criteria for good measurement properties using quality criteria. In the third step, the evidence is summarized and its quality graded using the GRADE approach.

The first step was performed by two authors (CDC and DC) independently, while the second and third steps were performed by one author (CDC) with other two authors (DC and MF) supporting to resolve any doubts. The COSMIN checklist has been used to provide a rating to each study(46, 47, 48, 49), with measurement properties divided into 3 domains: reliability, validity, and responsiveness (41). The COSMIN checklist contains different boxes and items that are filled out to assess if the included studies meet the standards for design and statistical analysis for each measurement property. Each item was rated with a 5-point scale as follows: NA (not applicable), inadequate, doubtful, adequate, very good. An overall score of each study was then determined by the “worst score counts” method.

In the second step, each study’s results regarding its measurement property were evaluated against the updated criteria for good measurement quality (39, 49). As suggested by Terwee et al. (2012), the outcomes of each study were rated as “sufficient” (+), “insufficient” (-), “inconsistent” (⊖), or “indeterminate” (?) (47). Subsequently, the results were synthesized for each metric with the Patient-Reported Outcome Measures (PROMs) according to the COSMIN guidelines, in order to assign an overall rating to each selected study: “sufficient” (+), “insufficient” (-), “inconsistent” (±), or “indeterminate” (?). (48).

In the third step, the aim was to grade the quality of the evidence following the Grading of Recommendations Assessment, Development and Evaluation (i.e. GRADE) approach developed by the COSMIN methodology (49, 47). The GRADE approach defines five factors to complete the quality of evidence: risk of bias (i.e. methodological quality of the studies), inconsistency of the study’s results, indirectness (i.e. different population and interventions), imprecision (i.e. wide confidence

intervals), and publication bias. In using the approach for measurement property per PROM (i.e. metrics), the quality of the evidence was graded as “high”, “moderate”, “low”, or “very low” (48).

2.6 Results

2.6.1 Study selection

The systematic search retrieved a total of 152 studies, of which 33 were duplicates. After screening titles, abstracts and full texts, 15 studies met the inclusion criteria (Fig.1). In addition, 7 studies were identified and included after checking the reference list of each previously included study. A total of 22 studies were included for the systematic review. The two authors (CDC and DC) demonstrated a fair agreement for title, abstract and full text screening (percentage agreement 77.5%; Cohen’s $k = 0.29$), with the third author (MF) solving all disagreements.

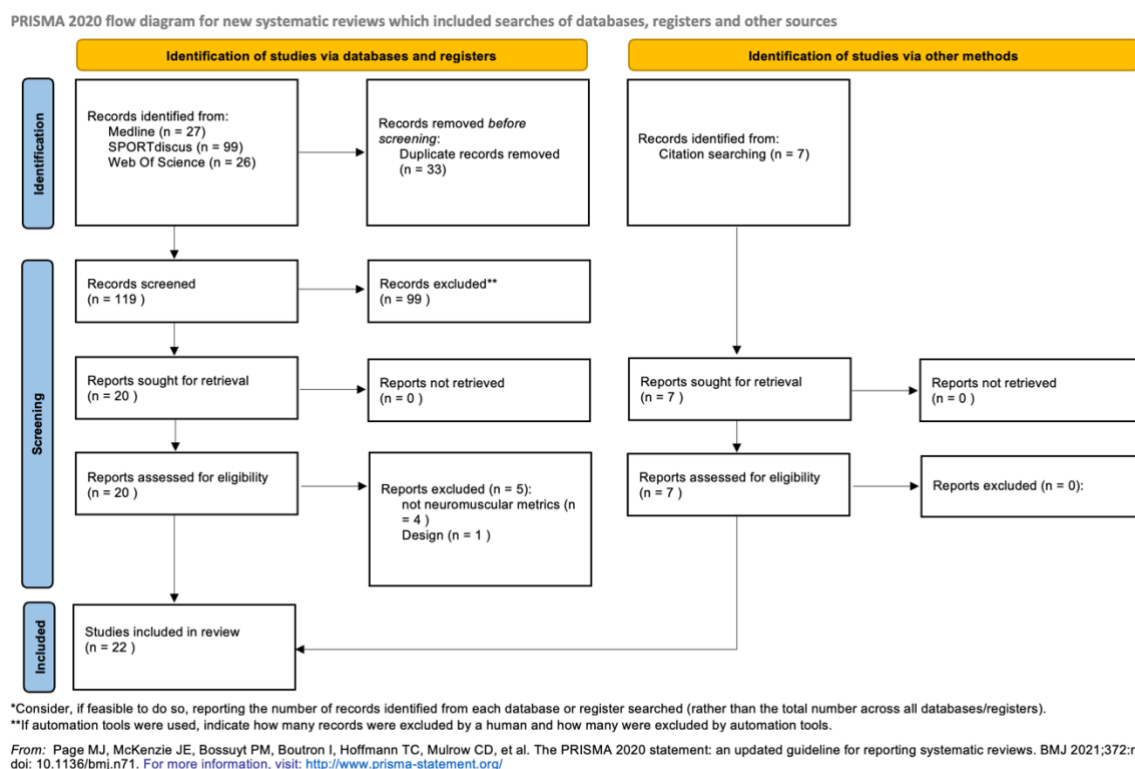


Figure.1. Prima flow chart for the selection process of the systematic literature review

2.6.2 Data extraction

The percentage of studies that investigated the validation process of HIAD metrics using GPS and IMU derived metrics were 58.3% and 41.3%, respectively. Out of the eight studies examining the validity, the majority focused on criterion validity and only 25% on construct validity. The statistical tests utilized to assess the validity of HIAD metrics were: Pearson’s correlation coefficient (50%), standardized differences expressed in percentage and Cohen’s d effect size (10%), typical error of estimate (TEE) expressed both in percentage and standardized units (20%), typical measurement

error expressed as TEM (10%), and linear mixed models (10%). Reliability represents the most frequently investigated measurement property, examined as relative (60.7%) and absolute reliability (39.3%). Absolute and relative reliability parameters were calculated with TE expressed as percentage of CV (14.3%), ICC (42.8%), and coefficient of variation CV (17.1%), Standard Error of Measure or SEM (14.3%), Bland and Altman plot (5.7%) and mean differences (2.9%). No studies investigated the responsiveness of the HIAD. The table of the data extraction is presented as supplementary material 2.

2.6.3 Assessment of Methodological Quality and Data Analysis

The methodological quality of the included studies and the items assessed are presented in Table 1. The first step of the methodological quality analysis revealed that the included studies ranged from “very good” to “inadequate”. All studies examining construct validity were rated of high quality (i.e., “very good”), whereas the majority of the studies focusing on criterion validity were rated as “very good” (83.3%) with 16.7% rated as “doubtful”. Studies examining relative reliability were mostly rated as “doubtful” (76.5%), “adequate” (11.8%), and “inadequate” (11.8%). Similarly, studies focusing on absolute reliability were rated as “doubtful” (60%), “adequate” (30%) and “inadequate” (10%) quality. Moderate agreement was shown in assessing methodological quality of studies between the two authors (CDC and DC) (agreement: 86.6%; Cohen’s $k = 0.42$). The results of the updated criteria for good measurement properties and the quality of the evidence (i.e. second and third step following COSIMN guidelines) are presented in the table 2.

The second step revealed that the outcomes of each study on criterion validity were rated 22.2% sufficient, 33.3% insufficient and 44.4% inconsistent. Construct validity showed ratings 100% insufficient. Studies that examined reliability were rated 52.4%, sufficient, 28.6% insufficient, 4.8% inconsistent and 14.3% indeterminate. Studies providing measurement error showed 100% indeterminate results.

The grade approach, used in the third step, revealed that the criterion validity of the metrics were 11.1% low, 44.4% moderate and 44.4% of high quality evidence. Construct validity has been found 100% low quality evidence. The reliability of the metrics showed 19% high, 19% moderate, 19% low and 42.9% very low quality of evidence. The measurement error of the metrics, were found 23.4% high, 23.5% moderate, 11.8% low and 41.2% very low quality of evidence.

All details and descriptions on the data analysis for each metric can be found in supplementary material 3 and 4-

Table 1 Assessment of methodological quality of the studies

Reference	Criterion validity	Construct Validity	Reliability	Measurement Error	Responsiveness
Armitage et al., 2021 [52]			doubtful	doubtful	
Barrett et al., 2017 [51]	doubtful				
Barrett et al., 2016 [53]			doubtful	doubtful	
Buchheit et al., 2015 [54]	very good		doubtful	adequate	
Burland et al., 2020 [55]			doubtful		
Hughes et al., 2019 [56]			doubtful	doubtful	
Meylan et al., 2016 [57]	very good		doubtful		
Waldron et al., 2021 [58]				doubtful	
Wilmes et al, 2023 [59]		very good	doubtful		
Wilmes et al., 2020 [60]	very good				
Aquino et al., 2020 [61]			doubtful	doubtful	
Beato et al., 2018 [62]			doubtful	doubtful	
Beato et al., 2017 [63]			doubtful	doubtful	
Bredt et al., 2016 [64]			inadequate	inadequate	
Dello Iacono et al., 2021 [65]			inadequate		
Barreira et al., 2016 [66]		very good	doubtful		
Buchheit et al., 2014 [67]	very good		doubtful		
Rampinini et al., 2015 [68]	very good				
Buchheit et al., 2018 [69]			adequate	adequate	
Sparkes et al., 2022 [70]			doubtful		
Fitzpatrick et al., 2019 [71]			adequate	adequate	
Beato et al., 2024 [72]			adequate	adequate	

Table 2 Updated criteria for good measurement properties and Quality of the Evidence for Measurement Properties of the metrics

Metrics	Criterion validity Overall rating / Quality of evidence	Construct Validity Overall rating / Quality of evidence	Reliability Overall rating / Quality of evidence	Measurement Error Overall rating / Quality of evidence	Responsiveness Overall rating / Quality of evidence
Player Load [51, 53, 66, 70, 71]	+/Moderate	-/Low	+/Low	?/Very Low	
Metabolic Power [54, 62, 63, 68]	±/Low		-/Low	?/Moderate	
Dist. PGPS > 20 W.kg – 1 [54, 63]	-/Moderate		±/Very Low	?/Low	
IMA COD [57]	-/Moderate		?/Moderate		
Knee Ang Vel [60]	+/High				
Hip Ang Vel [60]	+/High				
Peak Acc [64, 67]	-/High		-/Low	?/Very Low	
Time HMP>20W/kg [68]	?/Moderate				
PL/min [66]		-/Low	?/Very Low		
Hip Load [59]	+/High		+/Moderate		
Player Load anteroposterior plane (PLAP) [53, 71]			+/Very Low	?/Very Low	
Player Load mediolateral plane (PLML) [53, 71]			+/Very Low	?/Very Low	

PlayerLoad vertical plane (PLV) [53, 71]			+/Very Low	?/Very Low	
Impact Load [52, 55]			+/Moderate	?/Moderate	
Dist Acc (>3m/s ²)(>2m/s ²) [54, 61, 64]			-/Very Low	?/Moderate	
Dist Dec (>3m/s ²)(>2m/s ²) [54, 61, 64]			-/Very Low	?/Low	
IPA [56]			+/Moderate	?/Moderate	
N° Acc (>3m/s ²)(>2m/s ²) [62, 63, 64, 70]			-/Very Low	?/Very Low	
n° Dec (>3m/s ²)(>2m/s ²) [62, 63]			-/Low	?/Very Low	
HIE [65]			?/Very Low		
Fpeak (N) [69]			+/High	?/High	
Vertical stiffness (kN/m) [69]			+/High	?/High	
Propulsion efficiency (AU) [69]			+/High	?/High	
Dynamic Stress load (AU) [72]			+/High	?/High	

Abbreviations: ? indicates indeterminate; - insufficient; ± inconsistent; + sufficient; Dist.PGPS=distance metabolic power estimated from locomotor demands ; IMA COD= inertial movement analysis change of direction, Knee Ang Vel= knee angular velocity, Peak Acc= peak acceleration, Time HPM=, PL/min=, Dist Acc= distance acceleration, Dist Dec= distance deceleration, IPA= initial peak accelerations, N° Acc= number of acceleration , N°Dec= number of deceleration, HIE= high intensity effort, Fpeak= peak landing force, N= newton, AU= arbitrary units, k= vertical stiffness; m=meters

2.7 Discussion

The aim of this present study was to systematically review the literature on the methodological quality and statistical methods used to evaluate the validity, reliability, and responsiveness of GPS and IMU derived HIAD metrics. The main findings of the present systematic review showed that a limited number of studies focused on the validity of GPS and IMU devices in football. Very good methodological quality was found for criterion validity with only few studies that addressed construct validity. Relative reliability resulted the most investigated measurement property for the HIAD metrics, which presented a prevalence of doubtful (range: 60-76%) and a low percentage of adequate (range: 12-30%) methodological quality. Furthermore, no studies investigated the responsiveness of HIAD metrics. Overall, these results suggest the need for further studies of high methodological quality to assess the whole validation process of metrics assessed via GPS and IMU technology.

2.7.1 Methodological quality

According to the COSMIN guidelines (48), very good methodological quality was found in the two studies examining construct validity (59, 66). The methodology of these studies was rated as very good due to a clear definition of the constructs measured and the comparison instruments (i.e. hip load and player load/min metric, respectively) using appropriate hypothesis testing and comparison between subgroups. It should be noted that these studies assessed the construct validity overcoming the limitation of a lack of gold standard for IMU metrics (e.g., player load, body load, and dynamic stress load) measured on the field. A possible strategy to validate the metrics deriving from the accelerometer could be to assess their changes across activities that are different for the

construct of interest. For example, the evaluation of the player movement across changes of direction while controlling for distance and average speed. Considering that only two studies analyzed the construct validity of IMU metrics and no studies focused on GPS-derived metrics, more methodological approach for the assessment of the construct validity is needed.

Five (54, 57, 60, 67, 68) out of six included studies on criterion validity showed a very good methodological quality mainly due to their appropriate statistical analysis approach (i.e., correlation).

Barrett et al (51) have assessed the validity between live and post-downloaded player load data, however, due to methodological flaws and interpretation of the results the methodological quality was rated as doubtful.

Despite the studies focusing on the validity of the GPS and IMU metrics showed good methodological quality, their validity and accuracy of the metrics provided by the manufacturers need to be tested regularly, especially considering the frequent update of the software, hardware and firmware releases.

The Reliability of the HIAD metrics were more frequently examined compared to validity, however, only two studies (69, 71) were rated of adequate quality mainly due to their not clear definition of the test and re-test timepoints, the standardized conditions, and statistical analyses (ICC, SEM and limits of agreement).

Conversely, the heterogeneity across the methodological quality of the reliability studies was rated as doubtful (n= 13) due to a lack of stable and replicable conditions (i.e. not standardized protocol, changed environmental conditions, possible different physical and emotional status etc.), which may be a source of bias in the studies (52, 53, 54, 55, 56, 57, 59, 61, 62, 63, 66, 67, 70). Finally, Bredt et al. (64) and Dello Iacono et al. (65) were rated as inadequate for lack of reporting information on the time interval between the two assessments and the statistical tests used (e.g. the ICC parameter was not reported). In highlighting these methodological aspects, the present findings could stimulate scholars and manufacturers in engaging in strict GPS and IMU validation processes, which will allow practitioners to better monitor soccer players.

2.7.2 Quality evidence of metrics

Of the 24 HIAD analyzed in this review, non-uniform quality of evidence emerged regarding the criterion and construct validity of the metrics. Whilst an appropriate methodological approach was (64), questionable methodological quality of the studies and inconsistency of results determined the moderate level of evidence found for the criterion validity of player load (51), metabolic power (54), IMA (57), time HMP>20W/kg (68). Interestingly, the construct validity of player load and applied for criterion validity of knee and hip angular velocity, peak acceleration, and hip load metrics (59, 60, player load/min (73), showed low levels of evidence despite a good methodological quality. It could be speculated that the results of the scientific contributions included in the present systematic review might have been negatively affected by the small sample size (i.e., <50 participants) and the few studies present in literature. Due to the absence of non-invasive techniques for continuously monitoring internal loads on body structures, direct measurement is very difficult, thus severely

limiting the use of inertial accelerometers in assessing the biomechanical loading associated with specific football activities.

A heterogeneous quality of evidence was found in studies examining relative and absolute reliability with HIAD metrics resulting in high (n = 1), moderate (n = 6), low (n = 6) and very low (n = 10) quality of evidence. Buchheit et al (69) showed an adequate methodological quality of HIAD metrics and consistent outcomes derived from IMUs (i.e., force peak, vertical stiffness and propulsion efficiency). However, more research is needed to support these findings.

The studies investigating the relative reliability of IMA COD (57), hip load (59), impact load (52), and IPA (56) showed moderate level of evidence mainly due to doubtful methodological quality and a limited sample size (i.e. <100 participants). Player load (66, 70, 71), metabolic power (62, 63, 68), distance in PGPS>20W.kg⁻¹ (54, 63), player load/min (66), player load in different plane (53, 71), distance in acceleration and deceleration expressed in meters (54, 61, 62, 64, 70), and high intensity efforts (65) all showed a quality of evidence ranging between low and very low, due to doubtful and inadequate methodological quality, inconsistency of the outcomes and not adequate sample sizes. Studies that have assessed absolute reliability showed almost the same quality of evidence that was reported for the relative reliability. Nevertheless, the metabolic power (54, 62, 63) and distance in acceleration (54, 61), were reported to be at moderate quality of evidence, mainly substantiated by the consistency of the findings. The overall low quality of evidence suggests the urge to further study the reliability of the GPS and IMU metrics in football to allow an appropriate monitoring of players' external load during matches and training sessions.

Among the reviewed studies, it is important to note that none focused on the responsiveness of GPS and IMU metrics in football. This unexplored area is surprising when considering the large use of locomotor monitoring technology in professional and non-professional football environments. The assessment of responsiveness can provide information about the ability of a metric to detect meaningful changes and to detect if the players are changing their physical load (31). The assessment of each metric responsiveness is crucial to properly monitoring the physical load and assessing changes due to a particular condition (e.g., fatigue, playing home or away venues etc.) as well as to help coaches planning training and match strategies.

Based on the well-established COSMIN methodology for quality appraisal, the present work revealed important aspects related to the measurements of the external load in football.

Although the authors have applied this methodology in previous studies and have made the methodological process as up-to-date and reliable as possible (50), the present study is not free of limitations. In particular, the application of COSMIN requires several subjective decisions regarding any bias risk instrument, which are not standardized in sports science. Although there may be some subjectivity in the assessment, other reviewers may assess the selected instruments slightly differently, but we believe it is unlikely that the results will be substantially different. The purpose of this systematic review was to focus only on football, however, it would be appropriate to extend this methodological review process in all sports where technologies are used to monitor HIAD.

2.8 Conclusions

The present systematic review provides a comprehensive and detailed assessment of the methodological quality of the studies and the quality of evidence of the measurement properties of the most frequently used HIAD metrics in football. Despite the high expectations on wearable technology for monitoring physical load in football, some fundamental methodological aspects appeared to be undoubtedly neglected. Until appropriate validation studies are conducted, conclusions based on these HIAD metrics can be questionable and difficult to interpret. Future studies should improve the methodological quality of validation processes with the aim of raising the level of protocols used (e.g. protocol standardization, criterion measurement, validated results, sample size, statistical analysis). In addition, future studies should examine the responsiveness of the metrics.

2.9 References

1. Bangsbo J. 1994. The physiology of soccer – with special reference to intense intermittent exercise. *Acta Physiol Scand.* 151(619):1–155.
2. Shephard RJ. 1999. Biology and medicine of soccer: an update. *J Sports Sci.* 17:757–786.
3. Bloomfield J, Polman R, O’Donoghue P. 2007. Physical demands of different positions in FA Premier League soccer. *J Sports Sci Med.* 6:63–70.
4. Russell M, Sparkes W, Northeast J, Cook CJ, Love TD, Bracken RM, et al. 2016. Changes in acceleration and deceleration capacity throughout professional soccer match-play. *J Strength Cond Res.* 30(10):2839–2844.
5. Beato M, Drust B. 2021. Acceleration intensity is an important contributor to the external and internal training load demands of repeated sprint exercises in soccer players. *Res Sports Med [Internet].* 29(1):67–76.
6. Delaney JA, Cummins CJ, Thornton HR, Duthie GM. Importance, reliability, and usefulness of acceleration measures in team sports. *J Strength Cond Res.* 2018;32(12):3485–93.
7. Martín-García A, Casamichana D, Gómez Díaz A, Cos F, Gabbett TJ. 2018a. Positional differences in the most demanding passages of play in football competition. *J Sports Sci Med.* 17(4):563–570.
8. Riboli A, Semeria M, Coratella G, Esposito F. 2020. Effect of formation, ball in play and ball possession on peak demands in elite soccer. *Biol Sport.* 38 (2):195–205. doi:10.5114/biolSport.2020.98450.
9. Harper DJ, Carling C, Kiely J. High-intensity acceleration and deceleration demands in elite team sports competitive match play: a systematic review and meta-analysis of observational studies. *Sports Med.* 2019;49(12):1923–47.
10. Vanrenterghem J, Nedergaard NJ, Robinson MA, Drust B. Training load monitoring in team sports: a novel framework separating physiological and biomechanical load-adaptation pathways. *Sports Med.* 2017;47:2135–42.
11. Gastin PB, Hunkin SL, Fahrner B, Robertson S. Deceleration, acceleration, and impacts are strong contributors to muscle damage in professional Australian Football. *J Strength Cond Res.* 2019;33(12):3374–83.
12. de Hoyo M, Cohen DD, Sañudo B, Carrasco L, Álvarez-Mesa A, del Ojo JJ, et al. Influence of football match time–motion parameters on recovery time course of muscle damage and jump ability. *J Sports Sci.* 2016;34(14):1363–70.
13. Delves RIM, Aughey RJ, Ball K, Duthie GM. The Quantification of Acceleration Events in Elite Team Sport: a Systematic Review. *Sports Med Open.* 2021 Jun 30;7(1):45.
14. Dalen T, Ingebrigtsen J, Ettema G, Hjelde GH, Wisløff U. Player load, acceleration, and deceleration during forty-five competitive matches of elite soccer. *J Strength Cond Res.* 2016;30:351–9.
15. Verheul J, Nedergaard NJ, Pogson M, Lisboa P, Gregson W, Vanrenterghem J, et al. Biomechanical loading during running: can a two mass-spring-damper model be used to evaluate ground reaction forces for high-intensity tasks? *Sport Biomech.* 2019.

16. Harper DJ, Kiely J. Damaging nature of decelerations: do we adequately prepare players? *BMJ Open Sport Exerc Med.* 2018;4:e000379.
17. Young WB, Hepner J, Robbins DW. Movement demands in Australian rules football as indicators of muscle damage. *J Strength Cond Res.* 2012;26:492–6
18. Cardinale M, Varley MC. Wearable Training-Monitoring Technology: Applications, Challenges, and Opportunities. *Int J Sports Physiol Perform.* 2017 Apr;12(Suppl 2):S255-S262. doi: 10.1123/ijsp.2016-0423. Epub 2016 Nov 11. PMID: 27834559.
19. Malone JJ, Lovell R, Varley MC, Coutts AJ. Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *Int J Sports Physiol Perform.* 2017 Apr;12(Suppl 2):S218-S226. doi: 10.1123/ijsp.2016-0236. Epub 2016 Oct 13. PMID: 27736244.
20. Gómez-Carmona CD, Bastida-Castillo A, Ibáñez SJ, Pino-Ortega J. Accelerometry as a method for external workload monitoring in invasion team sports. A systematic review. *PLoS One.* 2020 Aug 25;15(8):e0236643. doi: 10.1371/journal.pone.0236643. PMID: 32841239; PMCID: PMC7447012.
21. Cummins C, Orr R, O'Connor H, West C. 2013. Global positioning systems (GPS) and microtechnology sensors in team sports: a systematic review. *Sports Med.* 43(10):1025–1042.
22. Akenhead R, Nassis GP. 2016. Training Load and Player Monitoring in High-Level Football: current Practice and Perceptions. *Int J Sports Physiol Perform.* 11(5):587–593
23. Varley MC, Gabbett T, Aughey RJ. Activity profiles of professional soccer, rugby league and Australian football match play. *J Sport Sci.* 2014;32:1858–66.
24. Aroganam G, Manivannan N, Harrison D. Review on wearable technology sensors used in consumer sport applications. *Sensors* 19: 1983, 2019.
25. Dellaserra CL, Gao Y, Ransdell L. Use of integrated technology in team sports: A review of opportunities, challenges, and future directions for athletes. *J Strength Cond Res* 28: 556–573, 2014.
26. Bartlett JD, O'Connor F, Pitchford N, Torres-Ronda L, Robertson SJ. Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an Individualized Approach. *Int J Sports Physiol Perform.* 2017 Feb;12(2):230-234. doi: 10.1123/ijsp.2015-0791. Epub 2016 Aug 24. PMID: 27194668.
27. Witte TH, Wilson AM. Accuracy of non-differential GPS for the determination of speed over ground. *J Biomech.* 2004;37(12):1891–8.
28. Scientific Advisor Committee. Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11:193– 205.
29. Atkinson G. Sport performance: variable or construct? *J Sports Sci.* 2002;20:291– 292.
30. Mokkink, L. B., Terwee, C. B., Patrick, D. L., et al. (2010a). International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *Journal of Clinical Epidemiology*, 63, 737–45.
31. Impellizzeri FM, Marcora SM. Test validation in sport physiology: lessons learned from clinimetrics. *Int J Sports Physiol Perform.* 2009 Jun;4(2):269-77. doi: 10.1123/ijsp.4.2.269. PMID: 19567929.
32. *Measurement in Medicine: A Practical Guide, Practical Guides to Biostatistics and Epidemiology*, Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, Dirk L. Knol.

33. Impellizzeri FM, Shrier I, McLaren SJ, et al. Understanding training load as exposure and dose. *Sports Med* 53: 1667–1679, 2023.
34. Scott MT, Scott TJ, Kelly VG. The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review. *J Strength Cond Res*. 2016 May;30(5):1470-90.
35. Crang ZL, Duthie G, Cole MH, Weakley J, Hewitt A, Johnston RD. The Validity and Reliability of Wearable Microtechnology for Intermittent Team Sports: A Systematic Review. *Sports Med*. 2021 Mar;51(3):549-565.
36. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*. 1998;26:217–238.
37. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016 Jun;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012. Epub 2016 Mar 31. Erratum in: *J Chiropr Med*. 2017 Dec;16(4):346.
38. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005 Feb;19(1):231-40.
39. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
40. Hopkins WG, Hawley JA, Burke LM. Design and analysis of research on sport performance enhancement. *Med Sci Sports Exerc*. 1999 Mar;31(3):472-85.
41. Mokkink, L. B., Terwee, C. B., Patrick, D. L., et al. (2010a). International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *Journal of Clinical Epidemiology*, 63, 737–45
42. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003;12: 349e62.
43. Dawson L, Beato M, Devereux G, McErlain-Naylor SA. A Review of the Validity and Reliability of Accelerometer-Based Metrics From Upper Back-Mounted GNSS Player Tracking Systems for Athlete Training Load Monitoring. *J Strength Cond Res*. 2024 Aug 1
44. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71
45. Mourad Ouzzani, Hossam Hammady, Zbys Fedorowicz, and Ahmed Elmagarmid. Rayyan — a web and mobile app for systematic reviews. *Systematic Reviews* (2016)
46. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539–549
47. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651–657. PubMed
48. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist manual. 2012. <http://www.cosmin.nl/>.
49. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157.

50. Jeffries AC, Wallace L, Coutts AJ, McLaren SJ, McCall A, Impellizzeri FM. Athlete-Reported Outcome Measures for Monitoring Training Responses: A Systematic Review of Risk of Bias and Measurement Property Quality According to the COSMIN Guidelines. *Int J Sports Physiol Perform.* 2020 Sep 21
51. Barrett S. Monitoring Elite Soccer Players' External Loads Using Real-Time Data. *Int J Sports Physiol Perform.* 2017 Nov 1;12(10):1285-1287.
52. Armitage M, Beato M, McErlain-Naylor SA. Inter-unit reliability of IMU Step metrics using IMeasureU Blue Trident inertial measurement units for running-based team sport tasks. *J Sports Sci.* 2021 Jul;39(13):1512-1518.
53. Barrett S, Midgley AW, Towlson C, Garrett A, Portas M, Lovell R. Within-Match PlayerLoad™ Patterns During a Simulated Soccer Match: Potential Implications for Unit Positioning and Fatigue Management. *Int J Sports Physiol Perform.* 2016 Jan;11(1):135-40.
54. Buchheit M, Manouvrier C, Cassirame J, Morin JB. Monitoring Locomotor Load in Soccer: Is Metabolic Power, Powerful? *Int J Sports Med.* 2015 Dec;36(14):1149-55.
55. Burland JP, Outerleys JB, Lattermann C, Davis IS. Reliability of wearable sensors to assess impact metrics during sport-specific tasks. *J Sports Sci.* 2021 Feb;39(4):406-411
56. Hughes T, Jones RK, Starbuck C, Sergeant JC, Callaghan MJ. The value of tibial mounted inertial measurement units to quantify running kinetics in elite football (soccer) players. A reliability and agreement study using a research orientated and a clinically orientated system. *J Electromyogr Kinesiol.* 2019 Feb;44:156-164.
57. Meylan C, Trewin J, McKean K. Quantifying Explosive Actions in International Women's Soccer. *Int J Sports Physiol Perform.* 2017 Mar;12(3):310-315.
58. Waldron M, Harding J, Barrett S, Gray A. A New Foot-Mounted Inertial Measurement System in Soccer: Reliability and Comparison to Global Positioning Systems for Velocity Measurements During Team Sport Actions. *J Hum Kinet.* 2021 Jan 30;77:37-50.
59. Wilmes E, Bastiaansen BJC, de Ruiter CJ, Vegter RJK, Brink MS, Weersma H, Goedhart EA, Lemmink KAPM, Savelsbergh GJP. Construct Validity and Test-Retest Reliability of Hip Load Compared With Playerload During Football-Specific Running, Kicking, and Jumping Tasks. *Int J Sports Physiol Perform.* 2022 Dec 1;18(1):3-10.
60. Wilmes E, de Ruiter CJ, Bastiaansen BJC, Zon JFJAV, Vegter RJK, Brink MS, Goedhart EA, Lemmink KAPM, Savelsbergh GJP. Inertial Sensor-Based Motion Tracking in Football with Movement Intensity Quantification. *Sensors (Basel).* 2020 Apr 29;20(9):2527.
61. Aquino R, Melli-Neto B, Ferrari JVS, Bedo BLS, Vieira LHP, Santiago PRP, Gonçalves LGC, Oliveira LP, Puggina EF. Validity and reliability of a 6-a-side small-sided game as an indicator of match-related physical performance in elite youth Brazilian soccer players. *J Sports Sci.* 2019 Dec;37(23):2639-2644.
62. Beato, M. Reliability of internal and external load parameters in 6 a-side and 7 a-side recreational football for health. *Sport Sci Health* 14, 709–714 (2018).
63. Beato M, Jamil M, Devereux G. Reliability of internal and external load parameters in recreational football (soccer) for health. *Res Sports Med.* 2018 Apr-Jun;26(2):244-250.
64. Breidt, Sarah & Praça, Gibson & Figueiredo, Lucas & Paula, Leandro & Ribeiro-Silva, Patrick & Andrade, A. & Greco, Pablo & Chagas, Mauro. (2016). Reliability of physical, physiological and

- tactical measures in small-sided soccer Games with numerical equality and numerical superiority. *Revista Brasileira de Cineantropometria e Desempenho Humano*. 18. 602-610.
65. Iacono AD, Unnithan V, Shushan T, King M, Beato M. Training load responses to football game profile-based training (GPBT) formats: effects of locomotive demands manipulation. *Biol Sport*. 2022 Jan;39(1):145-155.
 66. Barreira P, Robinson MA, Drust B, Nedergaard N, Raja Azidin RMF, Vanrenterghem J. Mechanical Player Load™ using trunk-mounted accelerometry in football: Is it a reliable, task- and player-specific observation? *J Sports Sci*. 2017 Sep;35(17):1674-1681.
 67. Buchheit M, Allen A, Poon TK, Modonutti M, Gregson W, Di Salvo V. Integrating different tracking systems in football: multiple camera semi-automatic system, local position measurement and GPS technologies. *J Sports Sci*. 2014 Dec;32(20):1844-1857.
 68. Rampinini E, Alberti G, Fiorenza M, Riggio M, Sassi R, Borges TO, Coutts AJ. Accuracy of GPS devices for measuring high-intensity running in field-based team sports. *Int J Sports Med*. 2015 Jan;36(1):49-53.
 69. Buchheit M, Lacombe M, Cholley Y, Simpson BM. Neuromuscular Responses to Conditioned Soccer Sessions Assessed via GPS-Embedded Accelerometers: Insights Into Tactical Periodization. *Int J Sports Physiol Perform*. 2018 May 1;13(5):577-583.
 70. Sparkes W, Turner AN, Weston M, Russell M, Johnston M, Kilduff LP. The between-week reliability of neuromuscular, endocrine, and mood markers in soccer players and the repeatability of the movement demands during small-sided games. *J Sports Med Phys Fitness*. 2022 Nov;62(11):1440-1448.
 71. Fitzpatrick JF, Hicks KM, Russell M, Hayes PR. The Reliability of Potential Fatigue-Monitoring Measures in Elite Youth Soccer Players. *J Strength Cond Res*. 2021 Dec 1;35(12):3448-3452.
 72. Beato M, Wren C, de Keijzer KL. The Interunit Reliability of Global Navigation Satellite Systems Apex (STATSports) Metrics During a Standardized Intermittent Running Activity. *J Strength Cond Res*. 2024 Feb 1;38(2):e49-e55.
 73. Oliva-Lozano JM, Conte D, Fortes V, Muyor JM. Exploring the Use of Player Load in Elite Soccer Players. *Sports Health*. 2023 Jan-Feb;15(1):61-66.
 74. Ferrari A, Cutti AG, Cappello A. A new formulation of the coefficient of multiple correlation to assess the similarity of waveforms measured synchronously by different motion analysis protocols. *Gait Posture*. 2010 Apr;31(4):540-2.

Chapter 3: Understanding inter unit reliability of Catapult Vector S7 in controlled situations

3.1 Overview

This chapter investigates the absolute and relative inter-unit reliability of GPS- and IMU-derived metrics obtained from the Catapult Vector S7 system during controlled football-specific drills. Using an electric golf cart to standardize movement patterns, four devices were tested across continuous, linear, and intermittent acceleration–deceleration conditions. Signal quality was strictly monitored through horizontal dilution of precision (HDOP) and satellite count to ensure high data accuracy. The analysis revealed excellent reliability for distance- and velocity-based metrics such as total distance, maximal speed, and metabolic power, whereas acceleration-, deceleration-, whereas IMU-derived variables showed higher variability between units. These findings confirm the reliability of the Catapult Vector S7 for general load monitoring but emphasize caution when interpreting high-frequency or direction-specific metrics.

3.2 Title Paper 2

“Inter-unit reliability of Catapult Vector S7 Device GPS and IMU Metrics for Simulated Football-Specific Drills”

Christian Di Claudio ^{1,2}, Maurizio Fanchini ^{2,3}, Carwyn Sharp ^{2,4}, Laura Capranica ¹, Daniele Conte ^{1,5}

1 Department of Movement, Human and Health Sciences, University of Rome “Foro Italico”, Rome, Italy

2 AS Roma Football Club, Rome, Italy

3 University of Verona, Italy

4 Department of Kinesiology and Sport Studies, University of Denver, Denver, U.S.A.

5 Department of Coaching Science, Lithuanian Sports University, Kaunas, Lithuania

(submitted to “European Journal of Sport Science”)

3.3 Abstract

Aim: This study aimed to evaluate the inter-unit absolute and relative reliability of GPS- and IMU-derived metrics from the Catapult Vector S7 system during simulated football-specific movements performed under highly controlled conditions using an electric golf car. Given the widespread use of wearable tracking technologies in elite football, understanding measurement consistency between units is crucial for accurate monitoring and training load interpretation.

Methods: Four Catapult Vector S7 units were mounted on different fixed positions of a motorized golf cart and tested across three movement drills simulating continuous running, straight-line accelerations, and intermittent accelerations/decelerations. Sixteen testing sessions were performed, with two excluded due to inadequate signal quality. GPS and IMU metrics, including total distance, speed zones, maximal speed, metabolic power variables, acceleration/deceleration metrics, and PlayerLoad measures, were extracted. Inter-unit reliability was quantified using the Intraclass Correlation Coefficient (ICC) and Typical Error expressed as coefficient of variation (TE%CV).

Results: Signal quality was consistently high across valid sessions (HDOP = 0.86 ± 0.01 ; satellites = 13.1 ± 0.09). Total distance, metabolic power, equivalent distance, and PlayerLoad showed excellent relative reliability (ICC > 0.96) and low values of absolute error (TE%CV < 10%). Maximal speed demonstrated moderate-to-excellent relative reliability (ICC = 0.45–0.97) with low value of absolute variability ($\leq 4.8\%$). Acceleration- and deceleration-based metrics showed heterogeneous reliability: while maximum acceleration/deceleration and density indices performed satisfactorily, threshold-based acceleration counts showed poor reliability (ICC = -0.07 to 0.53; TE%CV up to 106.5%). ANOVA revealed significant differences between device positions for PlayerLoad, with the centrally mounted unit recording higher values.

Conclusions: The Catapult Vector S7 exhibits strong inter-unit reliability for key GPS locomotor metrics and IMU-derived PlayerLoad during controlled football-specific simulated movements. However, acceleration and deceleration counts demonstrate substantial between-unit inconsistency, limiting their interchangeability. Signal quality monitoring (HDOP, satellite count) is essential for ensuring reliable data acquisition. Further research should assess intra-unit reliability and validate findings under athlete-driven, real-world conditions.

Keywords: Global Positioning System, Inertial Movement Unit, Monitoring Training Load, Workload, External Load

3.4 Introduction

The Global Positioning System (GPS) is one of the most widely used technologies for tracking football player movement (3), and allows a non-invasive monitoring of movement patterns during the activities (4, 6). The GPS is a satellite-based navigation technology that determines position and movement by calculating the time it takes for signals from multiple satellites to reach a ground-based receiver, it typically consists of an antenna, a processor, and software to compute location, speed, and trajectory. Modern GPS sport units are also equipped with an Inertial Measurement Unit (IMU), which integrates accelerometers, gyroscopes, and sometimes magnetometers to capture high-frequency data on acceleration, rotation, and orientation.

In modern football, monitoring training load is pivotal for the design of training sessions and optimizing players' performance. Training load has been conceptualised as internal and external training load, with the latter defined as the overall activity done by the players (5). Behind the plethora of external load parameters, GPS devices could enable the assessment of high-intensity efforts, acceleration, deceleration and change of direction that are critical in the fast-paced nature of football activity (5). However, these tools are not free of potential limitations that should be examined before its use in research and practice. The accuracy of the data is influenced by multiple factors, including the sensor's technical specifications, the precision of its calibration, the placement and stabilization of the sensor, and the methods used for signal processing (19). Furthermore, the quality of the GPS signal plays a pivotal role in determining the reliability of collected data. Factors such as horizontal dilution of precision (HDOP), interference from buildings or natural obstacles, atmospheric conditions, and the number of satellites during data collection can significantly affect GPS accuracy (3, 8, 9). A HDOP value of ≤ 1 represents the ideal positioning of satellites in the sky and results in greater accuracy, this also requires a low variability and high mean number (≥ 12) of satellites (9). Methodologies for improving signal quality, such as using advanced filtering techniques or integrating additional data sources, are essential for ensuring that GPS devices provide reliable results in challenging environments (10).

The use of GPS technology is widespread in football, where coaches and performance staff are used to analyse longitudinal data and/or compare different players of the team during the same activity. For example, it is very common to use the GPS data to discuss with players to push them to run more or more intensely during training sessions by comparing their values with a teammate's activity. Given the widespread use in practice, practitioners need to be confident on the GPS data quality. Therefore, GPS technology and derived metrics should undergo a strict validation process (7). Furthermore, to guide the interpretation of intra- and inter-player differences during sessions or drills, practitioners should be aware of the reliability of the measurement system, encompassing the intra- and inter-unit reliability, data processing, and computational components (11, 21). Reliability is defined as "the extent to which a measurement is free from errors" (11) and can be categorized as either absolute or relative reliability. Whilst absolute reliability refers to the amount of variation in repeated measurements for individuals allowing the assessment of changes over time, relative reliability measures how consistently the position of individuals in a sample is maintained across repeated measurements (12). Therefore, when using GPS and IMU technologies to monitor players' external load, it is crucial to assess both intra- and inter-unit reliability (6, 12).

To evaluate inter-unit reliability, that enables to understand the difference between devices of the same brand, a commonly used approach is to place multiple GPS units on a sled or rigid frame pulled by an athlete during specific movements (i.e., walking, jogging, acceleration and deceleration) over predefined distances (13-15). In ensuring identical motion profiles for each device simultaneously, numerous studies assessed technological variability in GPS measurements from one or more manufacturers (13-16).

Across different brands, the total distance covered and the peak/average speed metrics demonstrated high inter-unit reliability (particularly with 10 Hz devices) and low (<5%) coefficients of variation (CV), indicating good reliability and therefore, supporting the use of GPS for basic external load tracking (13-16). In contrast, acceleration and deceleration GPS-derived metrics exhibited high variability and heteroscedasticity (especially when based on threshold counts of accelerations >3 m·s⁻²). Brosnan and colleagues (15) reported CVs up to 12.8% for some acceleration indices, with significant between-unit differences between brands and within the same brand. Buchheit and colleagues (13) observed high variability (CVs up to 56%), with some GPS units recording 2 to 6 times more acceleration events compared to those of other units. Thornton and colleagues (14) confirmed poor inter-unit reliability for threshold-based acceleration and deceleration measures, while average acceleration metrics (e.g., mean absolute acceleration) showed moderate reliability (CV ≈ 6%). In simulated team sport movements, Crang and colleagues (16) reported a good inter-device reliability for distance, peak velocity, and average acceleration for the Catapult Vector S7 device, with coefficients of variation ranging from 0.1 to 3.9%. However, for most threshold-based acceleration and deceleration counts, the device exhibited inconsistent reliability, with significant variability observed across different sessions (16).

Considering the technological evolution, recently the Catapult Vector S7 device is based on a 10Hz GPS data acquisition frequency with implanted an accelerometer (- 3D +/- 16G sampled at 1kHz), a gyroscope (3D 2000 degrees/second at 100Hz), and a magnetometer (3D ±4900 μT @100Hz). However, at present there is no research available regarding the reliability of IMU-based measures, which are crucial to assess primary football conditioning as continuous running, or accelerations, decelerations, and rapid changes of directions (26). Therefore, the purpose of the present study was to investigate the inter-unit reliability of the GPS- and IMU- derived metrics in controlled continuous, accelerated and decelerated situations simulating low- and high-intensity running of elite football players (27) while considering the quality of signal during data collection. Overall, these novel findings could help the practitioners improving their quality of monitoring performance in football, ultimately enhancing training load management.

3.5 Methods

3.5.1 Experimental Approach to the Problem

The research protocol has been approved by the Institutional Research Committee of the University of Rome “Foro Italico” (CAR 212/2024). A total of 4 Catapult Vector S7 devices (Catapult Sports, Melbourne, Australia) (D1–D4) were randomly chosen from an overall sample of 25 units, and mounted on the vertical poles of the golf cart (EURECO ITALIA SRL, Pordenone, Italy) as follows: D1 at the central left rear pole, D2 on the left side pole, D3 on the central rear pole, and D4 on the front-

right pole (Fig 1). The antennas of each unit were pointed vertically to allow clear satellite reception. The electric golf car was driven by two members of the research team reproducing three different conditions (i.e drills) simulating the players' movements (27, 28), in the following order: 1) "Drill A" continuous running (velocity range from 0 to 24.6 km/h); 2) "Drill B" straight line-and-back running (velocity range from 0 to 22.5 km/h) ; and 3) "Drill C" and intermittent accelerations and decelerations (velocity range from 0 to 21.3 km/h).

The GPS signal quality was assessed by checking HDOP and the number of satellites connected for each unit during all the data collection, with a cut-off of HDOP higher than 1 and number of connected satellites lower than 10 established as exclusion criteria (8, 9). To ensure the best satellite's connections, the devices were activated 30 minutes before each session. The data collection was carried out over a period of one month (November-December 2024) for a total of 16 test sessions carried out in the same afternoon timeframe (14:00-15:30 h) on a 60x90 m football pitch made with artificial grass. After each session, data have been downloaded using the manufacturer's proprietary software (Open Field Software 8.1.0, Catapult Sports, Melbourne, Australia) and exported to an Excel database specifically created for further analysis. According to the Brosnan et al. (15), movements were identified when the GPS velocity/acceleration entered specific thresholds, being matched throughout the four devices considering the known start and end times for each drill. The considered metrics were: total distance, speed threshold-based distances covered between 15 and 20 km/h, 20 and 25 km/h; maximal speed, metabolic power (ref), metabolic power per kg (ref), equivalent distance (ref), equivalent distance index (calculated as the ratio between equivalent distance and the total distance), acceleration (i.e. $> 2.5 \text{ m}\cdot\text{s}^{-2}$) and deceleration- (i.e $> 2.5 \text{ m}\cdot\text{s}^{-2}$) based metrics (e.g., frequency of occurrence, distance, density), and density index, repetitive high intensity efforts (RHIE); and IMU derived metrics (e.g., player load, 2D player load and player load/m).

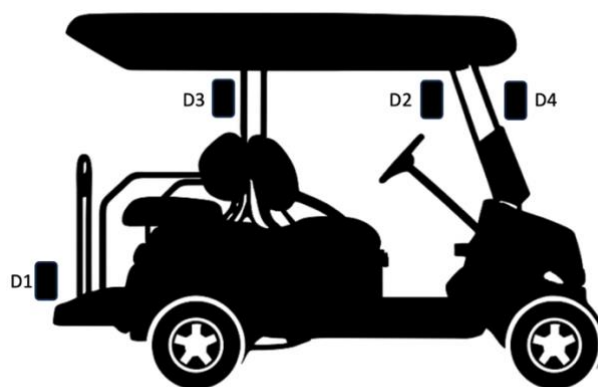


Figure 1. Schematic representation of the device positions on the electric golf car.

3.5.2 Procedures

The three drills of the experiment encompassed different simulated players' physical activities of low to high intensity interspersed by 30 s breaks. In particular, Drill A simulated continuous running, with a researcher driving the electric golf car around the course, passing outside 4 cones that delimited a distance of 880 meters (4 turns) continuous speed without stopping. When the 4 laps were completed, the driver stopped at the starting position for a 30-s break. Successively, Drill A was performed two more times for a total distance of 2640 m. Drill B resembled the straight line-and-back running, with a researcher driving the golf car for 4 repetitions of a 70 m straight line

progressive acceleration, a U-turn, and a 70 m progressive acceleration back to start line for a total distance of 280 meters. Drill C mirrored straight line intermittent accelerations and decelerations, which poses an emphasis on making the more game-related movements. In simulating high-speed accelerations, sudden breaks, and follow-up accelerations, from the start line, the researcher progressively accelerated the golf car to the middle line of the pitch (i.e. 35 meters) indicated by a horizontal white line, then the driver decelerated sharply to further accelerate to the end of the 70 meters where the car was stopped sharply. When 4 repetitions were accomplished, a 30-s break was allowed before restarting for another 3 rounds for a total distance of 840 meters.

3.5.3 Statistical analysis

Data from the three drills were considered together to avoid sample size limitation and presented as mean \pm standard deviation (SD). GPS sessions that showed HDOP higher than 1 and number of connected satellites lower than 10 were excluded due to low signal quality. A Shapiro–Wilk test was applied to assess the normal distribution of the dataset, showing all variables were normally distributed. A one-way analysis of variance (ANOVA) was performed with Jamovi (Jamovi 2.5.4.0) considering each GPS and IMU derived metric as the dependent variable and device position as the independent variable. When significant main effects were found, Tukey post hoc tests were conducted to identify pairwise differences between units. Statistical significance was set at $p < 0.05$. The Intraclass Correlation Coefficient (ICC) and Typical Error (TE) expressed as a percentage of Coefficient of Variation (%CV) were calculated for the assessment of relative and absolute inter-unit reliability, respectively. Whilst the ICC was applied to evaluate the ranking consistency between different units in the measurements, the TE assessed the difference between different units using the SD of the difference between measurements. Uncertainty was established as 95% confidence limits (CL) for ICC and TE. All reliability calculations were completed using Hopkins' reliability spreadsheet (Hopkins, 2017). The ICC was interpreted as excellent (>0.9), good (0.75–0.9), moderate (0.5–0.75) and poor (<0.50) (Koo & Li, 2016).

3.6 Results

Results of the following metrics are presented in Table 1, 2, 3: total distance, total distance between 15 and 20 km/h, total distance between 20 and 25 km/h, Maximal speed, metabolic power, equivalent distance, number of accelerations, number of decelerations and players load. Other metrics are presented as supplementary material.

Two test sessions that didn't meet the quality criteria (satellites: $n=12.1 \pm 0.2$; HDOP signal: 1.04 ± 0.01) were excluded, with the remaining showing good satellite connections ($n=13.1 \pm 0.09$) and HDOP signals (0.86 ± 0.01).

Descriptive statistics (mean \pm SD) for each device (D1–D4) for the 14 experimental sessions are reported in Table 1. The one-way ANOVA revealed a significant effect of device position on Player Load ($p = 0.014$), while no significant differences were found for the other GPS-derived metrics ($p > 0.05$). Post hoc Tukey comparisons indicated that the unit positioned at D1 recorded significantly higher Player Load values compared to D2 ($p = 0.002$), D3 ($p = 0.022$), and D4 ($p = 0.003$). Absolute and relative inter-unit reliability between devices are presented in Table 2 and 3, respectively. Descriptive values and reliability results from all drills are presented as Supplementary Material (Table 4 and 5).

Table 1				
Descriptive data (mean ± SD) of DRILL A, DRILL B, DRILL C				
KPI	D1	D2	D3	D4
Total Distance (m)	1316 ± 750	1320 ± 769	1362 ± 783	1341 ± 773
Total Distance 15-20 km/h (m)	401 ± 254	411 ± 247	397 ± 292	392 ± 248
Total Distance 20-25 km/h (m)	715 ± 691	718 ± 665	775 ± 777	755 ± 736
Max speed (km/h)	22.2 ± 0.8	22.1 ± 0.7	22.7 ± 2.1	22.4 ± 1.1
Metabolic Power (W/kg)	17.2 ± 2.8	17.1 ± 2.8	17.4 ± 3.0	17.4 ± 2.8
Equivalent distance (m)	1446 ± 735	1450 ± 742	1487 ± 765	1468 ± 750
Accelerations (n)	1.3 ± 2.1	1.4 ± 1.9	0.6 ± 1.3	1.8 ± 2.2
Decelerations (n)	5.7 ± 9.1	5.5 ± 9.2	5.8 ± 8.9	5.5 ± 9.1
Player Load (au)	45 ± 28	29 ± 16	32 ± 18	29 ± 16

Abbreviations: m=meters, s= seconds, au= arbitrary units, KPI=key performance indicators, D=device, SD=standard deviations

Table 1. Descriptive data of Catapult vector S7 in DRILL A, DRILL B, DRILL C

3.6.1 ANOVA AND POST HOC ANALYSIS

The one-way ANOVA revealed a significant effect of device position on Player Load ($p = 0.014$), while no significant differences were found for the other GPS-derived metrics ($p > 0.05$). Post hoc Tukey comparisons indicated that the unit positioned at D1 recorded significantly higher Player Load values compared to D2 ($p = 0.002$), D3 ($p = 0.022$), and D4 ($p = 0.003$).

3.6.2 ABSOLUTE RELIABILITY

Percentage typical errors (Table 2) were low and similar across all devices for total distance (range 1.3 to 2.1%), metabolic power (range 0.5 to 1.5 %) and equivalent distance (range 0.9 to 1.5 %). Results showed high TE% for the number of accelerations (range 61.7 to 106.5 %) and heterogenous TE% between devices for total distance 15-20 km/h (range 14.0 to 30.1 %), total distance 20-25 km/h (range 4.7 to 7.7 %), Max speed (range 0.6 to 4.8 %), decelerations (range 5.9 to 36.8 %), and player load (range 3.4 to 11.6 %).

Table 2

Inter device absolute reliability of Catapult vector S7
calculated by TE as a CV (%) in DRILL A, B, C

KPI	D1 vs D2	D3 vs D2	D4 vs D3	D1 vs D4	D3 vs D1	D2 vs D4
Total Distance (m)	1.5 (1.3 to 1.9)	1.3 (1.1 to 1.6)	1.4 (1.2 to 1.7)	2.1 (1.8 to 2.6)	1.4 (1.2 to 1.8)	1.9 (1.6 to 2.3)
Total Distance 15-20 km/h (m)	14.0 (11.8 to 17.5)	30.1 (25.0 to 38.0)	20.9 (17.5 to 26.2)	16.3 (13.7 to 20.4)	18.8 (15.7 to 23.5)	22.4 (18.7 to 28.1)
Total Distance 20-25 km/h (m)	4.7 (4.0 to 5.8)	7.7 (6.5 to 9.5)	4.9 (4.1 to 6.0)	5.4 (4.6 to 6.6)	6.6 (5.6 to 8.2)	5.6 (4.7 to 6.9)
Max speed (km/h)	0.6 (0.5 to 0.7)	4.8 (4.1 to 5.9)	4.9 (4.2 to 6.1)	2.0 (1.7 to 2.5)	4.7 (4.0 to 5.8)	2.0 (1.7 to 2.4)
Metabolic Power (W/kg)	0.5 (0.4 to 0.6)	1.1 (1.0 to 1.4)	1.5 (1.2 to 1.8)	0.9 (0.8 to 1.2)	1.1 (1.0 to 1.4)	1.0 (0.9 to 1.2)
Equivalent distance (m)	0.9 (0.8 to 1.1)	1.2 (1.0 to 1.4)	1.2 (1.0 to 1.5)	1.5 (1.3 to 1.9)	1.3 (1.1 to 1.6)	1.5 (1.2 to 1.8)
Accelerations (n)	81.5 (53.4 to 177.3)	106.5 (68.3 to 245.7)	87.3 (55.7 to 208.9)	83.8 (50.5 to 256.6)	93.5 (55.9 to 297.3)	61.7 (45.1 to 99.5)
Decelerations (n)	10.2 (7.7 to 15.6)	36.8 (27.0 to 59.2)	28.9 (21.6 to 44.9)	5.9 (4.4 to 8.8)	35.7 (26.7 to 55.1)	10.3 (7.8 to 15.7)
Player Load (au)	11.6 (9.8 to 14.4)*	3.4 (2.9 to 4.2)	6.3 (5.3 to 7.8)	9.9 (8.4 to 12.3)*	10.8 (9.1 to 13.3)*	6.3 (5.3 to 7.8)

Abbreviations: m=meters, s= seconds, au= arbitrary units, KPI=key performace indicators, D=device, SD=standard deviations, TE= Typical error

Notes: * indicates statistical significance ($p < 0.05$)

Table 2. Inter-device absolute reliability of Catapult vector S7 calculated as Percentage Typical Error

3.6.3 RELATIVE RELIABILITY

Relative reliability expressed with ICC was excellent and similar across all devices for total distance (ICC=1.00), total distance 20-25 km/h (m) (1.00), metabolic power (range 0.99 to 1.00), equivalent distance (ICC=1.00), number of decelerations (range 0.92 to 1.00) and player load (range from 0.96 to 0.99).

Results showed poor reliability for the number of accelerations (range -0.07 to 0.53) and heterogenous (from moderate to excellent) ICC results between devices for total distance 15-20 km/h (range 0.80 to 0.95), Max speed (range 0.45 to 0.97).

Table 3

Inter device relative reliability of Catapult vector S7 calculated by ICC in DRILL A, DRILL B, DRILL C

KPI	D1 vs D2	D3 vs D2	D4 vs D3	D1 vs D4	D3 vs D1	D2 vs D4
Total Distance (m)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)
Total Distance 15-20 km/h (m)	0.95 (0.91 to 0.97)	0.80 (0.68 to 0.87)	0.90 (0.83 to 0.94)	0.93 (0.89 to 0.96)	0.92 (0.86 to 0.95)	0.87 (0.79 to 0.92)
Total Distance 20-25 km/h (m)	1.00 (1.00 to 1.00)	1.00 (0.99 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (0.99 to 1.00)	1.00 (1.00 to 1.00)
Max speed (km/h)	0.97 /0.95 to 0.98)	0.45 (0.22 to 0.63)	0.51 (0.30 to 0.68)	0.79 (0.67 to 0.87)	0.50 (0.28 to 0.67)	0.78 (0.66 to 0.86)
Metabolic Power (W/kg)	1.00 (1.00 to 1.00)	1.00 (0.99 to 1.00)	0.99 (0.99 to 1.00)	1.00 (0.99 to 1.00)	1.00 (0.99 to 1.00)	1.00 (0.99 to 1.00)
Equivalent distance (m)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)
Accelerations (n)	0.25 (-0.85 to 0.76)	-.07 (-1.18 to 0.46)	0.16 (-1.05 to 0.61)	0.19 (-1.68 to 0.65)	0.04 (-1.59 to 0.60)	0.53 (0.11 to 0.74)
Decelerations (n)	0.99 (0.98 to 1.00)	0.92 (0.82 to 0.97)	0.95 (0.90 to 0.98)	1.00 (0.99 to 1.00)	0.92 (0.83 to 0.96)	0.99 (0.98 to 1.00)
Player Load (au)	0.96 (0.93 to 0.98)*	1.00 (0.99 to 1.00)	0.99 (0.98 to 0.99)	0.97 (0.95 to 0.98)*	0.97 (0.94 to 0.98)*	0.99 (0.98 to 0.99)

Abbreviations: m=meters, s= seconds, au= arbitrary units, KPI=key performance indicators, D=device, SD=standard deviations, ICC= Intraclass Correlation Coefficient

Table 3. Inter-device relative reliability of Catapult vector S7 calculated as ICC

3.7 Discussion

The primary aim of this study was to assess the inter-unit absolute and relative reliability of the Catapult Vector S7 system under controlled movement conditions using an electric golf car. Results showed that their-unit reliability is metric-dependent.

To the best of our knowledge, this is the first study to evaluate the inter-unit reliability of IMU derived metrics of Catapult Vector S7 (Player Load, Player Load/meter, 2D Player Load) throughout simulated low and high-intensity movements in a controlled condition. A key improvement in methodology for this research was achieved by incorporating signal quality management through the tracking of horizontal dilution of precision (HDOP) and the number of linked satellites during data collection. The analyses were restricted to sessions with high data quality. Previous research (3, 8) often overlooked (or not reported) this method, but it is crucial for validating GPS-derived metrics and should be a standard practice in applied sport science studies.

The main finding of the study was that GPS metrics (Total Distance, Average Speed, Peak Speed, Metabolic Power metrics, Threshold-based distance) exhibited low values for absolute inter-unit reliability across both linear and multidirectional drills, except for threshold-based distance between 15-20 km/h (range 14.0 to 30.1 %). In contrast, GPS acceleration- and deceleration-based metric showed more heterogeneity for absolute reliability, with acceleration density index, acceleration density, max acceleration and max deceleration performing consistently well whereas accelerations count, decelerations count, acceleration and deceleration distance metrics showing more heterogeneity (TE: range 5.9 to 106.5 % CV).

Furthermore, relative reliability revealed moderate to excellent inter-unit reliability for most distance and velocity-based metrics, including high-speed running distances and metabolic power and deceleration counts (ICC ranged from 0.80 to 1.00), across all drills. IMU derived metrics like

PlayerLoad showed excellent relative reliability with values ranging from 0.97 to 1.00. On the contrary, our results showed poor reliability for the number of accelerations (range -0.07 to 0.53) and heterogeneous (from moderate to excellent) ICC results between devices for Max speed (range 0.45 to 0.97), indicating limited consistency between units for these high-frequency, short-duration metrics.

Absolute inter-unit reliability of acceleration and deceleration-based GPS metrics are similar to findings of a previous study (16). The present findings showed a high variability across units for threshold-based acceleration counts exceeding $2.5 \text{ m}\cdot\text{s}^{-2}$ (TE: range 5.9 to 106.5 % CV). In the literature, Crang et al. (16) reported absolute repeatability as TE of acceleration counts stratified by intensity zones: low ($0\text{--}1 \text{ m}\cdot\text{s}^{-2}$), moderate ($2\text{--}3 \text{ m}\cdot\text{s}^{-2}$), and high ($>3 \text{ m}\cdot\text{s}^{-2}$). They similarly observed variability with TE ranging from 0.0 to 27.2% CV for low accelerations, 0.0 to 37.1% CV for moderate accelerations, and 3.7 to 32.6% CV for high accelerations. Factors that contribute to the instability of high-intensity Catapult Vector S7 GPS metrics could include contextual or environmental conditions, sensor technology limitations, and brand-specific computational algorithms for these variables (39).

Maximal speed also demonstrated good to excellent inter-unit reliability (ICC = 0.45–0.97) and good absolute reliability (TE%CV \leq 4.8%), corroborating earlier studies highlighting the robustness of peak velocity measurements with 10 Hz GPS devices from other brands (37, 38). In the study of Beato et al., (37), the interunit ICC for peak speed was 0.898 (95% CI: 0.831–0.939), indicating good reliability between GPS units. Although the model of GPS units used in this study is different from the one used in our study (Catapult S7 vs. STATSports Apex). Same results from Dawson et al., (38) showed that inter-unit reliability for maximal speed measured by STATSports Apex devices was excellent (ICC = 0.993), with a negligible percentage bias ($< 0.6\%$) and relatively narrow limits of agreement, indicating consistent reproducibility between units during shuttle run protocols with changes of direction.

Conversely, acceleration- and deceleration-based metrics presented greater heterogeneity, with certain indices such as acceleration density, maximal acceleration, and maximal deceleration showing moderate reliability, whereas acceleration and deceleration counts exhibited wide variability (TE%CV up to 106.5%; ICC = -0.07 to 0.53). The results of the present study align with findings of previous studies that identified threshold-based acceleration counts as among the least reliable GPS-derived variables (28, 35). The heterogeneity of the reliability values of these metrics may reflect the difficulty of consistently detecting high-frequency, short-duration acceleration events under dynamic conditions, compounded by device-specific filtering algorithms, environmental noise, and variations in antenna orientation. Interestingly, deceleration-based metrics showed higher reliability (ICC \geq 0.92), suggesting braking actions may be more consistently identified than acceleration events.

This study is also the first to evaluate the inter-unit reliability of IMU-derived metrics from the Catapult Vector S7. Player Load and Player Load/meter demonstrated moderate-to-excellent reliability (ICC = 0.97–1.00; TE%CV $< 10\%$). These results parallel those of Fitzpatrick et al. (25), who observed acceptable absolute and relative reliability for Player Load components in youth soccer players using the MinimaxX S4 device. Conversely, direction specific measures such as 2D Player Load showed greater variability (ICC = 0.13–0.74; TE%CV = 5.8–22.9%), indicating sensitivity to subtle differences in sensor orientation and attachment. Similar to findings by Mackay et al. (24), who reported very high reliability values for Player Load (ICC = 1.00; TE%CV = 1.5%) in indoor netball, the current results suggest that context and movement structure significantly influence IMU reliability. Controlled indoor conditions, stable surfaces, and human movement variability may all contribute to these differences. In contrast, the present study's outdoor setting and use of a motorized golf cart, though advantageous for motion standardization, could have introduced micro-vibrations and

positional discrepancies among devices. The significant ANOVA results for Player Load between device positions ($p = 0.014$), with the centrally mounted device showing higher values, further emphasize the sensitivity of IMU-based measures to mounting location and alignment. IMU-based metrics such as Player Load are valuable for quantifying the mechanical demands of training and match play, offering insight into non-locomotor activity not captured by GPS alone (33, 34). However, practitioners should interpret these data cautiously, as sensor placement, harness movement, and noise filtering algorithms can affect measurement consistency (10, 19). Integrating IMU data with GPS-derived metrics and subjective monitoring tools (e.g., RPE) may provide a more comprehensive view of external and internal load and fatigue status (35).

This integrated approach can improve load management strategies and contribute to both performance optimization and injury risk reduction.

The results of the present study are not free from limitations. The position of the GPS unit on the car could have influenced the results. The position for the unit D1 on the suspended posterior side of the car could have influenced the IMU accuracy with the device recording more “vibrations” compared to devices on other positions. Additionally, the thresholds used for certain metrics, such as accelerations and decelerations ($\pm 2.5 \text{ m}\cdot\text{s}^{-2}$), may not be ideal for all drills like the three used in this study. The same could be for the application of predefined thresholds for metrics like accelerations, decelerations, threshold-based distance for total distance (meters) covered between 15 and 20 km/h, and total distance (meters) covered between 20 and 25 km/h. However, this should not affect the assessment of inter-unit agreement in the current study, as all thresholds remained consistent. In addition, the adoption of a car movement lacks specificity compared to football activity. However, the aim of the present study was to assess the inter-unit reliability in a more controlled design. Future studies, should evaluate inter-unit reliability during specific football activities across a wider range of training conditions.

3.8 Practical application

Practical applications of these findings suggest that:

1. Catapult Vector S7 units can be used interchangeably when focusing on GPS-derived metrics such as total distance, maximal speed, and metabolic power.
2. Interchanging units should be avoided when tracking acceleration, deceleration, or IMU-based metrics due to higher variability between devices.
3. Practitioners should consistently monitor signal quality indicators such as HDOP and satellite count during both research and applied use, as these directly influence data reliability.

Future directions of this research should prioritize the investigation of intra-unit reliability, which may be more critical than inter-unit reliability in applied settings. Moreover, future studies should examine intra-unit consistency specifically in athlete populations, in order to validate findings obtained from controlled scenarios such as those involving a golf car.

3.9 Conclusions

The present findings support the use of Catapult Vector S7 devices for reliable GPS and IMU derived measurements of locomotor load in football environments, particularly when units are consistently worn. IMU-derived metrics such as Player Load can also be confidently used in performance analysis. However, caution is warranted when interpreting impact counts and high-frequency acceleration

data across units, especially in multi-directional contexts. Additionally, only one device model (Catapult Vector S7) was tested, thus limiting the generalizability of findings to other wearable tracking systems. Despite this study was conducted under controlled conditions using a motorized golf car to ensure movement standardization for enhancing the internal validity of inter-unit comparisons, further studies are needed to replicate the biomechanical variability present in human locomotion in a high ecological condition with soccer players. Therefore, to explore the impact of unit placement and software processing algorithms on measurement consistency, future research should examine intra-unit reliability across sessions, including free-running trials involving human participants.

3.10 References

- 1 Rago, V., Brito, J., Figueiredo, P., Costa, J., Barreira, D., Krstrup, P., & Rebelo, A. (2020). Methods to collect and interpret external training load using microtechnology incorporating GPS in professional football: A systematic review. *Research in Sports Medicine*, 28(4), 437–458. <https://doi.org/10.1080/15438627.2020.1792294>
- 2 Cummins, C., Orr, R., O'Connor, H., & West, C. (2013). Global positioning systems (GPS) and microtechnology sensors in team sports: A systematic review. *Sports Medicine*, 43(10), 1025–1042. <https://doi.org/10.1007/s40279-013-0071-2>
- 3 Malone, J. J., Lovell, R., Varley, M. C., & Coutts, A. J. (2017). Unpacking the black box: Applications and considerations for using GPS devices in sport. *International Journal of Sports Physiology and Performance*, 12(2), 18–26. <https://doi.org/10.1123/ijsp.2016-0078>
- 4 Cardinale M, Varley MC. Wearable Training-Monitoring Technology: Applications, Challenges, and Opportunities. *Int J Sports Physiol Perform*. 2017 Apr;12(Suppl 2):S255-S262. doi: 10.1123/ijsp.2016-0423. Epub 2016 Nov 11. PMID: 27834559.
- 5 Akenhead R, Nassis GP. Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *Int J Sports Physiol Perform*. 2016 Jul;11(5):587-93. doi: 10.1123/ijsp.2015-0331. Epub 2015 Oct 9. PMID: 26456711.
- 6 Rago V, Brito J, Figueiredo P, Costa J, Barreira D, Krstrup P, Rebelo A. Methods to collect and interpret external training load using microtechnology incorporating GPS in professional football: a systematic review. *Res Sports Med*. 2020 Jul-Sep;28(3):437-458. doi: 10.1080/15438627.2019.1686703. Epub 2019 Nov 22. PMID: 31755307.
- 7 Impellizzeri FM, Marcora SM. Test validation in sport physiology: lessons learned from clinimetrics. *Int J Sports Physiol Perform*. 2009 Jun;4(2):269-77. doi: 10.1123/ijsp.4.2.269. PMID: 19567929.
- 8 Delves RIM, Aughey RJ, Ball K, Duthie GM. The Quantification of Acceleration Events in Elite Team Sport: a Systematic Review. *Sports Med Open*. 2021 Jun 30;7(1):45.
- 9 Specht, M. (2022). Experimental Studies on the Relationship Between HDOP and Position Error in the GPS System. *Metrology and Measurement Systems*, 29, 17-36
- 10 Ellens S, Middleton K, Gatin PB, Varley MC. Techniques to derive and clean acceleration and deceleration data of athlete tracking technologies in team sports: A scoping review. *J Sports Sci*. 2022 Aug;40(16):1772-1800. doi: 10.1080/02640414.2022.2054535. Epub 2022 Apr 21. PMID: 35446231.

- 11 Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26:217–238.
- 12 Mokkink, L. B., Terwee, C. B., Patrick, D. L., et al. (2010a). International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *Journal of Clinical Epidemiology*, 63, 737–45.
- 13 Buchheit M, Al Haddad H, Simpson BM, Palazzi D, Bourdon PC, Di Salvo V, Mendez-Villanueva A. Monitoring accelerations with GPS in football: time to slow down? *Int J Sports Physiol Perform.* 2014 May;9(3):442-5. doi: 10.1123/ijsp.2013-0187. Epub 2013 Jul 30. PMID: 23916989.
- 14 Thornton HR, Nelson AR, Delaney JA, Serpiello FR, Duthie GM. Interunit Reliability and Effect of Data-Processing Methods of Global Positioning Systems. *Int J Sports Physiol Perform.* 2019 Apr 1;14(4):432-438. doi: 10.1123/ijsp.2018-0273. Epub 2019 Feb 20. PMID: 30204529.
- 15 Brosnan RJ, Watson G, Stuart W, Twentyman C, Kitic CM, Schmidt M. The Validity, Reliability, and Agreement of Global Positioning System Units-Can We Compare Research and Applied Data? *J Strength Cond Res.* 2022 Dec 1;36(12):3330-3338. doi: 10.1519/JSC.0000000000004139. Epub 2021 Sep 29. PMID: 34593728.
- 16 Crang ZL, Duthie G, Cole MH, Weakley J, Hewitt A, Johnston RD. The inter-device reliability of global navigation satellite systems during team sport movement across multiple days. *J Sci Med Sport.* 2022 Apr;25(4):340-344. doi: 10.1016/j.jsams.2021.11.044. Epub 2021 Nov 26. PMID: 34893434.
- 17 Eignor DR. The standards for educational and psychological testing. In: *APA Handbook of Testing and Assessment in Psychology, Vol. 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology.* Washington, WA: American Psychological Association, 2013. pp. 245–250
- 18 Sechrest L. Validity of measures is No simple matter. *Health Serv Res* 40: 1584–1604, 2005
- 19 Hughes GTG, Camomilla V, Vanwanseele B, Harrison AJ, Fong DTP, Bradshaw EJ. Novel technology in sports biomechanics: Some words of caution. *Sports Biomech* 23: 393–401, 2024
- 20 Impellizzeri FM, Shrier I, McLaren SJ, et al. Understanding training load as exposure and dose. *Sports Med* 53: 1667–1679, 2023
- 21 Scott PJ, Brown AW, Adedeji T, et al. A review of measurement practice in studies of clinical decision support systems 1998–2017. *J Am Med Inform Assoc* 26: 1120–1128, 2019
- 22 Clavel P, Leduc C, Morin JB, Owen C, Samozino P, Peeters A, Buchheit M, Lacombe M. Concurrent Validity and Reliability of Sprinting Force-Velocity Profile Assessed With GPS Devices in Elite Athletes.

Int J Sports Physiol Perform. 2022 Jul 11;17(10):1527-1531. doi: 10.1123/ijsp.2021-0339. PMID: 35894960

23 Cormier P, Tsai MC, Meylan C, Agar-Newman D, Epp-Stobbe A, Kalthoff Z, Klimstra M. Concurrent Validity and Reliability of Different Technologies for Sprint-Derived Horizontal Force-Velocity-Power Profiling. *J Strength Cond Res.* 2023 Jun 1;37(6):1298-1305. doi: 10.1519/JSC.0000000000004429. Epub 2023 Jan 18. PMID: 36727987

24 Mackay L, Sawczuk T, Jones B, Darrall-Jones J, Clark A, Whitehead S. The reliability of a commonly used (CatapultTM Vector S7) microtechnology unit to detect movement characteristics used in court-based sports. *J Sports Sci.* 2025 Mar;43(6):555-564. doi: 10.1080/02640414.2025.2468585. Epub 2025 Feb 24. PMID: 39994076

25 Fitzpatrick JF, Hicks KM, Russell M, Hayes PR. The Reliability of Potential Fatigue-Monitoring Measures in Elite Youth Soccer Players. *J Strength Cond Res.* 2021 Dec 1;35(12):3448-3452. doi: 10.1519/JSC.0000000000003317. PMID: 31498220.

26 Delaney JA, Cummins CJ, Thornton HR, Duthie GM. Importance, Reliability, and Usefulness of Acceleration Measures in Team Sports. *J Strength Cond Res.* 2018 Dec;32(12):3485-3493. doi: 10.1519/JSC.0000000000001849. PMID: 28195980.

27 Gualtieri A, Rampinini E, Dello Iacono A, Beato M. High-speed running and sprinting in professional adult soccer: Current thresholds definition, match demands and training strategies. A systematic review. *Front Sports Act Living.* 2023 Feb 13;5:1116293. doi:

28 Crang ZL, Duthie G, Cole MH, Weakley J, Hewitt A, Johnston RD. The Validity and Reliability of Wearable Microtechnology for Intermittent Team Sports: A Systematic Review. *Sports Med.* 2021 Mar;51(3):549-565. doi: 10.1007/s40279-020-01399-1. Epub 2020 Dec 24. PMID: 33368031.

29 West SW, Clubb J, Torres-Ronda L, Howells D, Leng E, Vescovi JD, Carmody S, Posthumus M, Dalen-Lorentsen T, Windt J. More than a Metric: How Training Load is Used in Elite Sport for Athlete Management. *Int J Sports Med.* 2021 Apr;42(4):300-306. doi: 10.1055/a-1268-8791. Epub 2020 Oct 19. PMID: 33075832.

30 Di Salvo V, Gregson W, Atkinson G, Tordoff P, Drust B. Analysis of high intensity activity in Premier League soccer. *Int J Sports Med.* 2009 Mar;30(3):205-12. doi: 10.1055/s-0028-1105950. Epub 2009 Feb 12. PMID: 19214939.

31 Di Salvo V, Pigozzi F, González-Haro C, Laughlin MS, De Witt JK. Match performance comparison in top English soccer leagues. *Int J Sports Med.* 2013 Jun;34(6):526-32. doi: 10.1055/s-0032-1327660. Epub 2012 Nov 26. PMID: 23184481.

- 32 Buchheit M, Mendez-Villanueva A, Simpson BM, Bourdon PC. Match running performance and fitness in youth soccer. *Int J Sports Med.* 2010 Nov;31(11):818-25. doi: 10.1055/s-0030-1262838. Epub 2010 Aug 11. PMID: 20703978.
- 33 Barreira, P., Robinson, M. A., Drust, B., Vanrenterghem, J., & Kalinowski, P. (2017). Validity and reliability of a commercially available accelerometer for measuring player load in team sports. *Journal of Sports Sciences*, 35(16), 1584–1591. <https://doi.org/10.1080/02640414.2016.1227461>
- 34 Boyd, L. J., Ball, K., & Aughey, R. J. (2011). The reliability of MinimaxX accelerometers for measuring physical activity in Australian football. *International Journal of Sports Physiology and Performance*, 6(3), 311–321. <https://doi.org/10.1123/ijsp.6.3.311>
- 35 Dawson L, Beato M, Devereux G, McErlain-Naylor SA. A Review of the Validity and Reliability of Accelerometer-Based Metrics From Upper Back-Mounted GNSS Player Tracking Systems for Athlete Training Load Monitoring. *J Strength Cond Res.* 2024 Aug 1;38(8):e459-e474. doi: 10.1519/JSC.0000000000004835. Epub 2024 Jul 8. PMID: 38968210.
- 36 Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2020). Accuracy of distance recordings in eight positioning-enabled sport watches: Instrument validation study. *JMIR mHealth and Uhealth*, 8(6), e17118. <https://doi.org/10.2196/17118>
- 37 Beato M, Wren C, de Keijzer KL. The Interunit Reliability of Global Navigation Satellite Systems Apex (STATSports) Metrics During a Standardized Intermittent Running Activity. *J Strength Cond Res.* 2024 Feb 1;38(2):e49-e55. doi: 10.1519/JSC.0000000000004613. Epub 2023 Oct 6. PMID: 37815238.
- 38 Dawson L, McErlain-Naylor SA, Devereux G, Beato M. Interunit reliability of STATSports APEX global navigation satellite system and accelerometer-derived metrics during shuttle run protocols of varied distances and change of direction frequency. *J Sports Sci.* 2025 Sep 10:1-13. doi: 10.1080/02640414.2025.2555554.
- 39 Altman DG, Royston P. The cost of dichotomising continuous variables. *Bmj* 2006;332(7549):1080.

Chapter 4: Assessment inter unit reliability of Catapult Vector S7 during game situations

4.1 Overview

This chapter examines the inter-unit reliability of the Catapult Vector S7 GPS and IMU devices during ecological, football-specific training sessions. Three elite youth players were monitored across 28 training sessions to assess the consistency of key performance indicators derived from GPS and IMU metrics, including distance, speed, metabolic power, and Player Load. Signal quality was carefully controlled through horizontal dilution of precision (HDOP) and satellite count to ensure robust data accuracy.

Results showed excellent relative reliability (ICC > 0.90) and good absolute reliability (TE as %CV < 2%) for total distance, maximal speed, and metabolic power. However, higher heterogeneity was observed for high-intensity acceleration and IMU-derived metrics. These findings support the use of Catapult Vector S7 units for reliable external load monitoring in football, while emphasizing the need for caution when interpreting high-frequency or direction-specific data.

4.2 Title Paper 3

“Inter-unit reliability of Catapult Vector S7 Device GPS and IMU Metrics for Simulated Football-Specific Drills”

Christian Di Claudio ^{1,2}, Maurizio Fanchini ^{2,3} Mattia Palmieri ², Matteo Forlini ², Laura Capranica ¹, Daniele Conte ^{1,4}

1 Department of Movement, Human and Health Sciences, University of Rome “Foro Italico”, Rome, Italy

2 AS Roma Football Club, Rome, Italy

3 University of Verona, Italy

4 Department of Coaching Science, Lithuanian Sports University, Kaunas, Lithuania

(submitted to “Journal of Sport Sciences”)

4.3 Abstract

Background: This study examined the inter-unit reliability of Catapult Vector S7 GPS and IMU devices in quantifying external load metrics during real-world football training, with emphasis on signal quality parameters.

Methods: Three elite U18 football players (mean \pm SD; age, 17.6 ± 0.4 years; height, 181.6 ± 6.0 cm; body mass, 71.3 ± 6.4 kg) each wore two GPS/IMU units across 33 training sessions. Metrics included total and threshold-based distance, acceleration and deceleration counts, metabolic power, PlayerLoad, and derived IMU metrics. Inter-unit reliability was assessed using intraclass correlation coefficients (ICC) and typical error (TE) expressed as the coefficient of variation (%CV). Linear mixed models (LMM) was used to test for differences between paired units. Sessions with poor GPS signal quality (HDOP > 1 or < 10 satellites) were excluded.

Results: Twenty-eight sessions (85%) met quality criteria. Relative reliability was excellent for most GPS-derived metrics (ICC > 0.90) and slightly lower for maximum acceleration (ICC = $0.64\text{--}0.84$) and PlayerLoad-derived metrics (ICC = $0.89\text{--}0.90$). Absolute reliability (TE as %CV) was good for total distance, maximal speed, and metabolic power (TE%CV $< 2\%$), but more variable for threshold-based distance, acceleration-deceleration, and IMU metrics (TE%CV = $0.0\text{--}35.4\%$). LMMs showed no significant differences between paired devices ($p > 0.05$).

Conclusions: The Catapult Vector S7 showed high inter-unit reliability for most GPS and IMU-derived metrics in football. However, caution is advised when interpreting high-speed, acceleration-deceleration, and IMU-based variables. Implementing signal quality controls (HDOP, satellite count) improves data accuracy and should be a standard in applied sports science.

Keywords: Global Positioning System, Inertial Movement Unit, Monitoring Training Load, Workload, External Load

4.4 Introduction

Football is a dynamic and physically demanding team sport that involves high-intensity activities such as running, sprinting, turning, and sudden changes of direction (1–3). Therefore, movement analysis is essential for assessing high-intensity actions performed during matches and training sessions, helping coaches to optimize training plans, enhance players' performance, and evaluate fitness and fatigue levels (4, 5).

To provide a comprehensive understanding of player dynamics during high-intensity accelerations and decelerations (HIAD), Global Positioning System (GPS) devices and Inertial Measurement Units (IMUs) are among the most commonly used technologies by practitioners (6–9). Specifically, GPS devices provide real-time tracking of players' speed and distance covered during training and matches, whereas IMUs capture inertial accelerations, decelerations, and rotational forces by integrating data from accelerometers and gyroscopes. Although GPS and IMU technologies are extensively used in elite sports (4), rigorous validity, reliability, and responsiveness assessments are essential to examine their measurement properties. Such evaluations are critical to making informed decisions regarding player performance profiles within sport-specific and research contexts (12–16). In this regard, assessing the inter-unit reliability of individual GPS and IMU devices is crucial to enable valid comparisons of data obtained from different players during the same training session. The inter-unit reliability of GPS devices has been extensively studied (13), however, a comprehensive overview remains challenging due to variability across studies in terms of movement types, metrics, manufacturers, models, and sampling rates (17). Furthermore, an important but often neglected aspect is the quality of the signal, reflected by the horizontal dilution of precision (HDOP) and the number of connected satellites, which determine the accuracy and interpretability of the collected data (8, 21).

The inter-unit reliability of the Catapult Vector S7 GPS device in team sports has been previously investigated (17–20). Crang et al. (17) examined the GPS reliability under simulated team-sport conditions, and reported good inter-device reliability for total distance, peak velocity, and average acceleration, with coefficients of variation (CVs) ranging from 0.1% to 3.9%. Conversely, the authors observed inconsistent reliability for most threshold-based acceleration and deceleration counts, with substantial variability across sessions. Two additional studies (18, 19) assessed the reliability of the Catapult Vector S7 GPS system for sprinting force–velocity (F–V) profiling (F_0 = theoretical maximal horizontal force; V_0 = sprinting velocity). Clavel et al. (18) reported a small inter-unit typical error, demonstrating high reliability of these devices for assessing F–V profile variables, with CVs ranging from 0.5% to 2.0% and intraclass correlation coefficients (ICC) between 0.93 and 0.99.

In a similar context, Cormier et al. (19) investigated the inter-unit reliability of different technologies for sprint profiling and reported high intra-system reliability of GPS/GNSS technology operating at 10 Hz, with accuracy bias <2% and strong ICC values (0.84–0.99), indicating consistent performance across devices and moderate-to-good accuracy for F–V profile metrics in single maximal sprint efforts.

During sport-specific movements in court-based sports, Mackay et al. (20) found good consistency among different Catapult Vector S7 units for inertial movement analysis (ICC > 0.95; CV = 4%) and moderate CVs (7.7%) during netball training. For IMU-derived PlayerLoad metrics, excellent reliability was observed (ICC = 1.00; CV = 1.5%), whereas peak and slow PlayerLoad values showed moderate reliability (ICC = 0.78; CV = 8.9%).

Although the inter-unit reliability of GPS and IMU systems has been well documented in the literature (14, 20), limited information is available regarding their reliability for metrics describing HIAD during football-specific movement patterns. This information is essential for football fitness

coaches to evaluate the feasibility of comparing GPS and IMU outcomes between players (i.e., units) during training sessions. Therefore, the aim of the present study was to evaluate the absolute and relative inter-unit reliability of Catapult Vector S7 GPS and IMU devices during football-specific training drills, providing valuable insights for football scientists and practitioners when interpreting and reporting between-player differences in external load.

4.5 Methods

Three elite young players (mean \pm SD; age, 17.6 ± 0.4 years; stature, 181.6 ± 6.0 cm; body mass, 71.3 ± 6.4 kg) of an Italian Serie A U18 team volunteered to participate in the study. The Research Ethics Committee of the University of Rome "Foro Italico" approved the study (protocol number: 212/2024). Participants were informed about the risks and benefits of the research before giving their informed consent and were asked to maintain their daily routine during the daily practice according to the team's schedule.

4.5.1 Design

This study encompasses an observational design in which players were required to wear two units of the Catapult Vector S7 (Catapult Sports, Melbourne, Australia) for 33 regular morning (10:00-12:00 h) football training sessions organized from October to December 2024. Specifically, devices 1 and 2 (D1 and D2) were assigned to player 1, devices 3 and 4 (D3 and D4) to player 2, and devices 5 and 6 (D5 and D6) to player 3. The devices were positioned one above and one below in a custom-made manufacturer-supplied vest at the center of the back, between the shoulder blades. Their position has been switched each session (Figure 1).

4.5.2 Procedures

All sessions were completed outdoor and planned by the coach; therefore, the content, intensity and duration of the sessions varied as per a normal training session (74 ± 24 min of session duration). The Catapult Vector S7 microtechnology unit has an acquisition frequency of 10Hz and it is housed with a 100 hz accelerometer sampled at 1kHz ($-3D \pm 16G$), a 100Hz gyroscope (-2000 degrees/second), and a 100Hz magnetometer ($- D \pm 4900 \mu T$).

The football training sessions were carried on a full-size artificial grass (68 m wide, 102 m long) football pitch, the GPS signal quality was checked through the number of satellites (< 10) and the HDOP (> 1) signal (Delves et al., 2022). After each session, data were downloaded with the manufacturers' proprietary software (OpenField Software 8.1.0, Catapult Sports, Melbourne, Australia). Afterward, files were exported to a secure database for analysis with data relative to drills activities and passive recovery periods identified by two researchers with more than five years' experience working with elite football athletes.

After checking the quality of the signal, five training sessions were excluded because they didn't meet the inclusion criteria (number of satellites = 11.5 ± 0.5 , HDOP = 1.04 ± 0.03), therefore a total of 28 training sessions were retained with a number of satellites of 14.71 ± 0.5 and HDOP signal of 0.79 ± 0.03 . The considered metrics encompassed: total distance (m) and distance in different speed threshold (i.e., 15- 20 km/h, 20-25 km/h); maximal speed (km/h); metabolic power (W/kg), equivalent distance (m) and equivalent distance index (%); acceleration and deceleration events ($n > 2.5 \text{ m/s}^2$) and distance (m), acceleration density (au), acceleration density index (au); and IMU derived player load (au), 2D player load (au).

4.5.3 Statistical analysis

Data are presented as means \pm standard deviations (SD). A Shapiro–Wilk test was applied to evaluate the normality assumption of the statistical distribution. To control for the potential effect of the device positioning (one upper and another lower positioned) the devices were alternatively positioned in one of the two pockets. Subsequently, linear mixed models (LMM) were used to examine the differences due to device positioning in the metrics. LMM analyses were performed using Jamovi software (version 2.3.21 available at <https://www.jamovi.org>). An alpha level of $P < 0.05$ was set a priori for statistical significance.

Relative and absolute inter-unit reliability of the coupled units was established using the intraclass correlation coefficient (ICC) and typical error (TE) expressed as percentage coefficient of variation (CV%). The ICC was used to evaluate the consistency between the measurements. The TE expressed the difference between the measurements in absolute terms using the SD of the difference between measurements. Uncertainty was presented as 90% confidence interval (CLs) for ICC and TE. All reliability calculations were completed using Hopkins' reliability spreadsheet (22). The ICC was interpreted as excellent (>0.9), good (0.75–0.9), moderate (0.5–0.75) and poor (<0.50) (23).



Figure 1. Graphic representation of the custom-made manufacturer-supplied vest.

4.6 Results

Descriptive statistics (mean \pm SD) of each metric and device (D1, D2, D3, D4, D5, D6) during the 28 training sessions analyzed, are reported in Table 1. The results showed that no difference in each metric due to their positioning, therefore not effect of the position of GPS devices in the two pockets were found (Table 2).

Table 1Descriptive data (mean \pm SD) of training sessions

KPI	D1	D2	D3	D4	D5	D6
	mean \pm SD	mean \pm SD	mean \pm SD	mean \pm SD	mean \pm SD	mean \pm SD
Total Distance (m)	6085 \pm 2350	6069 \pm 2375	5906 \pm 2331	5867 \pm 2327	6103 \pm 2441	6064 \pm 2444
Total Distance 15-20 km/h (m)	625 \pm 367	644 \pm 400	516 \pm 327	519 \pm 344	667 \pm 363	679 \pm 378
Total Distance 20-25 km/h (m)	246 \pm 223	243 \pm 213	214 \pm 224	214 \pm 227	324 \pm 220	317 \pm 208
Total Distance >25 km/h (m)	66 \pm 92	62 \pm 92	54 \pm 87	52 \pm 87	59 \pm 65	51 \pm 55
Max speed (km/h)	27.3 \pm 2.6	27.0 \pm 2.7	26.7 \pm 2.5	26.9 \pm 2.6	27.7 \pm 1.9	27.1 \pm 1.6
Metabolic Power (W/kg)	7.0 \pm 1.2	6.9 \pm 1.9	6.6 \pm 1.0	6.6 \pm 1.0	7.3 \pm 1.0	7.3 \pm 1.0
Equivalent distance (m)	7450 \pm 2647	7409 \pm 2637	7177 \pm 2838	7140 \pm 2372	7745 \pm 2767	7683 \pm 2732
Equivalent Distance Index (%)	1.2 \pm 0.0	1.1 \pm 0.0	1.1 \pm 0.0	1.1 \pm 0.0	1.2 \pm 0.0	1.2 \pm 0.0
Accelerations (n)	38 \pm 24	39 \pm 25	41 \pm 22	43 \pm 23	50 \pm 26	52 \pm 29
Decelerations (n)	39 \pm 22	38 \pm 21	25 \pm 15	26 \pm 16	51 \pm 28	51 \pm 25
Acceleration distance (m)	135.7 \pm 80.5	134.9 \pm 82.8	151.4 \pm 76.4	155.5 \pm 78.0	176.7 \pm 91.8	175.7 \pm 97.7
Deceleration distance (m)	53.1 \pm 33.8	50.1 \pm 32.6	38.7 \pm 27.2	38.3 \pm 25.2	75.1 \pm 39.3	72.5 \pm 38.3
Max acceleration (m·s ⁻²)	4.1 \pm 0.4	4.2 \pm 0.4	4.2 \pm 0.3	4.2 \pm 0.3	4.2 \pm 0.4	4.1 \pm 0.3
Max deceleration (m·s ⁻²)	5.1 \pm 0.9	4.9 \pm 0.9	4.4 \pm 0.7	4.4 \pm 0.7	5.2 \pm 0.5	5.0 \pm 0.6
Acceleration density index (au)	3.1 \pm 0.3	3.1 \pm 0.3	3.1 \pm 0.3	3.1 \pm 0.3	3.4 \pm 0.2	3.4 \pm 0.2
Acceleration density (au)	0.4 \pm 0.1	0.4 \pm 0.1	0.4 \pm 0.1	0.4 \pm 0.1	0.5 \pm 0.1	0.5 \pm 0.1
Player Load (au)	688 \pm 295	717 \pm 271	617 \pm 246	645 \pm 252	688 \pm 309	722 \pm 296
2 Dimensional Player Load (au)	478 \pm 188	503 \pm 107	419 \pm 141	450 \pm 148	480 \pm 193	508 \pm 182

m=meters, s= seconds, au= arbitrary units, KPI=key performance indicators, D=device, SD= standard deviations

Table 2

Position analysis Linear Mixed Model (LMM)

KPI	p value <0.05
Total Distance (m)	0.92
Total Distance 15-20 km/h (m)	0.71
Total Distance 20-25 km/h (m)	0.90
Total Distance >25 km/h (m)	0.66
Max speed (km/h)	0.56
Metabolic Power (W/kg)	0.90
Equivalent distance (m)	0.89
Equivalent Distance Index (%)	0.96
Accelerations (n)	0.85
Decelerations (n)	0.68
Acceleration distance (m)	0.87
Deceleration distance (m)	0.60
Max acceleration (m·s ⁻²)	0.95
Max deceleration (m·s ⁻²)	0.35
Acceleration density index (au)	0.59
Acceleration density (au)	0.59
Player Load (au)	0.59
2 Dimensional Player Load (au)	0.74

m=meters, s= seconds, au= arbitrary units, KPI=key performance indicators, D=device

Tables 3 and 4 show the absolute and relative inter-device reliability between the various devices.

Absolute reliability values are reported in Table 3. The lowest values of absolute reliability (expressed in TE as %CV) resulted for total distance (range: 0.7-1.7%), total distance covered between 15 and 20 km/h (range: 3.0- 4.3%), maximal speed (range: 1.4-2.6%), metabolic power (range: 0.4-1.5%), equivalent distance (range: 0.5-1.6%), equivalent distance index (range: 0.6 to 0.9%), acceleration distance (range: 5.5-9.6%), deceleration distance (range: 6.3 to 11.8%), max acceleration (range: 3.3-5.8%), max deceleration (range: 4.0-7.0%), acceleration density (range: 1.7-2.3%), acceleration density index (range: 2.8-3.0%), whereas the highest values were found for total distance covered between 20 and 25 km/h (range: 7.4-23.2%), total distance covered > 25 km/h (range: 14.9-35.4%), acceleration count (range: 8.1-12.7%), deceleration count (range: 3.2-14.1%), player load (range: 15.6-17.6%), 2D player load (range: 14.0-16.0%).

KPI	D1 vs D2	D3 vs D4	D5 vs D6
	TE as %CV (90% CLs)	TE as %CV (90% CLs)	TE as %CV (90% CLs)
Total Distance (m)	0.9 (0.8 to 1.2)	0.7 (0.7 to 1.1)	1.4 (1.4 to 2.1)
Total Distance 15-20 km/h (m)	4.3 (3.6 to 5.7)	3.0 (2.4 to 3.9)	3.9 (3.2 to 5.1)
Total Distance 20-25 km/h (m)	23.2 (18.6 to 30.9)	7.4 (6.1 to 9.7)	7.8 (6.3 to 10.2)
Total Distance >25 km/h (m)	17.7 (13.9 to 24.8)	14.9 (11.6 to 21.3)	35.4 (27.7 to 49.3)
Max speed (km/h)	1.9 (1.5 to 2.4)	1.4 (1.1 to 1.8)	2.6 (2.1 to 3.4)
Metabolic Power (W/kg)	0.6 (0.5 to 0.7)	0.4 (0.4 to 0.6)	1.5 (1.3 to 2.0)
Equivalent distance (m)	0.6 (0.5 to 0.8)	0.5 (0.4 to 0.6)	1.6 (1.3 to 2.1)
Equivalent Distance Index (%)	0.7 (0.6 to 1.0)	0.6 (0.5 to 0.8)	0.9 (0.8 to 1.2)
Accelerations (n)	12.7 (1.3 to 16.7)	8.1 (6.6 to 10.6)	9.5 (7.7 to 12.4)
Decelerations (n)	3.2 (10.7 to 17.3)	14.1 (11.4 to 18.6)	8.6 (7.0 to 11.3)
Acceleration distance (m)	9.6 (7.8 to 12.6)	5.5 (4.5 to 7.2)	7.7 (6.3 to 10.1)
Deceleration distance (m)	8.4 (6.8 to 10.9)	11.8 (9.6 to 15.5)	6.3 (5.1 to 8.2)
Max acceleration (m·s ⁻²)	5.8 (4.7 to 7.6)	3.3 (2.7 to 4.2)	5.7 (4.6 to 7.4)
Max deceleration (m·s ⁻²)	7.0 (5.7 to 9.2)	4.0 (3.2 to 5.2)	5.3 (4.3 to 6.9)
Acceleration density index (au)	3.0 (2.4 to 3.9)	2.8 (2.3 to 3.6)	2.9 (2.3 to 3.7)
Acceleration density (au)	2.3 (1.8 to 2.9)	2.0 (1.6 to 2.6)	1.7 (1.4 to 2.3)
Player Load (au)	15.6 (12.7 to 20.7)	15.8 (12.8 to 20.9)	17.6 (14.2 to 23.3)
2 Dimensional Player Load (au)	14.3 (11.6 to 18.8)	14.0 (11.3 to 18.4)	16.0 (13.0 to 21.2)

TEE as a CV (%)= Typical error as a Coefficient variation (%), m=meters, s= seconds, au= arbitrary units, KPI=key performance indicators
D=device, CLs=confident limits

Relative reliability values are reported in Table 4 Relative reliability showed excellent ICC values, except for maximal speed (good-excellent, 0.73-0.92) and acceleration density index (good-excellent, 0.72-0.92), in D5 and D6 devices. In addition, maximal acceleration (range: moderate-good, 0.64-0.84), and maximal deceleration (range: good-excellent, 0.80-0.96) showed lower values of ICC compared to other metrics.

KPI	D1 vs D2	D3 vs D4	D5 vs D6
	ICC (90% CLs)	ICC (90% CLs)	ICC (90% CLs)
Total Distance (m)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)
Total Distance 15-20 km/h (m)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)
Total Distance 20-25 km/h (m)	0.98 (0.95 to 0.99)	1.00 (1.00 to 1.00)	0.99 (0.99 to 1.00)
Total Distance >25 km/h (m)	0.98 (0.96 to 0.99)	0.98 (0.97 to 0.99)	0.94 (0.89 to 0.99)
Max speed (km/h)	0.97 (0.94 to 0.98)	0.98 (0.97 to 0.99)	0.85 (0.73 to 0.92)
Metabolic Power (W/kg)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	0.99 (0.97 to 0.99)
Equivalent distance (m)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)	1.00 (1.00 to 1.00)
Equivalent Distance Index (%)	0.92 (0.85 to 0.96)	0.94 (0.89 to 0.97)	0.84 (0.72 to 0.91)
Accelerations (n)	0.97 (0.94 to 0.98)	0.98 (0.97 to 0.99)	0.98 (0.96 to 0.99)
Decelerations (n)	0.97 (0.95 to 0.98)	0.97 (0.94 to 0.98)	0.98 (0.97 to 0.99)
Acceleration distance (m)	0.98 (0.97 to 0.99)	0.99 (0.99 to 1.00)	0.98 (0.97 to 0.99)
Deceleration distance (m)	0.99 (0.98 to 1.00)	0.99 (0.97 to 0.99)	0.99 (0.98 to 1.00)
Max acceleration (m·s ⁻²)	0.73 (0.54 to 0.85)	0.84 (0.71 to 0.91)	0.64 (0.41 to 0.79)
Max deceleration (m·s ⁻²)	0.87 (0.77 to 0.93)	0.96 (0.92 to 0.98)	0.80 (0.64 to 0.89)
Acceleration density index (au)	0.91 (0.83 to 0.95)	0.93 (0.88 to 0.96)	0.84 (0.72 to 0.92)
Acceleration density (au)	0.98 (0.96 to 0.99)	0.98 (0.96 to 0.99)	0.98 (0.97 to 0.99)
Player Load (au)	0.91 (0.83 to 0.95)	0.91 (0.83 to 0.95)	0.90 (0.82 to 0.95)
2 Dimensional Player Load (au)	0.91 (0.84 to 0.95)	0.89 (0.81 to 0.94)	0.90 (0.81 to 0.95)

*ICC Intraclass correlation coefficient, m=meters, s= seconds, au= arbitrary units, KPI=key performance indicators
D=device, CLs=confident limits

4.7 Discussion

The aim of this study was to evaluate the inter-unit reliability of Catapult Vector S7 GPS devices for monitoring high-intensity activities in an ecological and sport-specific environment of elite football, including metrics derived from both GPS and IMU, and substantiating the quality of the findings by considering the signal indicators HDOP and number of satellites. The results of the present study showed excellent inter-unit reliability of the Catapult Vector S7 GPS and IMU devices during football-specific training drills, with TE as a CV% generally below 5% for total distance, speed, and metabolic power-related metrics. Acceleration, deceleration, and Player Load derived metrics measures showed greater variability but maintained good-to-excellent intraclass correlation coefficients. After disregarding 15% of the training sessions for not meeting the quality criteria, the results indicate a high relative reliability between units for most of the analyzed metrics, with excellent intraclass correlation coefficients (ICCs: >0.90-1.00). In terms of absolute reliability, metrics such as total distance, maximal speed, metabolic power, and equivalent distance index demonstrated very low variability between devices (CV <2%), confirming their usefulness in routine monitoring the individual load of training sessions. However, CV values higher than 35.4% for high-intensity metrics such as distance covered higher than 20 Km/h and acceleration/deceleration counts (>2.5 m/s²) warrant caution when coaches compare data across units and differences between players. These findings are consistent with previous research on the validity and reliability of the Catapult Vector S7 system in other sports contexts (17, 18, 19, 20). Specifically, Crang et al. (17) reported inter-device reliability from good to poor across sessions for accelerations (CV range: 0.0-37.1%) and deceleration (range: 0.0-74.2%) counts categorized into intensity zones, whereas in the present study only events exceeding 2.5 m/s² were considered. The present findings are partially in line with the literature (17)

for the reliability of total distance and peak velocity (CVs ranging from 0.9% to 2.6%), the threshold-based distances >15 km/h reaching CVs up to 35.4% highlight the need for caution when comparing players based on high-speed running thresholds. In fact, differences in such metrics should always be interpreted in the context of the device's measurement error before drawing definitive conclusions.

To the best of our knowledge, this is the first study that evaluate the inter-unit reliability of accelerometer derived metrics (Player Load and 2D Player Load) throughout several football training sessions. The high ecological validity of our design is supported by the monitoring of professional football players during their regular training routines, providing practical and detailed insights for coaches and performance staff using GPS/IMU devices to assess external load in this athletic population. A recent study by MacKay and colleagues (20) examined the inter-unit reliability of the Catapult Vector S7 device, focusing specifically on metrics derived from the inertial measurement units (IMUs). They reported excellent reliability for both Player Load and 2D Player Load, with intraclass correlation coefficients (ICCs) ranging from 0.99 to 1.00, and coefficients of variation (CVs) ranging from 1.3% to 2.0%. These findings align with ours in terms of relative reliability (ICC range: 0.81–0.95), but differ markedly in terms of absolute reliability, with our CV% values ranging from 12.7% to 21.2%. This discrepancy may be attributed to the fundamental differences between the movement characteristics of the two sports analyzed. Variations in movement patterns, spatiotemporal demands, and game intensity could influence the way the accelerometer embedded within the GPS device detects and processes external load, thereby affecting the resulting data. The present study extends these findings to football, a sport characterized by complex, multidirectional movements performed in outdoor environments where GPS signal quality may be more variable, although the metrics derived from the accelerometer are not affected by signal quality.

A notable methodological advancement in this study was the inclusion of signal quality control through the HDOP and the number of connected satellites. By excluding sessions with HDOP >1 and satellite counts <10, the analyses were based on high-quality data, enhancing the robustness of the findings. This approach, often overlooked in previous literature (8, 21), proved to be essential for ensuring the validity of GPS-derived metrics and should become standard practice in applied sport science research, especially in considering that 15% of the training sessions were disregarded. However, the lack of significant differences between the coupled devices across all analyzed metrics ($P > 0.05$) substantiates the consistency of measurements even when the vertical positioning of the units was swapped mid-collection. This is an important practical outcome, supporting the robustness of the protocol adopted in real-world training conditions.

Even if this study provides interesting results for football coaches and practitioners, some limitations should be acknowledged. Firstly, a small sample size ($n = 3$ players) was recruited, although this was mitigated by the substantial number of training sessions analyzed ($n = 28$). Furthermore, the study focused exclusively on elite youth football players without considering sport-specific positional differences, which limits the generalizability of the findings to other populations and other sports. Therefore, future studies should examine the GPS inter-unit reliability in other football populations such as male and female professional players. Further research involving larger samples, diverse competitive contexts, and more varied movement protocols is recommended to confirm these findings and to enhance standardization in the use of GPS/IMU technology in sport performance monitoring.

4.8 Practical Application

The Catapult Vector S7 GPS/IMU system can be confidently used to monitor total distance, maximal speed, and metabolic power during football training sessions, as these variables show excellent inter-unit reliability. However, practitioners should interpret acceleration and deceleration counts, high-speed running, and IMU-derived metrics (e.g., PlayerLoad) with caution due to higher variability between units. To enhance data quality and comparability, the following practices are recommended:

1. Maintain consistent unit-athlete pairing across sessions.
2. Prioritize aggregated variables (e.g., total distance or total acc–dec counts) over isolated intensity bands for reliable monitoring.
3. Implement signal-quality thresholds ($\text{HDOP} \leq 1$; ≥ 10 satellites) before data analysis.

4.9 Conclusions

A significant methodological advancement in the current study is the systematic control of GPS signal quality. Excluding sessions with inadequate HDOP or low satellite counts not only improved data consistency but also represents a best-practice guideline for practitioners and researchers. This procedure strengthens the reliability of GNSS-derived variables and should be considered an essential step in both applied and research settings to ensure the comparability of data across sessions and athletes.

Taken together, these findings demonstrate that the Catapult Vector S7 GPS/IMU system provides high inter-unit reliability for the majority of GPS-derived metrics during football training sessions. The absolute and relative reliability parameters of total distance, maximal speed, and metabolic power variables supports their use for monitoring and comparing external load between players or across sessions. However, the larger heterogeneity observed in high-speed and IMU-based measures highlights the need for caution when interpreting acceleration–deceleration counts or PlayerLoad data between units. Ensuring consistent unit assignment and signal quality thresholds remains critical for accurate player load analysis in applied football contexts.

4.9 References

1. Bangsbo J. 1994. The physiology of soccer – with special reference to intense intermittent exercise. *Acta Physiol Scand.* 151(619):1–155.
2. Shephard RJ. 1999. Biology and medicine of soccer: an update. *J Sports Sci.* 17:757–786.
3. Bloomfield J, Polman R, O’Donoghue P. 2007. Physical demands of different positions in FA Premier League soccer. *J Sports Sci Med.* 6:63–70.
4. Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A., & Clubb, J. (2022). Tracking systems in team sports: A narrative review of applications of the data and sport specific analysis. *Sports Medicine - Open*, 8(1), 1–22. <https://doi.org/10.1186/s40798-022-00408-z>
5. Akenhead R, Nassis GP. Training load and player monitoring in high-level football: current practice and perceptions. *Int J Sports Physiol Perform.* 2016;11:587–93.
6. Cummins C, Orr R, O’Connor H, West C. 2013. Global positioning systems (GPS) and microtechnology sensors in team sports: a systematic review. *Sports Med.* 43(10):1025–1042.
7. Cardinale M, Varley MC. Wearable Training-Monitoring Technology: Applications, Challenges, and Opportunities. *Int J Sports Physiol Perform.* 2017 Apr;12(Suppl 2):S255–S262. doi: 10.1123/ijsp.2016-0423. Epub 2016 Nov 11. PMID: 27834559.
8. Malone JJ, Lovell R, Varley MC, Coutts AJ. Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *Int J Sports Physiol Perform.* 2017 Apr;12(Suppl 2):S218–S226. doi: 10.1123/ijsp.2016-0236. Epub 2016 Oct 13. PMID: 27736244.
9. Gómez-Carmona CD, Bastida-Castillo A, Ibáñez SJ, Pino-Ortega J. Accelerometry as a method for external workload monitoring in invasion team sports. A systematic review. *PLoS One.* 2020 Aug 25;15(8):e0236643. doi: 10.1371/journal.pone.0236643. PMID: 32841239; PMCID: PMC7447012.
10. Harper DJ, Carling C, Kiely J. High-intensity acceleration and deceleration demands in elite team sports competitive match play: a systematic review and meta-analysis of observational studies. *Sports Med.* 2019;49(12):1923–47.
11. Delves RIM, Aughey RJ, Ball K, Duthie GM. The Quantification of Acceleration Events in Elite Team Sport: a Systematic Review. *Sports Med Open.* 2021 Jun 30;7(1):45.
12. Impellizzeri FM, Marcora SM. Test validation in sport physiology: lessons learned from clinimetrics. *Int J Sports Physiol Perform.* 2009 Jun;4(2):269–77. doi: 10.1123/ijsp.4.2.269. PMID: 19567929.
13. Cragg ZL, Duthie G, Cole MH, Weakley J, Hewitt A, Johnston RD. The Validity and Reliability of Wearable Microtechnology for Intermittent Team Sports: A Systematic Review. *Sports Med.* 2021 Mar;51(3):549–565.
14. Dawson L, Beato M, Devereux G, McErlain-Naylor SA. A Review of the Validity and Reliability of Accelerometer-Based Metrics From Upper Back-Mounted GNSS Player Tracking Systems for Athlete Training Load Monitoring. *J Strength Cond Res.* 2024 Aug 1
15. Scott MT, Scott TJ, Kelly VG. The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review. *J Strength Cond Res.* 2016 May;30(5):1470–90.

16. Luteberget LS, Gilgien M. Validation methods for global and local positioning-based athlete monitoring systems in team sports: a scoping review. *BMJ Open Sport Exerc Med*. 2020 Aug 24;6(1):e000794. doi: 10.1136/bmjsem-2020-000794. PMID: 33062300; PMCID: PMC7520549.
17. Crang ZL, Duthie G, Cole MH, Weakley J, Hewitt A, Johnston RD. The inter-device reliability of global navigation satellite systems during team sport movement across multiple days. *J Sci Med Sport*. 2022 Apr;25(4):340-344. doi: 10.1016/j.jsams.2021.11.044. Epub 2021 Nov 26. PMID: 34893434.
18. Clavel P, Leduc C, Morin JB, Owen C, Samozino P, Peeters A, Buchheit M, Lacombe M. Concurrent Validity and Reliability of Sprinting Force-Velocity Profile Assessed With GPS Devices in Elite Athletes. *Int J Sports Physiol Perform*. 2022 Jul 11;17(10):1527-1531. doi: 10.1123/ijsp.2021-0339. PMID: 35894960.
19. Cormier P, Tsai MC, Meylan C, Agar-Newman D, Epp-Stobbe A, Kalthoff Z, Klimstra M. Concurrent Validity and Reliability of Different Technologies for Sprint-Derived Horizontal Force-Velocity-Power Profiling. *J Strength Cond Res*. 2023 Jun 1;37(6):1298-1305. doi: 10.1519/JSC.0000000000004429. Epub 2023 Jan 18. PMID: 36727987.
20. Mackay L, Sawczuk T, Jones B, Darrall-Jones J, Clark A, Whitehead S. The reliability of a commonly used (CatapultTM Vector S7) microtechnology unit to detect movement characteristics used in court-based sports. *J Sports Sci*. 2025 Mar;43(6):555-564. doi: 10.1080/02640414.2025.2468585. Epub 2025 Feb 24. PMID: 39994076.
21. Delves RIM, Aughey RJ, Ball K, Duthie GM. The Quantification of Acceleration Events in Elite Team Sport: a Systematic Review. *Sports Med Open*. 2021 Jun 30;7(1):45.
22. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc*. 2009;41(1):3-13.
23. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016 Jun;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012. Epub 2016 Mar 31. Erratum in: *J Chiropr Med*. 2017 Dec;16(4):346. doi: 10.1016/j.jcm.2017.10.001. PMID: 27330520; PMCID: PMC4913118.
24. Gómez-Carmona CD, Pino-Ortega J, Sánchez-Ureña B, Ibáñez SJ, Rojas-Valverde D. Accelerometry-based external load indicators in sport: too many options, same practical outcome? *Int J Environ Res Public Health*. 2019;16(24):5101.

Chapter 5: General Discussion and Conclusion

5.1 General Discussion

The overarching aim of this doctoral research was to contribute to the methodological understanding and practical application of wearable tracking technologies, specifically Global Positioning System (GPS) and Inertial Measurement Unit (IMU) devices, within elite football contexts. Through a progressive approach consisting of three complementary studies—a systematic literature review and two empirical investigations—the project sought to evaluate the validity, reliability, and responsiveness of key high-intensity acceleration and deceleration (HIAD) metrics. Collectively, the findings provide an integrated perspective on the current state of measurement quality in football load monitoring, emphasizing the critical importance of methodological rigor, signal quality management, and contextual interpretation of wearable-derived metrics (Cragg et al., 2021, Scott et al., 2016).

The first study, a systematic review, aims to provide a comprehensive overview of the methodological quality and statistical approaches adopted in previous research assessing GPS and IMU metrics (Cragg et al., 2021, Scott et al., 2016). Despite the widespread use of these technologies in team sports, the review revealed substantial heterogeneity in the validation protocols adopted in the studies, with many studies failing to fully meet the quality standards for measurement properties as defined by the COSMIN guidelines (Prinsen et al., 2018). Specifically, reliability was the most frequently investigated aspect, whereas construct validity and responsiveness were largely neglected. Moreover, the overall level of evidence across studies was often moderate to low, mainly for acceleration- and deceleration-based metrics. These findings underscore the need for more standardized methodological approaches, larger sample sizes, and transparent reporting practices. The review also highlighted that responsiveness—a key property for detecting meaningful changes in training load over time—remains an unexplored area in football GPS/IMU technology validation research.

Building upon these methodological gaps, the second study investigated the inter-unit reliability of GPS- and IMU-derived metrics using a highly controlled experimental design. The use of an electric golf cart allowed for the replication of linear and multidirectional movements under consistent and repeatable conditions. Importantly, this study incorporated a novel methodological consideration frequently overlooked (or not reported) in research studies and practice—signal quality monitoring through horizontal dilution of precision (HDOP) and satellite count—which ensured that only high-quality data were included in the analyses. The findings demonstrated excellent reliability for GPS-derived metrics related to distance and velocity (e.g., total distance, maximal speed, metabolic power), while acceleration- and deceleration-based metrics exhibited greater variability across units and a low level of reliability. IMU-derived metrics, particularly PlayerLoad and PlayerLoad/m, showed heterogeneity in absolute and relative reliability parameters, though direction-specific components were less consistent. These results reinforce the notion that GPS technology performs robustly for general locomotor monitoring, whereas caution should be adopted when interpreting high-frequency or threshold-based acceleration data (Malone et al., 2017).

The third study extended these findings to a highly ecological setting, examining the inter-unit reliability of the Catapult Vector S7 system during real football training sessions with elite youth players. This study provided valuable insights into the functioning of GPS and IMU metrics under real-world conditions, accounting for the inherent variability of human motion and environmental influences. By introducing strict quality control based on HDOP and the number of connected satellites, the study strengthened the reliability assessment process. Consistent with the controlled experiment (i.e. study n 2), distance- and velocity-based metrics showed excellent inter-unit reliability, while acceleration and deceleration-based metrics demonstrated higher variability. IMU-derived measures such as PlayerLoad and 2D PlayerLoad confirmed good relative reliability but resulted in not good absolute reliability expressed in TE%CV, possibly due to the complexity of multidirectional and high-intensity movements typical of football training. These findings highlight the delicate balance between methodological control and ecological validity in sports technology research. Whilst controlled conditions maximize precision, real-world testing ensures applicability and practical relevance.

The integration of findings across the systematic review and the two experimental studies reveals a coherent methodological narrative regarding the validation process of GPS and IMU derived metrics in football. The systematic review identified substantial variability in reported reliability outcomes across studies, emphasizing the influence of methodological factors such as device model, sampling frequency, movement task and statistical approach.

The ecological study (**Chapter 4**) showed that reliability outcomes are notably affected when measurements are performed in football specific conditions. Compared with the controlled setting study (**Chapter 3**), several metrics exhibited increased typical error and wider limits of agreement (i.e. High intensity threshold distance and Player Load derived metrics), despite similar device configuration. This finding supports the conclusion that reduced reliability in applied settings is not solely attributable to device limitations, but rather to the inherent variability of football movements and contextual factors such as player interactions and task constraints.

Across both experimental studies, indicators of signal quality (e.g., satellite count and HDOP) were considered as potential contributors to measurement variability. While signal quality was generally acceptable, the observed differences in reliability between controlled and ecological conditions suggest that movement complexity plays a more prominent role than signal degradation per se. This reinforces the importance of interpreting reliability outcomes within the specific context of data collection.

Taken together, the three studies converge on a key methodological message: reliability values should not be viewed as fixed properties of GPS or IMU derived metrics, but as context-dependent characteristics influenced by movement demands, testing conditions and statistical methodological choices.

Taken together, the three studies provide converging evidence that GPS and IMU devices particularly the Catapult Vector S7 are reliable tools for quantifying global locomotor loads in football. However, their accuracy decreases when capturing rapid, high-intensity, and multidirectional actions that generate high-frequency inertial data. This discrepancy may be explained by limitations in sensor sampling rates, proprietary data filtering algorithms, or the instability of device placement during complex movement patterns. Furthermore, the consistent consideration of signal quality across the

second and third studies represents a significant methodological contribution, demonstrating that even minor variations in HDOP or satellite connectivity can meaningfully affect data reliability. From a practical standpoint, these findings advocate for standardized monitoring protocols that integrate signal quality thresholds and device-specific calibration procedures to improve measurement consistency.

Beyond methodological implications, the outcomes of this research have clear applications in football environments. Reliable GPS and IMU data are fundamental for optimizing training load management, supporting individualized conditioning programs, and reducing injury risk. Coaches and sports scientists can confidently rely on distance, velocity, and metabolic power derived metrics for daily monitoring, while interpreting acceleration, deceleration and IMU-based variables with greater caution and contextual awareness. Moreover, the insights gained from this doctoral research can inform manufacturers in refining their hardware and software algorithms to enhance data stability under both controlled and ecological conditions.

5.2 Practical implications for football performance monitoring

The practical interpretation of GPS and IMU derived metrics should be grounded in their observed reliability outcomes under both controlled and ecological conditions (**Chapter 3 and 4**). In the present thesis, distance-based GPS variables demonstrated consistently high relative reliability (ICC values typically classified as good to excellent) and low values of absolute reliability across both experimental studies. These findings suggest that metrics such as total distance and speed can be used confidently for routine monitoring in applied football settings, including player and session-to-session comparisons.

Importantly, the relatively low TE%CV observed for these variables suggest that changes exceeding the typical measurement noise are likely to reflect true differences in external load. From a practical standpoint, this supports their use in decision-making processes related to training prescription, load management, and return-to-play progression, particularly when moderate or large changes are expected.

In contrast, variables associated with high-intensity locomotion and rapid changes in movement, including high-speed running and acceleration-deceleration based metrics, exhibited greater variability, especially in the ecological study. While relative reliability remained acceptable in some cases, the higher absolute error observed under football-specific conditions indicates that small fluctuations in these metrics should be interpreted cautiously. These results align with the notion that movement complexity and context substantially influence measurement precision.

IMU-derived metrics provided additional insight into the mechanical demands of football-specific actions, but their reliability was more sensitive to testing conditions. In the second study (**Chapter 3**), several IMU variables achieved acceptable ICC and CV values, indicating that the devices are capable of producing stable measurements under standardized movement patterns. However, in the ecological study (**Chapter 4**), increased variability was observed, likely reflecting the combined effects of multidirectional movement, player-specific execution, and signal quality. As a result, the present findings suggest that IMU-derived metrics may be particularly valuable for characterizing the mechanical demands of drills or sessions at the group level, while their use for detecting small changes over short time frames requires caution.

Overall, the practical application of the present results highlights the importance of selecting metrics not only based on their theoretical relevance, but also on their demonstrated measurement properties in the specific context in which they are applied.

5.3 General limitations of the studies

Across the three studies, several methodological and contextual limitations should be acknowledged.

First, the systematic review (**Chapter 2**) was constrained by the limited number and heterogeneous quality of available studies. Although the COSMIN checklist provided a structured evaluation of measurement properties, its criteria were originally designed for clinical tools rather than sport technology research, introducing a degree of subjectivity in its application. The inclusion of football-only studies further restricted the generalizability of findings to other sports. In addition, the substantial heterogeneity in device models, testing protocols, and statistical approaches prevented the use of quantitative synthesis, and the absence of studies examining responsiveness limited the completeness of the validation framework.

Second, the reliability study (**Chapter 3**) was conducted under artificial conditions using an electric golf cart to ensure standardized and repeatable movement patterns. The placement of the GPS devices on the vehicle may have influenced the results. Specifically, positioning unit D1 on the suspended rear section of the cart may have affected IMU accuracy, as this area is likely exposed to greater vibration than other mounting points. In addition, the thresholds used for certain variables—such as accelerations and decelerations ($\pm 2.5 \text{ m}\cdot\text{s}^{-2}$)—may not have been perfectly suited to the three drills included in the study. The same consideration applies to the predefined thresholds used to calculate distance covered within specific speed zones (e.g., $15\text{--}20 \text{ km}\cdot\text{h}^{-1}$ and $20\text{--}25 \text{ km}\cdot\text{h}^{-1}$). Nevertheless, these issues should not impact the assessment of inter-unit agreement, since all devices were evaluated using identical thresholds.

Finally, although vehicle-driven movement does not fully capture the complexity of football-specific actions, it allowed for a highly controlled testing environment aligned with the primary aim of the study: to examine inter-unit reliability. Future research should investigate this reliability during football-specific movements and across a wider range of real-world training conditions.

Finally, the ecological reliability study (**Chapter 4**) has practical value for coaches, but limitations exist. The small sample size and focus on elite youth players restrict generalizability, and roles positional differences were not considered. Future research should include larger, more diverse football populations, both male and female, across varied training conditions to strengthen findings and improve standardization of GPS/IMU use in sport performance monitoring.

A key limitation identified in this thesis concerns the use of some GPS and IMU derived metrics to compare different devices. Although several variables demonstrated acceptable relative reliability, their absolute reliability, as reflected by the coefficients of variation and standard error of measurement, may limit their sensitivity, especially for HIAD metrics.

For metrics characterized by higher ICC and CV values, small differences between different units are likely to fall within the range of measurement noise. In such cases, interpreting small differences as significant leads to erroneous decision-making. This issue is particularly relevant in high-

performance soccer environments, where staff aim to monitor variations between different players or different exercises while paying attention to HIAD metrics that reflect external load.

Conversely, variables that exhibit lower absolute error in both experimental studies can be interpreted with greater confidence when monitoring these differences. These considerations must be integrated and considered in the results of training and match reports.

Overall, while the thesis provides a robust evaluation of GPS and IMU reliability in football across different levels of experimental control, its findings should be interpreted in light of these methodological and contextual constraints.

5.4 Conclusions and Future Perspectives

This doctoral project provides an integrated evaluation of the reliability and methodological robustness of GPS and IMU technologies used to assess external load in football. The combined findings from the three studies reveal that distance- and velocity-based metrics demonstrate good inter-unit reliability and can be confidently used in performance monitoring. Conversely, acceleration, deceleration, and IMU-derived metrics show higher heterogeneity in absolute and relative reliability parameters (TE%CV and ICC), emphasizing the need for refined signal processing methods and standardized validation protocols. The consistent integration of signal quality measures (HDOP, satellite count) across the experimental designs represents a key advancement, ensuring more accurate and interpretable data.

Future research should aim to expand these findings by including larger and more diverse player samples, exploring intra-unit reliability over repeated sessions, and testing across varying environmental conditions. Furthermore, the responsiveness of GPS and IMU metrics remains a crucial yet underexplored dimension. Longitudinal investigations examining the sensitivity of these measures to training-induced adaptations, fatigue, or match-related performance changes are warranted. Ultimately, enhancing the methodological rigor and ecological validity of wearable technology research will contribute to more precise, data-informed decision-making in elite football performance monitoring.

5.5 References

1. Crang ZL, Duthie G, Cole MH, Weakley J, Hewitt A, Johnston RD. 2021. The validity and reliability of wearable microtechnology for intermittent team sports: A systematic review. *Sports Med.* 51(3):549-565.
2. Scott MT, Scott TJ, Kelly VG. The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review. *J Strength Cond Res.* 2016 May;30(5):1470-90.
3. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. 2018. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 27(5):1147-1157.
4. Malone JJ, Lovell R, Varley MC, Coutts AJ. 2017. Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *Int J Sports Physiol Perform.* 12(Suppl.2):S218-S226.

5.6 List of submitted publications

5.6.1 Articles submitted in peer review journal

1. Di Claudio C., Conte D. , Fanchini M. , Sharp C. , Capranica L. Validation process of GPS and IMU to assess high intensity accelerations and decelerations in football: a methodological systematic review (submitted to “European Journal of Sport Science”)
2. Di Claudio C., Fanchini M. , Sharp C. , Capranica L., Conte D. Inter-unit reliability of Catapult Vector S7 Device GPS and IMU Metrics for Simulated Football-Specific Drills (submitted to “European Journal of Sport Science”)
3. Di Claudio C., Fanchini M. , Palmieri M, Forlini M. , Capranica L., Conte D. Interunit reliability of Catapult S7 GPS during different football training sessions (submitted to “Journal of Sport Sciences”)

5.6.2 Contributes to conferences (Abstracts)

1. Di Claudio C., Conte D. , Fanchini M. , Sharp C. , Capranica L. Validation process of GPS and IMU to assess high intensity accelerations and decelerations in football: a methodological systematic review (presented 29th European College of Sport Science Congress in Glasgow July 2024, Session: CP-AP01 Training and Testing in Sports).