Article

Improving Event Data in Football Matches: A Case Study Model for Synchronizing Passing Events with Positional Data

Alberto Cortez [1,2,3], Bruno Gonçalves [1,2,3,*], João Brito [3] and Hugo Folgado [1,2]

1   Departamento de Desporto e Saúde, Escola de Saúde e Desenvolvimento Humano, Universidade de Évora, 7004-516 Évora, Portugal; alberto.v.cortez@protonmail.com (A.C.)
2   Comprehensive Health Research Centre (CHRC), Universidade de Évora, 7004-516 Évora, Portugal
3   Portugal Football School, Portuguese Football Federation, 1495-433 Oeiras, Portugal
*   Correspondence: bgoncalves@uevora.pt

Abstract

In football, accurately pinpointing key events like passes is vital for analyzing player and team performance. Despite continuous technological advancements, existing tracking systems still face challenges in accurately synchronizing events and positional data accurately. This is a case study that proposes a new method to synchronize events and positional data collected during football matches. Three datasets were used to perform this study: a dataset created by applying a custom algorithm that synchronizes positional and event data, referred to as the optimized synchronization dataset (OSD); a simple temporal alignment between positional and event data, referred to as the raw synchronization dataset (RSD); and a manual notational data (MND) from the match video footage, considered the ground truth observations. The timestamp of the pass in both synchronized datasets was compared to the ground truth observations (MND). Spatial differences in OSD were also compared to the RSD data and to the original data from the provider. Root mean square error (*RMSE*) and mean absolute error (*MAE*) were utilized to assess the accuracy of both procedures. More accurate results were observed for optimized dataset, with *RMSE* values of RSD = 75.16 ms (milliseconds) and OSD = 72.7 ms, and *MAE* values RSD = 60.50 ms and OSD = 59.73 ms. Spatial accuracy also improved, with OSD showing reduced deviation from RSD compared to the original event data. The mean positional deviation was reduced from 1.59 ± 0.82 m in original event data to 0.41 ± 0.75 m in RSD. In conclusion, the model offers a more accurate method for synchronizing independent datasets for event and positional data. This is particularly beneficial for applications where precise timing and spatial location of actions are critical. In contrast to previous synchronization methods, this approach simplifies the process by using an automated technique based on patterns of ball velocity. This streamlines synchronization across datasets, reduces the need for manual intervention, and makes the method more practical for routine use in applied settings.

Keywords: positional data; events data; synchronization; football; match analysis

1. Introduction

Over the years, sport analytics have contributed to a significant shift in the way performance analysis is conducted. While traditional approaches have relied on manual annotation and subjective assessments, advancements in analytics and computing have revolutionized the field [1]. Spectators and elite clubs now have access to a wide range of tools and technologies that facilitate real-time tracking and match analysis [2]. The

possibility of collecting and processing an enormous amount of data during matches has driven a transformation in sports science, enabling more sophisticated and data-driven approaches to evaluate performance [3]. This evolution aligns with broader big-data principles, where the volume, variety, and velocity of information can be harnessed to extract meaningful patterns and insights, ultimately enhancing tactical and performance analysis in football [4,5]. As big data continues to expand in volume and complexity, its principles are increasingly used to transform raw data through advanced analytics and machine learning models, supporting performance optimization and control [6,7]. This influx of data has created new opportunities for the development of advanced metrics and analytical tools with potential to transform the way the game is understood, techniques like expected passes or goals, Voronoi diagrams, or pitch control [8–11]. However, for these techniques to be applied effectively, it is crucial to ensure maximum reliability in data. The use of advanced tracking technologies, standardization of data collection protocols, data validation and filtering processes, integration of multiple data sources, and the application of machine learning and statistical methods all contribute to this goal.

One of the key issues in this field of research is the integration of data from different systems and sources, such as technical event data and positional data. Independent data sources pose a significant challenge in maintaining consistency and accuracy while integrating data [12]. Several challenges impact the accuracy and reliability of synchronizing positional and event data in sports. Previous studies have shown that ensuring precise alignment between these datasets is crucial for meaningful insights [13]. However, issues such as those regarding the collection of both types of data can lead to discrepancies in data alignment [8]. Furthermore, the lack of validation in the synchronization process raises concerns about the reliability of synchronized data [14]. These types of challenges can directly impact the quality and performance of an analysis, even more so for training models on this data [15].

In football, determining the precise moment and location of key events such as passes is crucial for evaluating players and teams' performance. In the modern era of football, where data-driven insights are of extreme importance, pass analysis provides a lens through which teams can fine-tune their strategies and optimize their performance. By analyzing passing activities, teams can identify patterns of play, positional tendencies, and areas of improvement. The significance of passes in football has been explored through two main approaches, notational and experimental studies. More recently, the proliferation of football data has opened new paths for pass analysis, such as the risks and benefits of passes [16], the classification of pass quality [17], and the evaluation of passing effectiveness along with player involvement in creating scoring opportunities [18]. Inspired by these works, meticulous records of all on-the-ball actions such as shots, passes, and tackles became commonly collected across most professional football leagues. Utilizing these events data, numerous studies have conducted pass analyses on a much larger scale than was previously possible with experimental studies. While certain studies have assessed the value of a pass solely using event data [19,20], integrating manually tagged event data with automatically collected positional data allows for a more detailed analysis of the pass's value. Numerous studies have approached this quantification of pass value in various ways, typically evaluating how a successful pass would enhance the probability of scoring [21–23]. Even though manually gathered event data provides valuable insights into individual players during specific ball actions, recent developments in computer vision have enabled accurate tracking of all 22 players and the ball throughout the match; commonly known as tracking or positional data, with this technology, notable improvements have been achieved. [8,24].

Analytics have increasingly focused on the integration of events and tracking data to gain deeper insights into player performance and team tactics. For instance, a fine-grained framework for evaluating the instantaneous expected value of possessions (EPV) [25] has been proposed, revealing that even subtle spatial shifts such as receiving a pass a few meters closer to the center or in a less congested area can significantly change the expected outcome. Building on this foundation, different authors revisited EPV modeling using deep learning approaches and introduced a novel evaluation benchmark that incorporates the reward and risk of individual passes [26]. This demonstrates that pass events are highly sensitive to both timing and location, reinforcing the need to accurately pinpoint these moments to distinguish between high-risk, high reward passes and safer alternatives.

However, existing tracking systems rely on different methods of data capturing, leading to a significant challenge in achieving accurate synchronization between event data and positional data. This spatio-temporal synchronization of positional and event data presents crucial improvement for football analysis, and more specifically of passes. There are authors interested in the problems regarding this synchronization [27]. They emphasized the importance of the synchronization step and referenced existing methodologies, extending the approach used for shot events [9] to the synchronization of passes [8]. However, specific details or evaluations of their implementation were not extensively discussed.

This study explores a new methodology to synchronize event data with positional data, to adjust the precise moments of pass occurrence, using the ball positional data. Regarding this, we propose a custom algorithm that integrates events and positional data gathered from football matches, aiming to synchronize the moment of passing actions, with the event data from that match being presented. To evaluate the results, a dataset was prepared based on the manual identification of all passing actions from the same football match compared with the data from the proposed algorithm and the existing event data.

To address the gap found in the literature, the objective of this study was to develop and validate a simplified and automated synchronization method for aligning positional data with event data. Specifically, the study aimed to (1) reduce the complexity and time required for data synchronization, (2) improve reproducibility and accuracy compared to other methods, and (3) provide a scalable solution adaptable to various data sources and sport contexts.

2. Materials and Methods

2.1. Data Source

For this case study, access to data was provided through the official FIFA Data Platform. Positional data was collected for all players during one match of the 2022 FIFA World Cup using a multicamera computerized tracking system (TRACAB, Chyron Hego, New York, NY, USA), with high-definition cameras operating at 25 Hz. The validity and reliability of the TRACAB systems have been previously established [28]. Original event data was manually collected by trained operators in real time during the match, with both data being provided by the FIFA Data Platform (https://fdp.fifa.org/, 8 November 2023) and representing the official data of the competition.

2.2. Procedures to Synchronization

To synchronize the positional data with the passing event data, an algorithm is proposed to correct the potential delay between events and positional data, thereby improving the identification of the moment of the passing actions. To achieve this objective, the distance travelled by the ball is first calculated, followed by the estimation of its velocity. This enables the detection of speed variation, which in turn allows for the identification of the actual moment the pass occurs.

To accurately identify passing actions, we developed a rule-based algorithm that combines ball tracking data with event annotations system. The process begins by monitoring ball speed to detect potential passes. An event is flagged when the ball's speed exceeds 8 m/s (meters per second) and is preceded by an increase above 5 m/s, ensuring that only sharp and deliberate changes in velocity are considered. To prevent multiple detections of the same event, the algorithm applies a temporal filter: if several speed threshold crossings occur within 0.4 s, only the first instance is retained. In addition to speed thresholds, the algorithm detects moments where the ball undergoes a sudden and significant change in velocity, which often signals a purposeful action like a pass or shot. For each of these moments, a unique sequence ID is generated. To confirm whether the detected action is indeed a pass, the algorithm compares the timestamp of the identified event with the closest "Pass" label from the event data. If the time difference between the two is less than 0.5 s, the event is classified as a valid pass; otherwise, it is discarded. This approach allows for more accurate identification of passing actions by combining mechanical features of ball movement with contextual information from annotated event data (Figure 1).
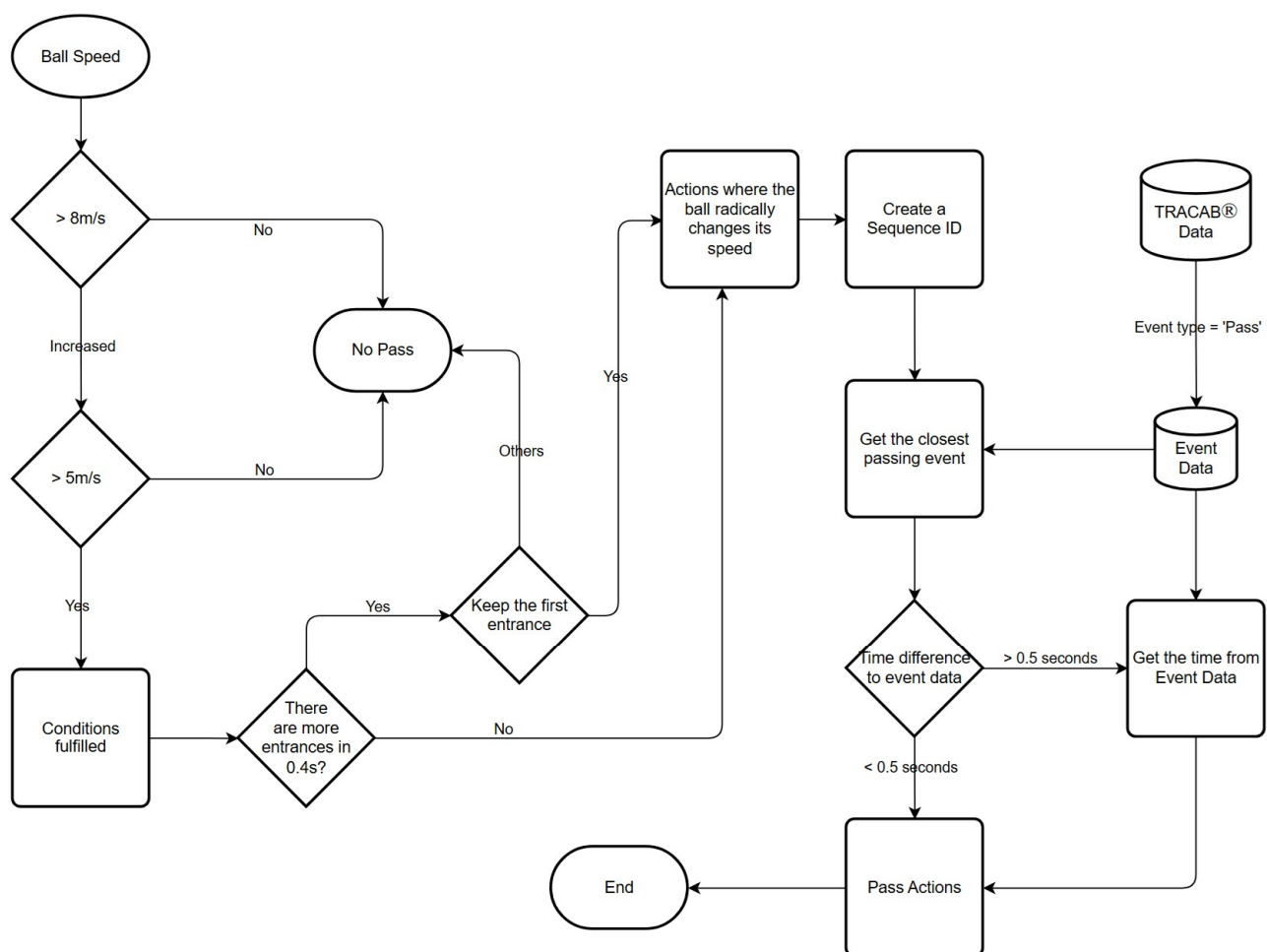

Figure 1. Flowchart of the algorithm.

At this point, a potential pass is identified when the ball's speed exceeds 8 m/s, preceded by an acceleration above 5 m/s, conditions that typically signal a deliberate ball displacement. However, these kinematic thresholds are not exclusive to passes; other actions such as shots, long clearances, or even fast goal kicks may also satisfy these criteria, as actions in which the ball radically changes its speed.

To refine detection, the algorithm implements a two-stage filtering process. First, it eliminates redundant detections by retaining only the first qualifying event within 0.4 s.

Then, to validate whether the detected action is indeed a pass, the algorithm queries the event dataset for the nearest labeled "Pass" event and compares its timestamp to the detected event. If the time difference is less than 0.5 s, the event is classified as a valid pass; otherwise, it is discarded. As such, while the algorithm does incorporate a basic validation mechanism via event matching, its reliance on speed thresholds means that without the event data, it may not reliably distinguish between ball actions with similar mechanical profiles.

2.3. Datasets and Data Treatment

Three datasets were used in this study, based on data from a single match of the 2022 FIFA World Cup. The first was a dataset created by applying the previously presented custom algorithm to synchronize the positional and event data, referred to as the optimize synchronization dataset (OSD). After being exported, the raw positional data from the match was processed in Python 3.8 and divided into two datasets: one for players and one for the ball. Using the ball dataset, the x- and y-coordinates on the field were used to compute the distance the ball travelled between each frame. With this distance, and the known time intervals between frames, the ball's speed was then calculated for each frame. Since the positional and event data were recorded using different time units, a conversion was applied to unify the time scale. Specifically, each event's timestamp (recorded in milliseconds) was adjusted to match the frame-based structure of the positional data (25 Hz, or 40 ms per frame).

The second dataset was a synchronization of the positional and event data to the common time unit, referred to as raw synchronization dataset (RSD). This dataset consisted of a basic integration of the events and the positional data from the match, in which both data types were converted to the same temporal resolution, as presented previously.

Finally, a third dataset, referred to as the manual notational dataset (MND), was used as "gold standard". This dataset was constructed by manually annotating interactions between players and the ball. The annotation was conducted by identifying all instances in which a player received or passed the ball, to precisely determine the moment at which each pass occurred. These actions were registered using an analysis software—LongoMatch Open-Source version 1.3.2—considering the video footage from the match with tactical camera [29]. After annotation, the recorded events were exported as a time series and adjusted to match the temporal resolution of the other datasets, ensuring comparability across all data sources. Manual annotations were conducted by a single expert analyst. While this ensured consistency, the absence of inter-rater validation represents a limitation. Future studies should incorporate multiple annotators and assess reliability to reduce subjectivity.

2.4. Methodology

For comparing the datasets, an inter-method accuracy was calculated by the root mean square error (*RMSE*) and mean absolute error (*MAE*) for each method. Recognizing the susceptibility of certain methodologies to the disruptive effects of outliers, a *modified Z-score* technique was implemented [30]. This approach leveraged the mean absolute error (*MAD*) as a robust measure of dispersion, thereby mitigating the impact of extreme values. By calculating the Z-scores of individual data points relative to the mean and employing the *MAD* as a reference, the method effectively identified and excluded outliers from subsequent analyses. Such meticulous attention to statistical robustness not only enhances

the reliability of the findings but also underscores the commitment to methodological precision [31].

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{(Optimize\ Synchronization\ dataset - Manual\ Notational\ dataset)^2}{n}}$$

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |Optimize\ Synchronization\ dataset - Manual\ Notational\ dataset|$$

$$MAD = \frac{\sum |differences\ between\ datasets - mean\ difference|}{n}$$

$$Modified\ Z - score = \frac{0.6745 \times (differences\ between\ datasets - mean\ difference)}{MAD}$$

The *RMSE* was selected as the metric to quantify the inter-method linear error due to its capability to provide a comprehensive assessment of the deviation between predicted and observed values, considering both the magnitude and direction of errors. On the other hand, the *MAE* was utilized to estimate the average absolute errors between methods, offering a straightforward measure of the magnitude of discrepancies without considering their directionality.

Finally, the spatial differences (i.e., distance of identified event) of the determined location of events of OSD were also compared to the RSD data and to the original data provided by the tracking system provider.

To evaluate the accuracy of the novel technique, both *RMSE* and *MAE* were computed under two conditions: with and without outliers. Outliers were excluded by applying a *modified Z-score*. The comparisons were made between MND and both RSD and OSD, as these results will allow an understanding of the differences when compared to the "gold standard". Finally, a comparison between both synchronized datasets will provide the real differences between procedures for non-manual application.

2.5. Statistical Analysis

To compare the accuracy among the three synchronization methods (Manual Notational Dataset (MND), Optimized Synchronization Dataset (OSD), and Raw Synchronization Dataset (RSD)), a statistical approach based on the distributional characteristics of the data was adopted. Initially, the normality of the variables was assessed using the Shapiro–Wilk test and visual inspection through Q–Q plots. The normality was not verified, and therefore the non-parametric Friedman test was used. Post hoc pairwise comparisons were conducted using Durbin–Conover tests with appropriate correction for multiple comparisons. The rank biserial correlation (r_{rb}) was used as effect size and interpreted with the following thresholds: <0.1 trivial, 0.1–0.3 small, 0.3–0.5 moderate, >0.5 large [32]. Statistical calculations were carried out using Jamovi software 2.4 [33], and the statistical significance was set at $\alpha = 0.05$.

3. Results

To evaluate the robustness of the algorithm to changes in parameterization, a sensitivity analysis was conducted on the ball speed threshold used to segment pass sequences. Thresholds of 4 m/s, 5 m/s, and 6 m/s were tested. At the baseline threshold of 5 m/s, the algorithm identified 1044 pass sequences. Lowering the threshold to 4 m/s resulted in 1100 sequences (5.37% increase), while raising it to 6 m/s reduced the count to 971 sequences (7.00% decrease). Importantly, the total number of passes identified remained constant across all threshold levels due to a protective mechanism within the algorithm

that ensures each pass is always detected. The variation with the threshold reflects the segmentation of each sequence and the exact timing of each pass, as determined by when the ball velocity exceeded the defined threshold. Regarding the differences in the timestamp, the matched passes differed by 23.5 ± 16.9 ms between the 4 m/s and 5 m/s thresholds, and by 33.5 ± 21.7 ms between the 6 m/s and 5 m/s thresholds. These represent mean relative deviations of 4.2% and 6.2%, respectively. This suggests that, while the core event detection is stable, sequence framing and temporal precision are sensitive to parameter changes. This reinforces the need for careful calibration of velocity thresholds to maintain consistency and interpretability in analyses.

The performance of the synchronization algorithm was evaluated by comparing passing events against ground truth labels using a confusion matrix. While the algorithm showed high overall accuracy, several unmatched events were observed (123 events). These mismatches largely stem from the notational dataset containing only completed passes, which omits certain actions necessary for full alignment. The confusion matrix highlights both the algorithm's effectiveness (717 events) and the limitations imposed by incomplete event data (Figure 2).
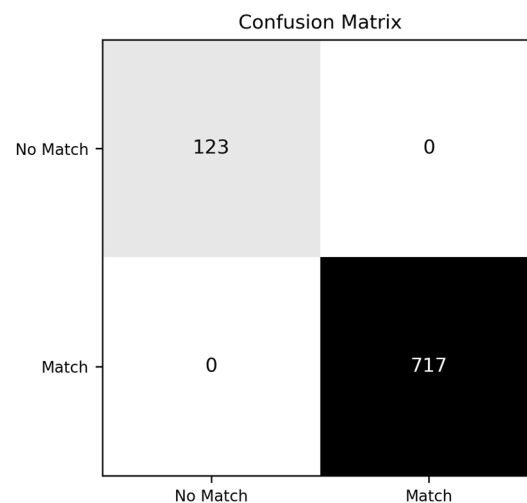

Figure 2. Confusion matrix of synchronized events to notational data.

Comparisons between the notational dataset and the two synchronized datasets revealed significant differences (Friedman test, $\chi^2 = 358$, $p < 0.001$). MND to RSD showed a mean difference of -21.0 ms (95% CI = [-27.0, -15.0]), whereas MND to OSD showed a mean difference of 13.0 ms (95% CI = [-6.5, 19.5]). This analysis represented a small positive effect for both OSD ($r_{rb} = 0.18$) and RSD ($r_{rb} = 0.29$). The results showed a similar *RMSE*, with a lower error for OSD (299.49 ms) compared to RSD (300.46 ms). Similarly, it was identified for *MAE*, with 82.63 ms for OSD and 83.42 ms for RSD. Given the potential influence of outliers on the results, a *modified Z-score* method was employed to identify and exclude anomalous values. This procedure resulted in the removal of 8 outliers from the RSD (-1192.63 ± 2481.37 ms) and the OSD (-1181.63 ± 2482.65 ms) dataset. While these data points were statistically identified as outliers using Z-scores, it is acknowledged that their origin has not been fully explored. The outliers presented show similar values for both datasets and could reflect data anomalies or measurement errors. The removal of these outliers helped stabilize variance and improve the robustness of subsequent analyses. Normality of the remaining data was then assessed using the Shapiro–Wilk test. The results improved for both *RMSE* (RSD = 75.98 ms; OSD = 73.53 ms) and *MAE* (RSD = 61.13 ms; OSD = 60.39 ms), indicating a modest but meaningful enhancement in synchronization accuracy. The reduction in *RMSE* suggests fewer large alignment errors, which is particularly important in high-tempo game scenarios where even slight timing discrepancies can lead to

misinterpretation of player actions or physical outputs. Similarly, the lower *MAE* reflects a more consistent alignment across all data points, enhancing the reliability of time-sensitive performance metrics. These improvements, while relatively small in magnitude, contribute to greater confidence in the temporal accuracy of the data and support the use of the proposed method.

Additionally, as these previous results demonstrated an improvement when compared to the "gold standard" for the optimize method, the same procedure was applied to analyze the OSD data in comparison to the RSD with a mean difference of 42.8 ms (95% CI = [37.5, 48.2]). The results showed small differences between the two datasets with 47.08 ms of *RMSE* and 33.81 m/s of *MAE*. The same method was applied without considering outliers (9 outliers were removed with 127.44 ± 156.12 ms), with results of 41.58 ms of *RMSE* and 31.97 ms of *MAE*.

An essential component of the analysis involved examining ball speed over time and identifying the precise moment when a pass occurred. Figure 3 illustrates the ball speed plot during 30 s of a game, highlighting fluctuations and where six passes occur during this window.
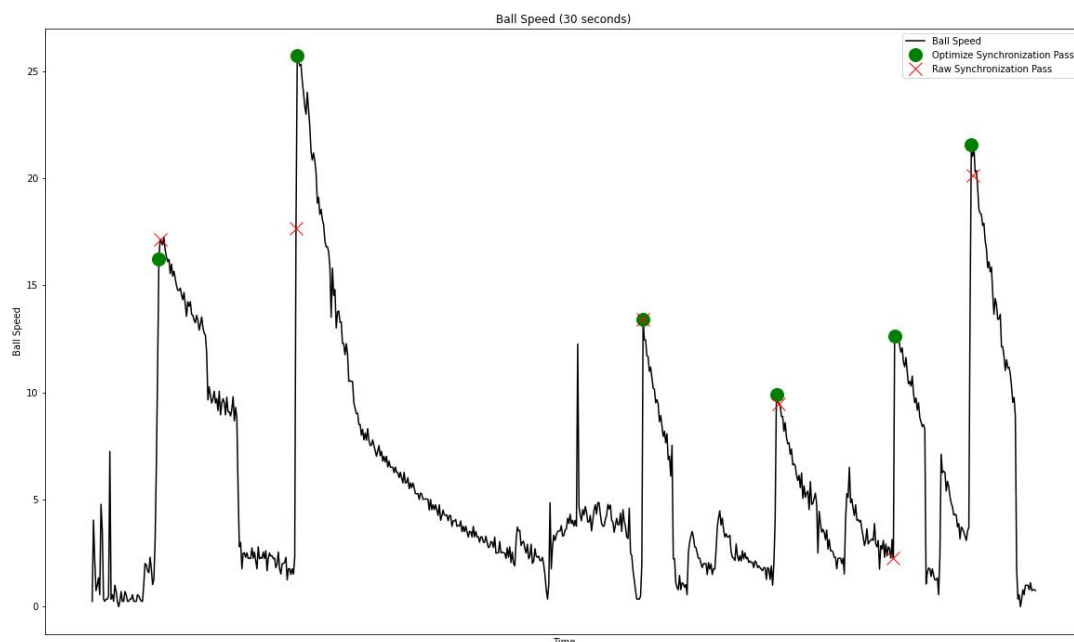

Figure 3. Ball Speed with the specific moment where the pass occurs in OSD and RSD.

After the results scored in the previous tests, where an improvement by the OSD was observed, it was important to comprehend the pass moments in both datasets (OSD and RSD). A ball speed plot was created, and the specific point where the pass happened was marked for both synchronized datasets.

The presented Figure 3 showed that the novel technique appears with a more coherent definition of the moment of the pass in OSD when comparing to RSD. It was marked frequently in the moment when the ball reaches the most speed in that sequence, where the RSD had fluctuations in the position where it's marked.

After comparing the benefits of the algorithm regarding time then it was important to compare the data regarding their locations in meters (m). The spatial accuracy of pass events was also examined by comparing pass locations recorded in the OSD against those from RSD and raw event data. In this context, three different reference points were used to evaluate pass location. First, the OSD refers to the moment identified by the algorithm based on a sharp increase in ball velocity, typically interpreted as the likely initiation of a pass. Second, the RSD marks the location at which the algorithm officially classifies an

action as a pass, integrating both ball movement characteristics and temporal alignment with event data. Finally, the raw event data corresponds to the timestamp provided by the provider dataset, representing the annotated pass location without necessarily capturing the precise physical initiation of the action. Figure 4 presents a football field visualization of pass locations across datasets, illustrating areas of convergence and divergence.
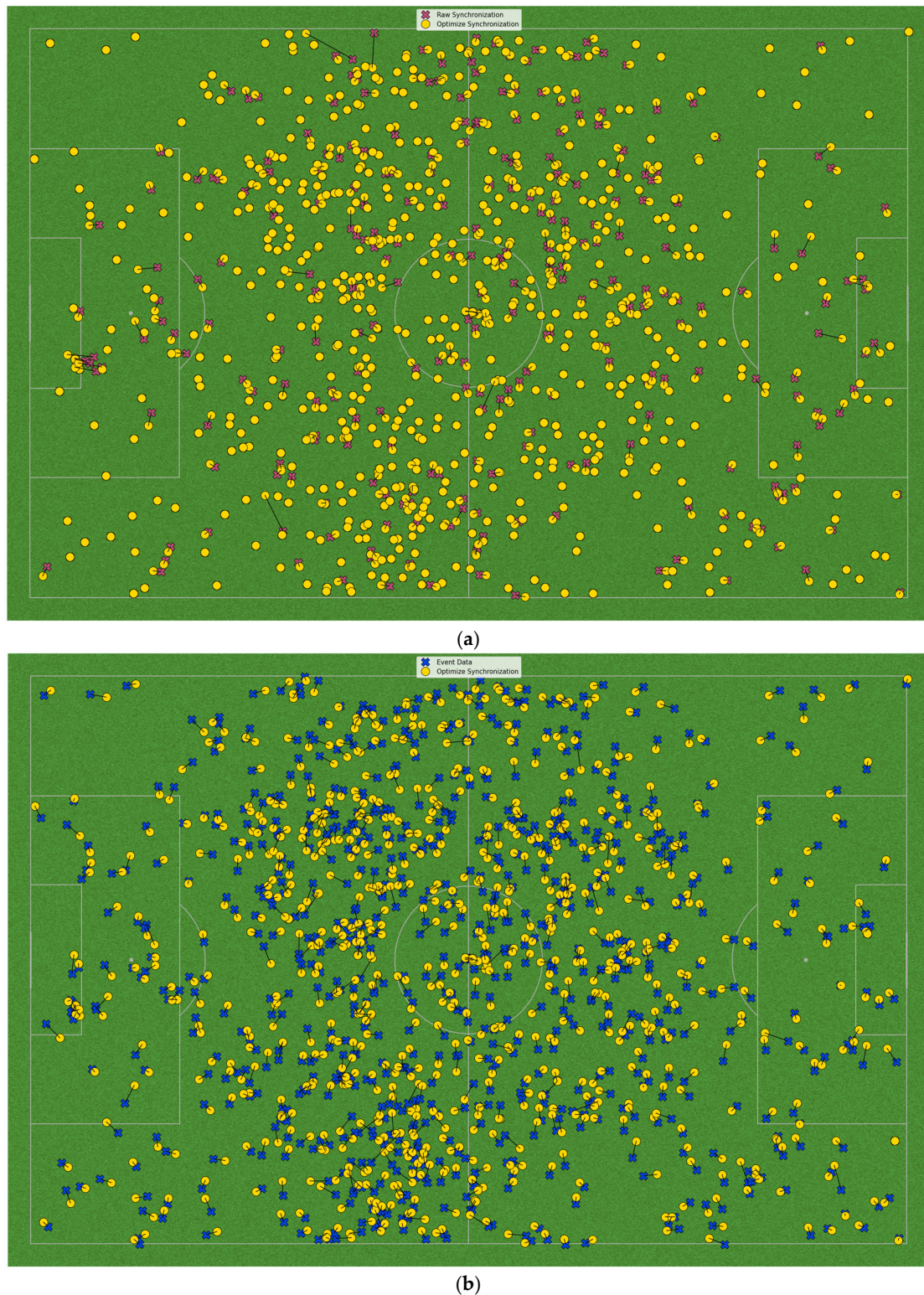

(**a**)

(**b**)

Figure 4. Pass Locations Comparison (**a**) OSD (yellow dots) compared to RSD (red crosses); (**b**) OSD (yellow dots) compared to Event Data (blue crosses)).

Quantitative results indicate a mean positional deviation of 0.41 ± 0.75 m (95% CI = [0.36, 0.46], *p* < 0.001) (Figure 4a) between OSD and RSD data (*RMSE* = 0.861 m, *MAE* = 0.410 m) with most differences in the range of 0 to 1 m, but with 108 passes within the range of 1 to 2 m. Additionally, between OSD and event data (*RMSE* = 1.785 m, *MAE* = 1.586 m) there is a mean positional deviation of 1.59 ± 0.82 m (95% CI = [1.53, 1.64], *p* < 0.001) (Figure 4b), with most differences within the range of 1 to 2 m (475). The larger discrepancy in event data likely results from lower spatial resolution and manual annotation errors inherent in event-based tracking.

4. Discussion

Recent advances in data integration approaches have addressed a long-standing issue: data source independence. As has been pointed out, ensuring consistency and accuracy during the integration of different data is extremely difficult [12]. To improve the integration of different data sources, this method leverages ball speed and event data accuracy to enhance the harmonization. By incorporating precise measurements of ball speed and meticulous event data, this method ensures that inconsistencies are identified and corrected. This approach not only enhances data consistency and accuracy but also allows a more refined and reliable integration process, effectively mitigating the issues previously encountered. Additionally, this approach presents a validation within the synchronization process to address the concerns regarding the reliability of synchronized data [14]. This mechanism involves a validation framework that utilizes ball speed metrics and event accuracy data to cross-reference synchronized data against notational dataset to verify its integrity.

One notable constraint of this methodology is its reliance on the accuracy of positional data. Despite advancements in tracking technology over the past decade, the precision of ball tracking remains a topic lacking comprehensive validation within existing literature [8]. The synchronization of spatio-temporal data between positional and event datasets, often collected through distinct systems, is essential for enhancing the analysis of passes. This study addresses this challenge by combining the intrinsic value of the positional data, using this to understand the moment where the ball increases speed, and the accuracy of the event data, through the identification of the player and the approximation of the time of the action.

While the temporal enhancement discussed in this paper offers valuable benefits, the advancement in spatial accuracy remains of paramount importance. Frequently, analyses rely on positional locations from event data, Brooks and colleagues [19] constructed a model that predicts shot opportunities based in pass origins and destinations, while Bransen and Van Haaren [20] proposed to value each pass by computing the difference between the values of the possession sequence before and after the pass, and [16] through a supervised learning approach. The study proposes a methodology to estimate the risk and reward of all passes, with the intention of enhancing the understanding and analysis of the game of football. Recently, several authors have developed models to evaluate possessions where the moment and location of the pass influence greatly the results, revealing that these models are highly sensitive to subtle spatial displacements [25,26]. Hence, improving the precision of spatial data significantly contributes to the robustness and reliability of analytical findings and modeling outcomes in football research.

A previous study discussed the development of a novel approach to pass synchronization [9], building upon the foundation laid by their previous work [8], although the framework requires further clarification. This model aims to bridge the gap between theoretical development and practical implementation. This could enhance accessibility and facilitate replication/validation by other researchers but also fosters collaboration and

advancement within the field of football analytics. Moreover, by improving the synchronization of passes, this model contributes to the refinement and accuracy of analytical insights derived from tracking data, thus enhancing the overall understanding of football dynamics.

An important methodological limitation of this study lies in the algorithm's reliance on ball velocity derived from positional tracking data. Since the detection of pass events is triggered by sharp changes in ball speed, any inaccuracy in ball position (whether due to tracking noise, interpolation artifacts, or system latency) can directly affect velocity calculations. These errors may lead to both false detections (e.g., identifying a pass where none occurred) and missed events, particularly in situations involving subtle or low-velocity passes. Given that the ball tracking system may vary in precision depending on factors such as camera calibration, occlusions, or frame rate, the robustness of the algorithm is inherently tied to the quality of the positional data. This dependency should be considered when interpreting the results, especially in comparative studies across different data providers or match contexts. Future improvements could involve integrating contextual variables (e.g., player–ball distance, directionality) to reduce reliance on velocity thresholds alone.

5. Conclusions

This paper presented a novel technique for synchronizing event data with positional data, addressing a critical challenge in the integration of separately collected data. Through comprehensive evaluations of both temporal and spatial outcomes, this approach demonstrated significant improvements in alignment accuracy. The results underscore the importance of precise synchronization, particularly for models that typically consume such data without any preprocessing or correction, and will facilitate a more accurate analysis of player interactions, team strategies, and subsequent game dynamics. Improving the reliability of input data directly enhances models that critically depend on precise temporal and spatial information. Future work may explore the generalization of this technique across domains with varying data resolution and structure.

Author Contributions: A.C. and H.F.; Responsible for conception and design of the project. A.C.; Data collection. A.C., B.G. and H.F.; Performed data analysis, interpreted the data. A.C., B.G., J.B. and H.F.; writing, review and editing. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available on the FIFA Data Platform, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data is available from the authors upon reasonable request and with permission of Portuguese Football Federation and FIFA. If someone wants to study further with the data, please contact the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Gudmundsson, J.; Horton, M. Spatio-Temporal Analysis of Team Sports. *ACM Comput. Surv.* **2017**, *50*, 1–34. [CrossRef]
2. Link, D.; Lang, S.; Seidenschwarz, P. Real Time Quantification of Dangerousity in Football Using Spatiotemporal Tracking Data. *PLoS ONE* **2016**, *11*, e0168768. [CrossRef]

3. Memmert, D.; Lemmink, K.A.P.M.; Sampaio, J. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Med.* **2017**, *47*, 1–10. [CrossRef]

4. Goes, F.R.; Meerhoff, L.A.; Bueno, M.J.O.; Rodrigues, D.M.; Moura, F.A.; Brink, M.S.; Elferink-Gemser, M.T.; Knobbe, A.J.; Cunha, S.A.; Torres, R.S.; et al. Unlocking the Potential of Big Data to Support Tactical Performance Analysis in Professional Soccer: A Systematic Review. *Eur. J. Sport Sci.* **2021**, *21*, 481–496. [CrossRef]

5. Gandomi, A.; Haider, M. Beyond the Hype: Big Data Concepts, Methods, and Analytics. *Int. J. Inf. Manage* **2015**, *35*, 137–144. [CrossRef]

6. Jaspers, A.; De Beéck, T.O.; Brink, M.S.; Frencken, W.G.P.; Staes, F.; Davis, J.J.; Helsen, W.F. Relationships between the External and Internal Training Load i Professional Soccer: What Can We Learn from Machine Learning? *Int. J. Sports Physiol. Perform.* **2018**, *13*, 625–630. [CrossRef]

7. Cortez, A.; Trigo, A.; Loureiro, N. Football Match Line-Up Prediction Based on Physiological Variables: A Machine Learning Approach. *Computers* **2022**, *11*, 40. [CrossRef]

8. Anzer, G.; Bauer, P. A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Front. Sports Act. Living* **2021**, *3*, 624475. [CrossRef]

9. Anzer, G.; Bauer, P. Expected Passes: Determining the Difficulty of a Pass in Football (Soccer) Using Spatio-Temporal Data. *Data Min. Knowl. Discov.* **2022**, *36*, 295–317. [CrossRef]

10. Emanuel, B.; Figueira, N.; Alexandre, D.; Coutinho, M.; Caldeira, N.; Lopes, R.J.; Fernandes, D.; Araujo, D. From Optical Tracking to Tactical Performance via Voronoi Diagrams: Team Formation and Players' Roles Constrain Interpersonal Linkages in High-Level Football. *Sensors* **2022**, *23*, 273. [CrossRef] [PubMed]

11. Routray, A.; Sivakumar, N.; Hur, S.H.; Bang, D.J. A Comparative Study of Optimal Individual Pitch Control Methods. *Sustainability* **2023**, *15*, 10933. [CrossRef]

12. Fernández, J.; Barcelona, F.C.; Fernandez, J.; Bornn, L. Wide Open Spaces: A Statistical Technique for Measuring Space Creation in Professional Soccer. *Sloan Sports Anal. Conf.* **2018**, *2018*, 1–19.

13. Arbues-Sanguesa, A.; Martin, A.; Fernandez, J.; Ballester, C.; Haro, G. Using Player's Body-Orientation to Model Pass Feasibility in Soccer. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2020, Seattle, WA, USA, 14–19 June 2020; pp. 3875–3884. [CrossRef]

14. Pettersen, S.A.; Johansen, D.; Johansen, H.; Berg-Johansen, V.; Gaddam, V.R.; Mortensen, A.; Langseth, R.; Griwodz, C.; Stensland, H.K.; Halvorsen, P. Soccer Video and Player Position Dataset. In Proceedings of the 5th ACM Multimedia Systems Conference, MMSys 2014, Singapore, 19–21 March 2014; pp. 18–23. [CrossRef]

15. Brefeld, U.; Lasek, J.; Mair, S. Probabilistic Movement Models and Zones of Control. *Mach. Learn.* **2019**, *108*, 127–147. [CrossRef]

16. Power, P.; Ruiz, H.; Wei, X.; Lucey, P. Not All Passes Are Created Equal: Objectively Measuring the Risk and Reward of Passes in Soccer from Tracking Data. In Proceedings of the Knowledge Discovery and Data Mining 2017, Halifax, NS, Canada, 13–17 August 2017; pp. 1605–1613. [CrossRef]

17. Horton, M.; Gudmundsson, J.; Chawla, S.; Estephan, J. Classification of Passes in Football Matches Using Spatiotemporal Data. *ACM Trans. Spat. Algorithms Syst.* **2014**, *3*, 6. [CrossRef]

18. Bransen, L.; Van Haaren, J.; Van De Velden, M. Measuring Soccer Players' Contributions to Chance Creation by Valuing Their Passes. *J. Quant. Anal. Sports* **2019**, *15*, 97–116. [CrossRef]

19. Brooks, J.; Kerr, M.; Guttag, J. Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 49–55.

20. Bransen, L.; Van Haaren, J. Measuring Football Players' on-the-Ball Contributions from Passes During Games. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11330 LNAI, pp. 3–15.

21. Goes, F.R.; Kempe, M.; Meerhoff, L.A.; Lemmink, K.A.P.M. Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches. *Big Data* **2019**, *7*, 57–70. [CrossRef]

22. Gómez-Jordana Martín, L.; Ric, Á.; Amaro Silva, R.; Gómez-Jordana, L.I.; Milho, J.; Silva, R.; Passos, P. Landscapes of Passing Opportunities in Football-Where They Are and for How Long Are Available? In Proceedings of the Barça Sports Analytics Summit, Camp Nou, Barcelona, 13 November 2019.

23. Steiner, S.; Rauh, S.; Rumo, M.; Sonderegger, K.; Seiler, R. Outplaying Opponents—A Differential Perspective on Passes Using Position Data. *Ger. J. Exerc. Sport Res.* **2019**, *49*, 140–149. [CrossRef]

24. Stein, M.; Seebacher, D.; Marcelino, R.; Schreck, T.; Grossniklaus, M.; Keim, D.A.; Janetzko, H. Where to Go: Computational and Visual What-If Analyses in Soccer. *J. Sports Sci.* **2019**, *37*, 2774–2782. [CrossRef]

25. Fernández, J.; Bornn, L.; Cervone, D. A Framework for the Fine-Grained Evaluation of the Instantaneous Expected Value of Soccer Possessions. *Mach. Learn.* **2021**, *110*, 1389–1427. [CrossRef] [PubMed]

26. Overmeer, T.; Janssen, T.; Nuijten, W.P.M. Revisiting Expected Possession Value in Football: Introducing a Benchmark, U-Net Architecture, and Reward and Risk for Passes. *arXiv* **2025**, arXiv:2502.02565.

27. Spearman, W.; Basye, A.; Dick, G.; Hotovy, R.; Pop, P. Physics-Based Modeling of Pass Probabilities in Soccer. Research Papers Competition. In Proceedings of the 11th MIT Sloan Sports Analytics Conference, Boston, MA, USA, 3–4 March 2017.

28. Linke, D.; Link, D.; Lames, M. Football-Specific Validity of TRACAB's Optical Video Tracking Systems. *PLoS ONE* **2020**, *15*, e0230179. [CrossRef] [PubMed]

29. Folgado, H.; Bravo, J.; Pereira, P.; Sampaio, J. Towards the Use of Multidimensional Performance Indicators in Football Small-Sided Games: The Effects of Pitch Orientation. *J. Sports Sci.* **2019**, *37*, 1064–1071. [CrossRef] [PubMed]

30. Crawford, J.R.; Howell, D.C. Comparing an Individual's Test Score against Norms Derived from Small Samples. *Clin. Neuropsychol.* **1998**, *12*, 482–486. [CrossRef]

31. Yaro, A.S.; Maly, F.; Prazak, P.; Maly, K. Outlier Detection Performance of a Modified Z-Score Method in Time-Series RSS Observation With Hybrid Scale Estimators. *IEEE Access* **2024**, *12*, 12785–12796. [CrossRef]

32. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Routledge: New York, NY, USA, 1988.

33. The Jamovi Project. Jamovi. (Version 2.4) [Computer Software]. 2023. Available online: https://www.jamovi.org (accessed on 2 August 2025).