Deakin University Higher Degrees by Research thesis examinations
Response to Examiners' Reports

Candidate name	**Andrew J. Simmons**
Thesis title	Computational Pipelines for Spatio-Temporal Analysis of Team Invasion Games (Previously: Computational Pipelines for Spatio-Temporal Analysis of Team Invasion Sport)
Degree	Doctor of Philosophy (Information Technology)
School	Applied Artificial Intelligence Institute (A^2I^2)
Date	June 2019

Examiner Key:
B = Examiner 1, A/Prof. [redacted] (Statistician / Head Performance Analyst)
C = Examiner 2, Dr. [redacted] (Applied Sport Scientist)
A = Examiner 3, **Anonymous** (Computer Scientist)

Exam-iner	Examiner's comment	Response	Location or page in revised thesis
		General	
C	Throughout the individual chapters of the thesis the advancement in the field of either computer science or sports science is not clear.	The contributions of the initial chapters, while motivated by sport, are written from a computer science / software engineering perspective, and thus best judged by an examiner with a background in computer science or software engineering. The application to sport is covered in Chapter 7. I have added a new paragraph to the "Thesis structure" section to clarify which fields the chapters contribute to. As the thesis already summarises the contributions at the end of each chapter, I have extended these to outline opportunities for future work to improve the evaluations and explain how the advancements could be utilised by sport scientists.	Thesis Structure (p.11) End of Chapter 3,4,5,6 (p.77, p.122, p.183, p.219)
A	This thesis encapsulates a rare cross-disciplinary research outcome and makes several worthwhile contributions to data processing pipelines, sports data analysis and geo-spatial data analysis. ... The evaluations undertaken can all be improved and at a minimum effort should be made within this thesis on how these evaluations can be undertaken.		

C	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.	Each component is functional. I have included additional screenshots and demos: - Ch2 Background (Not applicable) - Ch3 Modelling (Theoretical) - Ch4 Computational Pipelines (Applied in Chapter 7) - Ch5 De-identification (added screenshots, demo: deidentify.org) - Ch6 Spatio-Temporal (added screenshots, demo of my "XYT" tool as extension to geojson.io GIS editor: demo.visualmodel.org/gps_xyt/#demo - Ch7 Platform (utilises work built in other chapters to demonstrate sport platform).	p.182, p.218
C	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 the examiners have opposing views regarding the organisation of the thesis and appropriate level of level of literature support, I have decided to leave the literature reviews/surveys for each chapter as part of the thesis. However, I have moved the less relevant parts of the related-work section in Chapter 7 to an Appendix.	pp. 238-239, Appendix D.2
A	The thesis demonstrates strong written communication skills, the work is well organised, the references are grounded and complete (to the context), and the work is presented in a writing style that is quick to parse		
C	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.**	The examples from other domains are intended to demonstrate the generality of the approach taken. In light of the examiner's suggestion, I have decided to rename the thesis from "Computational Pipelines for Spatio-Temporal Analysis of Team Invasion **Sport**" to "Computational Pipelines for Spatio-Temporal Analysis of Team Invasion **Games**" to reflect its applicability to abstract games and scenarios beyond sport.	Title Page, Cadidate Declar- ation

C	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)	Existing systems combine GPS and video as an animation. My approach goes further than this by integrating GPS, video, and game statistics at a deep level and showing that it can be used for team analysis. This is highlighted by the Motivating Scenario (Section 4.2.3) that explains pain points with current video annotation tools due to their lack of full data provenance capture. Section 6.2.4 "Support for Coordinate Transformations within Sport Player Tracking Software" reviews existing commercial platforms and highlights their limited ability to truly integrate information. Feedback from analysts at Geelong Football Club (Section 7.3) confirms that existing commercial systems (including systems only available to clubs that are not publicalicised) are lacking compared to the system that I demonstrated to them.	
C	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.	The framework is for strategic team-level analysis, not conventional management of individual players. The inability to analyse data at the individual player level is a direct consequence of de-identification to preserve player privacy (a core aspect of this thesis) These trade-offs are discussed in Section 5.6 "Analysis of Trade-Off between Participant Privacy and Data Quality" I have introduced a new Limitations subsection within the De-identification Chapter Summary (Section 5.8) that reiterates the sacrifice of individual player analysis in exchange for privacy preserving team-level analysis.	p.181
B	The club itself will need something in return! So this is a rather delicate juggling act and I am not sure that I am convinced of this, especially given Chapter 7 details some nice broad statistics, but nothing down the level of detail I am aware of in clubland.		

C	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.	Footnotes are used for comments and further references (e.g. URLs for further background information), whereas references are used to cite peer-reviewed content / books. This convention is common in computer science / software engineering literature, but may be unfamiliar to readers in the sport science / medical field.	
B	Numbers less or equal to ten in words. Easy fix. The use of 's' instead of 'z'.	The use of 'z' and 'we' occurred in sections relating to my international publications, where US spelling is more common. I have revised the thesis for consistency to always prefer Australian/UK English. In some cases I have deliberately retained US spelling (e.g. quotes and code snippets).	All
C	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").		
C	Throughout the thesis there is inconsistent use of terms such as "2D" and "2-dimensional".	Fixed	All
C	Please review the thesis and ensure that all acronyms presented within the body of work and figures/tables are defined.	Done (where appropriate)	All
C	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).	I have re-formatted the narrow table columns in LaTeX to use "raggedleft" in order to improve the spacing.	Table 5.1, 5.2, 5.3 (pp. 152-154), Table 5.6 (p.168)
Chapter 1			
B	I was expecting some insight into the rationale for de-identification, as the author does make it a focal point. From a coaching/analyst's perspective this would be a rather big hurdle in a pragmatic sense, and I believe that there needs to be, even if brief, some mention of why this is an issue… I know this covered in great detail in Chapter 5, but you could talk about it, or refer to it, here	I have included a cross-reference to the more detailed discussion in Chapter 5.	p.6

B	AFL is mentioned first on page 10, in the Considerations, without its explanation, which appears first in Chapter 2. Needs to be detailed i.e. AFL (Australian Football League) or replaced with Australian Rules Football upon its first mention.	Fixed. I have reworded to avoid the use of "AFL" until introduced in Chapter 2.	pp. 9-10
Chapter 2			
C	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?	I have renamed the literature review to "Evolution of player tracking devices for use in Australian Rules Football" to clarify its scope. It is not intended to be a major contribution in itself, its primary purpose is to provide background to frame the rest of the thesis.	p.18
3	The work proceeds carefully by setting the context of prior technological and analysis methods in Chapter 2. This chapter is quite long and can be trimmed with a more expansive Appendix. Although the content of the chapter itself is helpful to a reader with minimal background in this area, the knowledge captured is a summative exposition and not critical to the structure of the overall thesis.		
C	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.	Fig 2.2. was captioned "Timeline of Positioning Devices in Sport". I agree that this was misleading, as it only relates to developments relevant to AFL. I have renamed the figure to "Timeline of Geopositioning Devices for AFL analysis".	Fig 22.2 (p.37)
B	p.13. Please add ["traditionally" / "typically"] as it is not always the case that a pass is always a player decision. Coaches can advise for specific connections to occur despite a player's intuition. Furthermore, tactics are often driven from players on the field – this point needs to made as I feel the author has a fairly compartmentalised approach to coach and player roll	The purpose of these statements was to show that decisions can be made top-down from the coach, bottom up from the players, or a combination of both. I provided common examples, but in no way were they intended to suggest that a particular action can only be top down / bottom up. I've accepted the examiners suggestion to clarify this as traditionally/typically.	p.13

B	p.14 – 'thus this form of communication..." I am not sure about this statement – but I understand the idea – I assume you mean the coach cannot say anything as it happens, however direction from a coach is possible in terms of things to do next. **If this the case all good**	The examiner's interpretation is more-or-less as intended.	
B	p.14 – "Coaches provide video...their own style". This is a generalised perception and speaking from experience there are multiple utilities of video feedback. This needs expansion... [detailed breakdown of how video is used to review play/set up in team invasion games] Don't assume it is always done this way. Clubs don't always disclose what they do as that is their competitive edge.	The context of this section (p.14, paragraph 2) is for closed skills (such as Olympic target shooting) rather than team invasion games (discussed later). While I acknowledge that coaches / sport analytics software providers may have deeper analysis methods that they don't share for competitive reasons (p.35), the thesis background can only review what is publicly known rather than speculating on what clubs may or may not be doing in practice.	
B	p.15 "Currently some coaches use a range of summary statistics..." Statement as written is asserted as if law and this is definitely not true.... I simply don't agree. In my experience at clubland, summary statistics are broken into team based and individual based parameters and if the coach/players aren't aware that these are products of performance, not causative, then there is certainly a problem. You are probably attempting to state that stats out of context, or without some spatial reference, are somewhat misleading. That is certainly true.... Please rewrite this paragraph to assert this point, or rejoinder my argument.	I have reworded the statement to clarify that it is the "traditional" approach, and incorporated the point raised by the examiner.	p.15

B	p.16 “Teams head in ...” Define that a bit better. (ie play towards differing ends). You could easily use some visuals to help clarify this section.	I have reworded this sentence to be more precise.	p.16
B	You also receive 1 point if you hit the post, or rush a behind...these little details are completed nicely in the Ryall PhD thesis if you need ideas.	The introduction to the game of AFL in this thesis has been carefully written to discuss *only* the aspects that are relevant to the thesis (in particular to prepare an unfamiliar reader for the analysis in Chapter 7). It is not intended to be a full description of the game.	
B	Also, please calculate the occurrence of draws – shouldn’t be hard to evaluate.	Done. I’ve calculated them by running a query over public match data. Draws represent just 1.03% of all matches.	p.17
B	p.20 the goal assists paragraph – I am again uncertain I agree…. how can this be a bad incentive? … Please rewrite or remove	Measuring handballs and goal assists can incentivise a player to touch the ball for the purpose of improving their metrics without actually contributing in any meaningful way. Anything other than rewarding the team for scoring goals has the potential to result in a misalignment of incentives. However, I’ve decided to take the examiner’s advice to remove this paragraph due to lack of strong literature support.	p.20
B	p.22 and p.29 The Clarke model and commentary around gambling etc is important…. Beating a gambling market really is not important to teams	The examiner’s comments seem to be agreeing with Gap 1: “While it is trivial to compare accuracy of predictive models for betting purposes, these do not necessarily deliver insight to players and coaches”	
B	p.40 GPS v Video indoors – This is not quite true. RF systems indoors do provide detailed positional based info too, and are often suitable over video systems due to occlusion. However, video systems have the distinct advantage of not requiring the athletes to wear anything which is a big deal in codes that do not have rules around everyone wearing the same tech (ie netball) and information sharing.	I’ve clarified the statement to say “GPS is *typically* better suited...” and provided additional explanation as per the examiner’s comments.	p.40
B	p.13 – optimizing (inconsistent use of s and z)	Fixed (minor)	p.13

B	p.25 - 2 minute – change to two minute	Left as is, as minutes are a unit of measurement. “Always use numerals for numbers written with units. Otherwise, spell out numbers below 11" (IEEE Editorial Style Manual for Authors, 2019)	
Chapter 3			
C	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.	The information theoretic perspective itself is not novel, but the application to sport performance is new. While not practical to utilise at this stage, it is included as a foundation for further work and is used as a lens to motivate the rest of the thesis. I’ve added a future work section at the end of the end of the chapter outlining the existing limitations of the information theoretic approach and how they could be addressed in future.	p.77
3	The information theoretical perspective is an interesting lens in terms of motivating how performance analysis works within sports. Although this is theoretically interesting, the underlying knowledge is not directly novel. This aspect can be improved by providing a stronger motivation of how this lens can be better utilised either in practice or how this can lay the foundation of further works.		
B	This chapter develops the model framework for the thesis through the definition of solutions. This was a well-developed chapter and was impressive in its breadth and care. I thought 3.3 was excellent; and 3.3.1 was nice in its idea around unknowns.		
C	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.	These were Prof. Paul Gastin (Head of Sport & Exercise Science, LaTrobe University) and Dr. Clare MacMahon (Senior Lecturer Health Sciences, Swinburne University). I’ve reworded my statement to clarify that this was intended to describe the iterative nature of model development and acknowledge the input of sport researchers rather than	p.52

		to serve as expert validation of its correctness.	
C	Page 60: f is listed within the figure description but appears to be missing from the presented equation.	'f' and 'g' are arbitrary functions. I have expanded the definition of the function composition operator to explicitly state this.	p.60
B	p.66 the talk of holistic approach – I was a little confused here. The knowledge of a team performance would help inform player feedback if the system was holistic? So is the proposal inclusive of performance variables?	I've deleted this statement, as its framing as a limitation was confusing. A clearer framing of the same concept is already present in Section 3.3.6 when discussing future work.	p.66
B	p.68 I thought AFL used T6/clearsky which is RF &/or GPS, but I could be wrong! You should say this is at time of publishing too. T6 are 100 Hz?	The examiner is correct that the AFL currently uses Catapult Clearsky (e.g. the sample data shared by Hawthorn Football Club was at 100Hz). However, the dataset provided by Geelong Football Club was from the 2015 season, which utilised older 10Hz GPS devices at the time. The data was exported at 5Hz, then further down-sampled to 1Hz to prevent re-identification. The thesis specifies the device sample rates and technology, but doesn't mention the particular model because there were a mix of devices in use by different clubs during the 2015 season.	
B	p.56 – "This is represented.... Actions take **place**"	Fixed (typo)	p.56
B	p.56 - "...stochastic process". Please refer to the PhD thesis by Don Forbes – "Dynamic prediction of Australian Rules football using real time performance statistics"	Don Forbes' thesis focuses on the details of creating a Markov Model of AFL, but does also contain some discussion of the rationale for why the game should be considered a stochastic process. I have decided to cite it as per the examiner's request.	p.56
B	p.62 – AFL equity ratings – reference?	Have already cited Karl Jackson's work when introducing AFL equity rankings	p.62

		in Chapter 2. However, I've decided to re-cite it again here for clarity.	
Chapter 4			
C	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.	While this chapter is motivated by the needs of the sport domain, the contributions in this chapter are described from a computer science and software engineering perspective. The practicality of the notation in evidenced through the use of the notation within Chapter 7. This chapter already undertakes a detailed theoretical analysis. I have added a future work section detailing the planned development of further software tooling to support practical applications of the notation and the need for an empirical user study.	p.122
B	This chapter was polished and a nice intro into the methods needed to achieve the desired outcomes expressed in chapter 3. A lot of this chapter was new to me so I was quite impressed by the W3C idiom		
A	Chapter 4 is a deep and thought out unit of work. Surprisingly, the content here has not been put for publication as this will make a great stand-alone contribution by itself, and I strongly encourage that the candidate considers this recommendation. ... This chapter should aim to present a short outline of an evaluation protocol as part of further work to be done.		
Chapter 5			
B	I am a bit perplexed (and this might be at the heart of the problem!) as to how this [data de-identification] can be managed. If you obtain consent from players etc. and have an open environment then it is case closed! ... If the data isn't released, and the analysis is completed in house, then there is a layer of protection? ... I am still coming to terms with the removal of the levels of player-based information, however I get it if this is for researchers outside of a clubland situation wishing to play with team data.	This chapter deals with the case of a club attempting to share data with an external researcher. This results in a more complex situation than when the analysis is all conducted in-house. I have added a new paragraph to the introduction of the chapter to clarify the scenario.	p.125

C	Page 147, paragraph 3 (study selection): it may be beneficial to reference or mention the two studies which examined the same dataset.	This was already indicated by the references in Table 5.1. I have modified the text to explicitly state the two studies here as well.	p.150
C	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....."	I have reworded the statement to "may be due to journal page limitations" to acknowledge the speculative nature of this statement, and removed overly general claims. The risk of re-identification occurs when two publications publish similar statistics for a slightly different subset of participants, thus allowing inference of participant identities. I have added a reference to a CSIRO report describing "differencing attacks" in support of my argument.	pp. 155-156
C	Page 153, paragraph 2: it may be beneficial to further explain why the risk of [re]-identification increases when multiple papers utilise the same dataset.		
C	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.	I have added a footnote to acknowledge the examiner's remark that IMUs can provide insights beyond just complementing GPS position, and to discuss sensor fusion.	p.158
C	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?	Yes, "attacker" is anyone that plans to re-identify the data. This includes an (unethical) researcher or research assistant. This is why I went to lengths to ensure that nobody (including myself) would be able to re-identify the data. This has already been covered in Sec 5.4, Threat Model ("attacker or unethical researcher") as well as in the Sec 5.3, Prevalence of Improper De-identification Methods, that goes on to explain the importance of preventing re-identification attacks "even if one does not deliberately attempt to undermine the de-identification scheme" (p.145)	
B	p.148-151 Greenham paper – Did they acknowledge Champion Data as there are other places one could get that data (ie ProWess, AFL Tables, or independently)	Yes. Greenham et al. 2017 state "These included eight variables routinely measured and reported by **Champion Data**".	

C	Page 163: within figure 5.9, it would be beneficial to further define the equation.	This 'equation' is used conceptually (hence why listed as a figure rather than an equation). It is only included to help remind the reader that sets are invariant to the order elements are listed, as per the figure caption.	Fig 5.9 (p.166)
B	The mixed sets do clearly make sense, as does a downsizing in resolution.		
C	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. ... 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?	In contrast to traditional sport science analysis that focuses on monitoring players at an individual level, this thesis analyses sport from a high-level team strategy perspective, hence the choice of a lower sampling rate compared to other studies. The rate of 1Hz was sufficient for the high-level team strategy analysis conducted in Chapter 7. Note that it would have been possible to utilise a higher sampling rate (up to the capability of the device); however, for the purpose of the thesis, a rate of 1Hz was decided in order to ensure strong de-identification taking into account the needs of the application. I have included a footnote to clarify that any threshold is possible and that the sampling rate chosen for strategic analysis of AFL in this thesis is not appropriate for all sports or contexts. I have also added a new limitations sub-section at the end of the chapter to elaborate on these issues and how future work could help address them.	p.169, p.181
B	You are losing huge amounts of information. This is probably ok in an AFL context as it is an open sport with anyone allowed to do anything; however, this would not work that well for say Netball with specific roles and zones endemic to the rules of the game.		
A	Given that the protocol developed has been used and evaluated in practice there is confidence that it is helpful. Although not a gold standard, the primary efficacy seems to be there and hence it can be deemed to be a worthwhile contribution upon which further refinement is possible.		

C	The effective under sampling created by this [de-identification] 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 thesis studies data at a team level, not at the individual level. I discuss the limitations as part of the discussions in Sec. 5.6, "Analysis of trade-off between Participant Privacy and Data Quality", and highlight that the ability to analyse individual players' movement patterns runs counter to the goals of team-level analysis of de-identified data.	
B	Was happy with the way you argued the limitations.		
C	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.	As examiners had conflicting views, I have left the presentation of the de-identification model as is. This section (the interaction model) is based on one of my publications, so has already been peer-reviewed.	Sec. 5.7 (pp. 170-180)
B	The proposed interaction model in 5.7.1 is very nice and tidy, and the model proposed is excellent. It kind of reminds me of the current catapult openfield/cloud system, whereby you have differing access levels with different outcomes (although this is much better!)		
C	Given that the process was rather time consuming, is this de-identification process something that could be adopted within practice?	Without this process, it would likely not have been possible at all. For industry projects where the client has commercial incentive to have the project completed within a short deadline, or there are multiple people who can authorise, they might reply to requests faster. As noted in the protocol future work section, there may be ways to reduce the number of iterations needed. This is already discussed in Sec. 5.7.8.	p.180

A	The chapter can be improved by outlining briefly, how the protocol can be further refined and improved or areas that need specific future attention.	I have added a limitations and future work section.	pp. 181-184
B	p.136, 138: two instead of 2	Fixed (minor)	p.138, p.141
B	p.152: 2 of 3 – two of three	Fixed (minor)	p.143
C	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.	The effects are summarised by the "Analysis Constraints" column of the table. I have moved the column to the right hand side to make it more obvious. The approach is only for team-level analysis -- by design, the de-identification approach prevents individual player analysis.	Table 5.6 (p.168)
B	Section 5.6 is really well done. I like the process and it makes heaps of sense...		
C	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...."	The thesis follows IEEE citation style, which does not require stating the researcher name in-text. However, for clarity, I have added the researcher's surname (Kissoon Curumsing).	p.173
C	Page 175, paragraph 3: it would be beneficial to outline what "upload encrypted data" is referring to (i.e. what is encrypted data).	I have modified the text to clarify that the "encrypted dataset" is an encrypted version of the raw identifiable data (in this case the high frequency player GPS tracking data)	p.179
C	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.	I created the deidentify.org website and backend. I have added screenshots of the finished web-based tool to make this clear.	p.182

Chapter 6			
C	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.	Examples from animal tracking demonstrate the generality of the approach. Applications to sport are covered in Chapter 7. I have added a new "Applications to Sport" subsection to Sec. 6.4, and a screenshot of an interactive demo at: https://demo.visualmodel.org/gps_xyt/#demo	p.218
C	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.	Both f and b are both used, albeit not directly (the formula uses 'e', 'e' uses 'b', and 'b' uses 'f') I've re-ordered the definitions in the equation to make this more apparent.	Eq. 6.2.1 (p.193)
C	Figure 6.5: it would be beneficial to define each of the symbols utilised within this figure.	Meaning of ticks (supported), crosses (not supported) and wavy lines (partial support) is as one would intuitively expect. I've added a key for purposes of formality.	p.200
B	p.194 footnotes precede the text on the next page	Fixed (minor).	p.203
C	Page 207, paragraph 1: it would be beneficial to define "t" within ".... Each containing x,y,t fields that.....".	't' stands for time. The same sentence indicates that it is "relative offset time from the start of the reference event". I've updated the text to clarify x,y,t are *respective* to the definitions that follow.	p.214
B	Section 6.3 needs a tidy up as it reads like a paper rather than a part of the thesis – this is covered in the annotations [(see below)]	It is based on a paper I published (and is acknowledged accordingly, with an Authorship Statement provided). I've performed additional editing to improve consistency with the style of the rest of the thesis.	Sec 6.3 (p.205 -)
B	p.199 sport analyst then sport performance analyst – changing titles? Please pick one	Fixed (minor). Changed to "Sport performance analyst" rather than shortening to "Sport analyst"	p.206
B	p.199 Use of the word 'we'	Fixed (minor).	Sec 6.3

B	p.199 “In this paper...”	Fixed (was based on a paper I published)	Sec 6.3
C	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.....”.	This is based on a theoretical argument in terms of Bloom’s taxonomy to evaluate the level of learning necessary. I’ve clarified the need for an empirical user study with sport performance analysts as future work.	p.219
A	Chapter 6 is a novel method that shows the value in treating spatio-temporal reference frames as geographical objects. The novelty and value of this work is validated by the SIGSPATIAL work arising from this unit of research.		
Chapter 7			
C	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. ... 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.	While limited, the largest existing published study at the time consisted of only a single match. Note that while only five games were used, each game consists of many events. I have added a new "Limitations" subsection within the Chapter 7 Summary (Sec 7.4) to acknowledge the threats to validity due to the availability of only a small number of matches, and added a "Future Work" subsection to discuss how this could be addressed in future.	pp. 258-259
B	On a technical note, the sampling is very poor in the end. To have only six matches of data is a big concern!		

C	Page 214, paragraph 1: it would be beneficial to outline the number of matches that were collected/utilised.	I have reworded this section to remove the ambiguity. Rather than state the number of matches, I've referred the reader to the selection flow diagram (Fig 7.6) in Sec. 7.1.4 for details (as the number differs vastly between initial number and the number after removing matches with data issues)	Sec. 7.1.1 (p.221)
B	p.214 for 2015 matches ... rewrite this as it sounds like a quantity not a year		
C	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.	I have reduced this section to focus on the most relevant related work and moved the review of team spread metrics to Appendix Sec. D.2.	Sec 7.2.2 (pp. 238-239), Appendix Sec. D.2 (pp. 325-237)
C	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.	There were six matches of potentially usable GPS data, but only five home team matches. The rationale for only utilising the home team matches is provided in the Experiment Design subsection. I have edited the Datasets subsection (within Sec. 7.2.3) to clarify this, and added a footnote to point the reader to the relevant sections for further details.	p.240
B	Either the users aren't adept at using it or the tech is inadequate or the players aren't wearing it properly or a combination as you outline nicely in Fig 7.6.		
C	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.	The red crosses, blue and black lines in the bean plot (violin plot) represent median, mean, and data points respectively (generated by Python Statsmodels). I have added a caption to explain. The t-test result table columns (e.g. 't', 'p') follow conventions. I've expanded the caption to explicitly state the type of test and the meaning of all symbols. I have also updated the table to report results of Welch's t-test rather than the standard t-test to avoid the assumption of equal variance.	Fig. 7.23 (p.250), Table 7.1 (p.251) Table 7.2: (p.252)
A	This chapter can be improved by captions and legends in the diagrams.		
C	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.	Fixed (minor layout issue). I have reworded the text to introduce figure 7.7 earlier.	p.228

C	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?	The intent of this section is to examine new team-level metrics, such as team spread, rather than traditional approaches that focus on individual players or distance. The related work in Sec. 7.2.2 and Appendix Sec. D.2 shows that multiple techniques are found in previous literature, and that previous literature has not yet settled on a standard. The intent of the forward distance measure is just as a proxy of how close the team are to scoring a goal in order to provide additional context to the analysis of team spread. Other measures, such as "apparent width" (cited), are indeed possible alternatives for forward distance. However, re-running the analysis with alternative definitions/features is outside the scope of minor revisions.	p.241
B	p.217 Now I have just read about the process and importance of de-identification yet you have done precisely that yourself. You mention the football club and the club numerous times, yet you de-identify the team as Geelong (I think!)... This whole chapter talks about the club etc. Either **identify it as Geelong** (or whichever club it is) or remove the information that leads me at least to think it is Geelong. It is hypocritical to be critical of other works in Chapter 5 for doing this when this occurs in this chapter – so please make a choice either way.	Examiner is correct that the primary team worked with in this thesis is Geelong Football Club. Chapter 5 is about de-identification of *players*, not the club. The aim of Chapter 5 is not to criticise individual authors of other works, but rather point out general issues with current practices. Based on the examiner's feedback, I have decided to explicitly state that the club is Geelong (as this can easily be inferred). Note however, that no individual player data is identifiable. I have added a footnote to note that "while the name of the club is identified here (as it would not be practical to suppress it), none of the individual player data (beyond public video footage) are identifiable, even to myself"	p.221

B	From experience, wouldn't it be better to use end-on footage rather than broadcast? Most coaches I have worked with rarely to never use broadcast	The club could utilise the tool with "behind the goals" (end-on) footage (or any other footage) if desired. I've added a new paragraph about this in Sec. 7.1.2 (Video Data). The club had access to behind the goals footage (end-on footage) of the team. However, for the purpose of this thesis, I only used non-identifiable data (e.g. de-identified GPS traces) or public data (e.g. broadcast). Behind the goals footage cannot (reasonably) be de-identified and is not public.	Sec. 7.1.2 (p.223)
B	The team spread is interesting – I can't help but think that you could look at the outcomes as a function of the matched paired spreads too. [Detailed proposal] This would describe the interactive nature of this index as it happens, rather than a group- based approach you have taken. Whilst differences were found, I wonder if this is not of value given that this is not considered for. The location as a reference starting point is value, however as the data is detailed one should be able to deduce this.	In principle yes; however, it would require a larger dataset to detect the subtle influence on goals (if any), particularly when broken down at a fine granularity. Further, I only had access to one team's data, not the opposition (as clubs are generally unwilling to share their data with other clubs). Chapter 8 (Conclusions) already includes some discussion along these lines in Sec 8.3 (Future Work). Specifically, Sec. 8.3.3 (Two Team Perspective) and Sec. 8.3.4 (Information Gain). But as noted, this would require more data.	
B	[Game speed and Team Spread] shows us some nice results... What is notable is that whilst both are linear, the defence gradient is gentler than attack. Also, the quantity of goals scored is more of a cluster than a correlation? Maybe you could address this in the text.	Given the limited size of the dataset (5 matches), it is important not to draw too much from figures other than general trends. I've decided not to include this suggestion as it would be speculating beyond what can be deduced with statistical significance.	
B	p.250 First paragraph is pretty close to the mark. I think that the How is now really important. You have demonstrated it is possible, however I wonder if there is still a way to go to make this accessible, dynamic and interactive as or near-as the play goes.	I envision an interactive exploratory environment (similar to Jupyter notebooks) to explore hypotheses layered on-top of the pipeline developed in this thesis. A prototype of this was presented to the club (Sec. 7.3). I've included some additional discussion in the Chapter 7 Summary of Future Work (Sec 7.4).	p.259

B	p.217,8 you use numbers for words when you should use words – anything 10 or less should be written in words – please correct throughout this chapter.	Fixed (minor)	Ch 7, all
B	p.241 during the goal kick – should that be goal attempt?	Fixed (minor)	p.248
Chapter 8			
B	p.256 Sounds good, but I am sure the enemy doesn't wear GPS and share	This is in the context of war simulation exercises. Both the 'blue' (friendly) and 'red' (enemy) teams are tracked and analysed after the exercise to extract insights into the cause of mission success/failure in order to help prepare for a real scenario. Even in real warfare scenarios, it may be possible for intelligence to (partially) track the enemy.	
Appendix C			
	p.313 – That is a shocker! But one could draw similar conclusions about this thesis, Deakin being aligned with Geelong so be careful	My conclusion that "removing the name of the club from the published paper provided only superficial privacy," is intended to highlight problems with current norms, not criticise the individual author for attempting to do so. As per the examiner's feedback, I've decided to explicitly state that the data used in this thesis was for Geelong Football Club, but not to reveal any information about individuals (beyond what is already public).	
Overall			
C	I would like to congratulate the candidate for taking on the challenges associated with bridging the gap between computer science/engineering and professional sport.	N/A	
B	It is a large piece of work, with detailed writing across a number of disciplines highlighting the process needed to align data, anonymise, work it over and add a hint of analytics at the end.		
A	… a rare cross-disciplinary body of work.		

Other Changes			
Geelong Football Club	If an acknowledgement is required, can it be a generic "Geelong Football Club" rather than any individual please.	I have removed the names of individual sport performance analysts at the football club from the Acknowledgements, and acknowledged the club instead.	Acknowledgements
Geelong Football Club	... We assumed that teams would be spread wide in good attack and more condensed in good defence. More explanation ...	I have clarified the sport performance analyst's position based on the email from Geelong Football Club.	p.238, p.254
Geelong Football Club	We don't believe there's anything we'd object to overall.	N/A	Chapter 7
Self		Removed reference to smart-home draft paper (rejected, re-submission to new venue in progress). Fixed minor self-identified typos.	Sec. 8.2.2