Thank you for the opportunity to read this thesis from an applied sports science perspective. The work comprises a series of chapters which address computational pipelines that could be utilised in the spatio-temporal analysis of team invasion sports. I would like to congratulate the candidate for taking on the challenges associated with bridging the gap between computer science/engineering and professional sport. Collectively, the thesis comprises a body of work around a central theme however, I recommend a number of changes that I ask that the candidate and supervisory team to consider.

Specifically, throughout the individual chapters of the thesis the advancement in the field of either computer science or sports science is not clear. Each chapter of the thesis should explicitly highlight or state the usability and/or validity of the individual computational platforms. In its current format it is difficult to ascertain the volume of work that has been undertaken and the functionality of the individual components. This aspect could be addressed through further work that explores the application of each platform or potentially a significantly increased analysis and breakdown of the case study within chapter 7. If chapter 7 is to be explored further, this should be integrated throughout each chapter of the thesis.

Throughout the thesis the connection between computer science and sports science needs to be addressed. Whilst this has been described within the abstract, such relationship is not clear throughout the thesis. As such the individual chapters of the thesis must be revised in order to highlight the sports science/data analysis limitations that have been overcome through the development of the computational pipeline. Furthermore, how this work has advanced the fields of both computer science and sports science must be further stated throughout each chapter.

From a sports science perspective, whilst the de-identification process is novel there are inherent flaws in down sampling data to 1HZ. Specifically, it has been well elucidated that microtechnology devices sampling at 10HZ have an increased accuracy when compared to 1Hz and 5HZ devices. As such this must be addressed throughout the thesis and at a minimum highlighted as a limitation. Unfortunately, this level of data (i.e. 1HZ) is not likely to be utilised by sports scientists within either an applied or research based setting. As the effective under sampling created by this process discards useful data that is required in the determination of athlete movement patterns and associations between physiological responses, performance and injury occurrences. With this in mind, the thesis should address the impact that the data de-identification process has upon data integrity and therefore, the ability of such data to be utilised in both research and practice.

The literature overviews at the start of each chapter are somewhat not clear to a lay reader and to a degree detract from the work of the thesis and the contribution to the field. In my opinion the significant emphasis and reliance on existing literature should be reduced, so that the candidate can further explain and emphasis their work, methods, findings and contributions to the field. As previously mentioned at times these do not appear to be explicitly stated or clear.

The thesis title is “computational pipelines for spatio-temporal analysis of team invasion sport”. In my opinion the utilisation of chess, backgammon and albatrosses throughout the thesis detracts from the ability of this work to be utilised in sport. With this in mind, I would suggest that such examples be altered to have a focus upon team sports or the title of the thesis be made broader.

Finally, I am not discrediting the inherent complexities in the computer science processes that have been developed however, to the best of my knowledge commercial platforms exist that address the integration of datasets (e.g. GPS, video and game statistics). As it stands the ability of the platform to further address real-world problems that provide evidence-based information for the preparation and

management of athletes is unclear. Therefore, as per my previous comments the novelty, validity and useability of the processes developed throughout this thesis need to be further highlighted.

General comments

- Throughout the thesis there are inconsistencies in the formatting of references, with some references being reported as [1] and others being reported as [1]. Furthermore, superscript references are being reported at the bottom of pages throughout the thesis. Please go through the thesis and ensure that all references are formatted consistently.
- Please go through the thesis carefully and correct typographical and formatting errors (including the utilisation of "we" within some chapters and not others as well as inconsistencies in the use of capital letters and the use of "z" and "s"). I have noted several errors within the current version of the thesis, but you should be sure that the final version is without such errors.
- Throughout the thesis there is inconsistent use of terms such as "2D" and "2-dimensional". Such inconsistencies should be revised throughout the thesis.
- Please review the thesis and ensure that all acronyms presented within the body of work and figures/tables are defined.
- Please review the formatting of the tables presented within the thesis. At times the text appears to be spread and has significant gaps between words (potentially this is due to the utilisation of "justified" formatting).
- Within chapter 7 it becomes apparent that the de-identification process that has been devised throughout this PhD is associated with a website (deidentify.org). This should be made clear and highlighted throughout the relevant de-identification chapter. Such output supports and strengthens the real world application of this chapter and associated model.

Specific comments

Chapter 2

Collectively, this chapter conducts a review of the technological and analytical developments that have led to the increase of spatio-temporal data in sport.

- In my opinion this chapter is lacking sufficient references to support statements. Furthermore, the chapter is lacking depth and fails to adequately review and synthesise existing literature. Perhaps this could be addressed through reducing the scope of the review?
- Figure 2.2: does not appear to be a complete or thorough timeline and therefore, presents a limited overview on the introduction of positioning devices within sport.

Chapter 3

Collectively, this chapter presents on models that could be used to describe the state of a game and the performance feedback process. Despite this at times the chapter reads like a high level review. Whilst the novel processes proposed and the resulting applications/outcomes within this chapter are somewhat unclear.

- Page 52, paragraph 1: In order to strengthen this section, it would be beneficial to provide further details on the qualifications/experience of the sport researchers.

- Page 60: f is listed within the figure description but appears to be missing from the presented equation.

Chapter 4

This chapters discusses the foundations for describing the computational pipelines for sport analysis. Despite this at times the chapter reads like a high level review. Whilst the novel processes proposed and the resulting applications/outcomes within this chapter are somewhat unclear.

Chapter 5

Collectively, this is an informative chapter that utilises both case studies and data to highlight issues with data de-identification. Throughout this chapter it would be beneficial to provide further detail on the model that you have developed/proposed to improve the ethical conduct of research. Such detail would facilitate the use of such methodology by other sporting codes, teams and researchers.

5.3.4

- Page 147, paragraph 3 (study selection): it may be beneficial to reference or mention the two studies which examined the same dataset.

5.3.5

- Page 152, paragraph 2: it may be beneficial to include a reference to support the statement that "this appears to be due to journal page limitations which have the effect of preventing authors revealing too much….."
- Page 153, paragraph 2: it may be beneficial to further explain why the risk of de-identification increases when multiple papers utilise the same dataset.

5.4

- Page 155, paragraph 2: although it is somewhat alluded to throughout the remaining paragraph, it is important to acknowledge that IMU's provide greater insight than just complementing GPS position for short distances.

5.5

- This section specifically alludes to an "attacker" being able to re-identify data, but would this concept not also hold true for a researcher or research assistant who may have access to de-identified data?

5.5.2

- Page 163: within figure 5.9, it would be beneficial to further define the equation.

5.6

- Table 5.6: it may be beneficial to include further information within this table. For example it would be beneficial for researchers and practitioners alike to understand the effect that each de-identification method will have on data integrity and their ability to undertake analysis at both the individual player and team level.

5.7.3

- Page 169, paragraph 2: it would be beneficial to include the names of specific researchers or work within the following sentence: “In our work, we look at the techniques proposed by [51,50] to introduce….”

5.7.7.

- Page 175, paragraph 3: it would be beneficial to outline what “upload encrypted data” is referring to (i.e. what is encrypted data).
- It would be beneficial to outline or at least mention the potential data integrity issues when down sampling GPS data to 1 HZ. Does such down sampling hinder the data analysis process and subsequent results?
- Given that the process was rather time consuming, is this de-identification process something that could be adopted within practice?

Chapter 6

Collectively, this chapter proposes a novel method for representing spatio-termporal reference frames as geographic objects. Although this method may be translatable to other fields such as animal tracking (as noted within the chapter summary), a stronger focus/emphasis on the application of such method within sport would help to strengthen this chapter.

6.2.1

- Equation 6.2.1: f is listed within the figure description but appears to be missing from the presented equation.
- Equation 6.2.1: b is listed within the figure description but appears to be missing from the presented equation.

6.2.3

- Figure 6.5: it would be beneficial to define each of the symbols utilised within this figure.

6.3.6

- Page 207, paragraph 1: it would be beneficial to define “t” within “…. Each containing x,y,t fields that…..”.

6.3.7

- Throughout the study/PhD was the model evaluated or tested by someone without a computer science or engineering background. This would greatly support the conclusion that the “system solves the fundamental problem of enabling non GIS experts to define and make comparisons…….”.

Chapter 7

This chapter integrated the techniques proposed in earlier chapters to construct a computation pipeline for AFL analysis. Unfortunately, in my opinion the strength of the findings presented within this chapter are somewhat hindered by the very small sample size of matches that were utilised in the analysis. As such, the specific findings highlighted within this chapter require careful consideration and revision.

7.1.1.

- Page 214, paragraph 1: it would be beneficial to outline the number of matches that were collected/utilised.

7.1.3

- It is apparent that the de-identification process that has been devised throughout this PhD is associated with a website (deidentify.org). This should be made clear and highlighted throughout the relevant de-identification chapter.

7.1.5

- It appears that figure 7.8 is mentioned within text prior to figure 7.7. If this is the case, it would be beneficial to consider changing the order of the respective figures.

7.2.1

- This section on related work feels out of place and detracts from the current chapter. Please consider removing this section or more subtly integrating it throughout the thesis.

7.2.3

- Figure 7.6 states that there were 6 matches with useable GPS data whilst pages 233 and 237 state that 5 matches were utilised. As such this section related to the number of utilised matches is somewhat unclear.

7.2.6

- Figure 7.23: it would be beneficial to further describe aspects of the plots (e.g. red crosses, blue and black lines).
- Table 7.2: it would be beneficial to describe each of the symbols utilised within the table headings.
- Page 245, paragraph 1: it would be beneficial to further justify the use of forward distance only. Are there any other research outputs that utilise such method or variable? Surely, the total distance throughout a play also contributes to game speed?
- Page 247, paragraph 1: is it possible that the findings of this applied chapter are limited due to the utilisation of a very small sample of games (n=5). Indeed, if a larger sample of matches was utilised it is likely that the results may be different, as the current results could be skewed due to a range of factors including but not limited to player availability (i.e. players not available for selection due to injury) and game strategy.