

Spatio-temporal metrics that distinguish play outcomes: A pilot study

CIARÁN MCINERNEY , SIMON GOODWILL, LEON FOSTER, SIMON CHOPPIN


Centre for Sports Engineering Research, Sheffield Hallam University, Sheffield, United Kingdom

ABSTRACT

In team invasion sports, tactical behaviour can be examined using spatio-temporal data, i.e. the position of the players at a given time. A review of the spatio-temporal metrics used in team invasion sports performance analysis indicated that thousands of variations of metrics being used. Information about the distribution of metrics' individual effects can inform us of the best variable-selection method. The aim of this pilot study was to estimate the distribution of strong marginal effects of spatio-temporal metrics of field hockey plays. With institutional ethical approval, the Womens' and Mens' gold medal games from the EuroHockey 2015 field hockey tournament were recorded. Best, acceptable and worst-case outcomes for plays were described by 1,837 spatio-temporal metrics. Each metric's marginal effects were estimated using Cramér's V, Mutual Information and the I-score. Values for Cramér's V of 0.2 and 0.4 to mark the boundaries of small, moderate and large effects. Less than 1% of metrics show large effects with > 87% of all metrics showing small effects as per the Cramér's V thresholds. These large effect metrics were all within the 98th percentile of Mutual Information values and within the 96th percentile of the I-score values, which supports the Cramér's V distribution of marginal effects. Therefore, according to the recommendations of Tibshirani (1996), univariate variable-selection methods will be the most appropriate for selecting important metrics. **Key words:** TEAM INVASION SPORTS, VARIABLE-SELECTION, MARGINAL EFFECTS.

Cite this article as:

McInerney, C., Goodwill, S., Foster, L., & Choppin, S. (2017). Spatio-temporal metrics that distinguish play outcomes: A pilot study. *Journal of Human Sport and Exercise*, 12(2proc), S502-S507. doi:<https://doi.org/10.14198/jhse.2017.12.Proc2.03>

 **Corresponding author.** Centre for Sports Engineering Research, Sheffield Hallam University, Collegiate Campus, Sheffield, S1 1WB, United Kingdom.

E-mail: c.mcinerney@shu.ac.uk

6th ISPAS International Workshop, 22-23 March 2016. International Society of Performance Analysis of Sport. Carlow, Ireland.
JOURNAL OF HUMAN SPORT & EXERCISE ISSN 1988-5202

© Faculty of Education. University of Alicante

doi:10.14198/jhse.2017.12.Proc2.03

INTRODUCTION

The goal of sports performance analysis is to create a competitive advantage by providing useful advice to athletes and support staff. To achieve this the analyst must appropriately measure and evaluate behaviour. There is no shortage of metrics to measure behaviour (Karcher and Buchheit, 2014; Lanzoni et al., 2011; Podgórski and Pawlak, 2011) and an abundance of data is being collected in some high profile sports like association football and basketball. Analysts must select the most appropriate metrics for their purposes; available resources limit the amount and type of data which can be collected.

Team invasion sports are characterised by two teams attempting to score points by invading their opponent's territory, e.g. football sports, hockey sports and basketball (Read and Edwards, 1997). A team's behaviour is often quantified using metrics relating to player locations at a given time (Gudmundsson and Horton, 2016). This spatio-temporal data is easily translated into many metrics that describe the point, line and surface relations of players and the ball over time. The number of spatio-temporal metrics quickly becomes large as the number of players and the duration of play increases, and as newly applied analysis methods produce novel metrics.

Analysts need a method to help decide which metrics are most important. Methods can either evaluate metrics based on their individual effects, also known as marginal effects, or based on their combined effect with other metrics, also known as interaction effects. It is important that the methods used reflect the true relationships in the dataset because each method works best under certain assumptions. For example, selecting metrics based on their marginal effects would assume that metrics' relationships to performance are independent. In contrast, selecting metrics based on their interaction effects allows for the idea that metrics might be measuring the same or similar relationships. Selecting metrics based on marginal effects is typically less computationally expensive than using interaction effects so it might be preferred when resources are constrained. The disadvantage is that, although one metric might show a strong effect, there might be a combination of metrics that provide greater information. The subtleties of the interaction effects might make the important difference in the top end of elite sport performance where differences between teams are small.

Tibshirani (1996) suggested some guidance for deciding which methods to use. If there are an obvious subset of inputs that strongly relate to the output then univariate selection methods work best. Haws et al. (2015) suggested that this could be evidenced by having a small number of strong marginal effects. Examples of univariate variable-selection methods include measures of marginal effects or interaction effect methods like minimum-redundancy-maximal-relevance (Peng et al., 2005; see Ng and Breiman, 2005 for discussion on univariate and bivariate variable selection methods). If, on the other hand, there is no obvious subset of inputs that clearly and strongly relate to the output, then sparse regression methods, such as LASSO regression (Tibshirani, 1996), are best. This scenario could be evidenced by moderately amount variables with strong and/or moderate marginal effects. The first step toward appropriately selecting metrics is therefore to estimate the distribution of strong effects. The aim of this pilot study was to estimate the distribution of strong effects of spatio-temporal metrics of field hockey plays.

MATERIAL AND METHODS

Data Collection

With institutional ethical approval, the Womens' and Mens' gold medal games from the EuroHockey 2015 field hockey tournament were recorded using a static 4K camcorder (Sony FDR-AX1). The camcorder was positioned 14 m from the side of the pitch, approximately in line with the halfway line and at an elevation of

7 m. A 0.3x fisheye lens was used to maximise the pitch in the field of view. An HD camcorder was used to record the games from an elevation behind one of the goals. This second camcorder was allowed to pan and zoom to follow the ball during the game.

The performances of interest in this study were plays: a duration of gameplay starting when an attacker with possession crosses the 23 m line and ending when one of three outcomes occurred. These outcomes represented Best, Acceptable and Worst case scenarios for an offensive field hockey play (Table 1). At the moment of an outcome, the locations of all players on the pitch were digitised, calibrated and undistorted using in-house software. The second camcorder view was used to help locate players that were obscured in the primary view. Data for 233 plays were collected.

Table 1. Explanations of the Best-, Acceptable- and Worst-case outcomes of plays.

Scenario	Explanation
<i>Best</i>	When the ball is dribbled or passed into the circle. The circle, also known as "the D", is the semi-circular area in front of the goal from within which a goal can be scored.
<i>Acceptable</i>	When the play ends without a circle entry but possession is maintained. We use a defined list of situations: <ol style="list-style-type: none"> 1. A penalty corner is awarded. 2. A free hit is awarded. 3. The balls cross back over the 23 m line by passing or dribbling. 4. When the ball is hit over the backline by a defender following an uncontrolled interception or attempted steal.
<i>Worst</i>	When a defender regains and maintains control of the ball by intercepting or stealing, or any event where the rules of the game dictate that possession must be forfeited to the opposition, e.g. foul committed or the ball goes out of bounds.

Data Processing

Using the spatial and temporal data extracted from the footage, 1,837 spatio-temporal metrics were computed to describe each play; The complete list of each metric is available at doi.org/10.17032/shu-160004. We will not describe the metrics here because the purpose of this study is instead to understand the distribution of marginal effects amongst the metric set. All metrics were discretised because the methods used to estimate marginal effects require discrete data.

Analysis

We used three methods to estimate metrics' marginal effects: Cramér's *V*, Mutual Information and the I-score. A metric's marginal effect is that metric's isolated effect on the outcome when all other variables are ignored. These statistics are effect sizes and were our proxy for metrics' true effects. Each method estimates a marginal effect in a different way. Truly important metrics should score highly for all methods. Cramér's *V* is a measure of association between nominal variables, which in our case are our discrete metrics. Values less than 0.2 indicate small effects, values between 0.2 and 0.4 indicate moderate effects, and values above 0.4 indicate large effects. Mutual Information is a measure of mutual dependency between two variables (Peng et al., 2005). It measures the amount of information about one variable that can be determined from the other. The I-score measures the difference between the actual and expected count of a particular outcome

(Chernoff et al., 2009). It has a minimum of 0 meaning that the independent variable has no influence on the dependent variable. In this study, the independent variables are each metric and the dependent variable is the outcome type. The I-score is similar to Cramér's V but is not dependent on the degrees of freedom, which makes it useful when extended to multiple variables. This characteristic makes the I-score particularly useful for studying interaction effects.

Each of the three outcomes were compared in pairs: Best vs. Worst, Best vs. Acceptable, Acceptable vs. Worst. For each comparison, the marginal effect of each metric was first determined by Cramér's V because it has prescribed thresholds for large, moderate and small effects. The Mutual Information and I-score for each large effect metric was examined and its percentile calculated. If the large effect metrics (as per Cramér's V) are truly important then they should all be within high percentiles of Mutual Information and I-score, also.

RESULTS

In all comparisons, $\leq 1\%$ of metrics show large effects and $> 87\%$ of metrics show small effects, as determined by Cramér's V (Figure 2). This small number of large effect metrics were all within the 98th percentile of Cramér's V values, the 98th percentile of Mutual Information values and within the 96th percentile of the I-score values.

DISCUSSION

All three methods for estimating marginal effects of metrics suggested that there are relatively many metrics with small marginal effects, few with moderate effects and relatively very few with large marginal effects. Importantly, those metrics that were deemed to have large effects based on their Cramér's V also scored highest for Mutual Information and I-score. The agreement between the measures of marginal effect suggests that the distribution of marginal effects is representative.

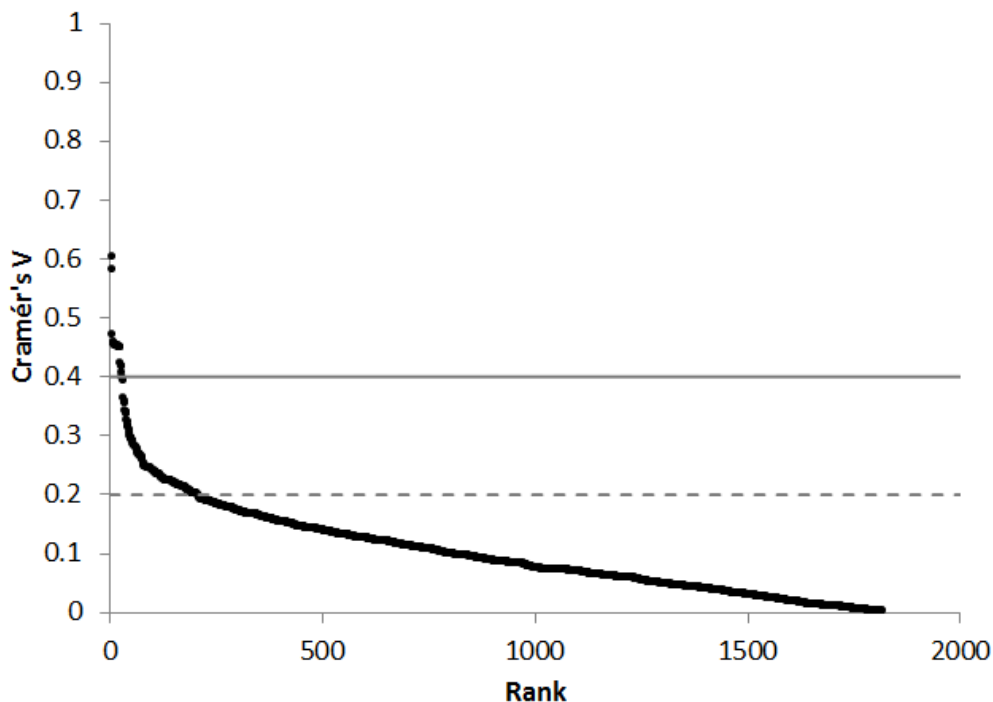


Figure 1. Cramér's V scores for all metrics, ordered by value, for the Best vs Worst comparison. The solid horizontal line (—) marks our threshold for a large effect and the dashed line (- -) marks our threshold for a moderate effect.

Haws et al. (2015) worked on a problem similar to ours in the area of genomic selection, i.e. very many independent variables and relatively few observations. The study of Haws et al. (2015) highlights the influence that the choice of analysis method has on the conclusions that can be made, which is why we took care to triangulate our estimates of marginal effect. It must be noted that our conclusions are limited to the dataset, metrics and statistics used.

The purpose of this study was to estimate the distribution of marginal effects because of its influence on the appropriateness of variable-selection methods (Tibshirani, 1996). Using appropriate methods to select important variables is not just a pedantic concern but a practical one. Resources are limited in the elite sport for both athletes and support staff. Committing such limited resources to inappropriate recommendations can be detrimental in the short duration of an Olympic cycle where there is little time to make the most of interventions or correct mistakes.

Our results suggest that univariate variable-selection methods would be most appropriate, based on Tibshirani's (1996) suggestions. It is important to note that the method which Tibshirani's used to represent univariate variable-selection methods was only appropriate for variable counts up to 30. In the past two decades, advances have been made in the field of genomic selection, which attempt to determine influential gene expressions from a set of tens of thousands of candidate genes. Methods such as the I-score (Chernoff et al., 2009) and minimum-redundancy-maximal-relevance (Peng et al., 2005) are variable-selection methods that have been specifically design for problems with a large number of variables and few observations. Many of the best genomic selection methods take interaction effects into account making them particularly useful for team sports where the adversarial and cooperative performances of players are of interest.

CONCLUSIONS

The dataset investigated had a small number of metrics strongly related to outcome of plays. According to the recommendations of Tibshirani (1996), univariate variable-selection methods will be the most appropriate for selecting important metrics.

DATA AVAILABILITY STATEMENT

Due to confidentiality agreements with research funders, the underlying research data can be accessed from SHU Repository for Data at doi.org/10.17032/shu-160004 by request only. Data can only be accessed with the permission of all funders.

ACKNOWLEDGMENTS

We would like to acknowledge the Engineering and Physical Sciences Research Council for partly funding this research (grant number EP/K504543/1).

REFERENCES

1. CHERNOFF, H., LO, S. H., & ZHENG, T. (2009). Discovering influential variables: A method of partitions. *Ann. Appl. Stat.*, 3(4), 1335–1369. <http://doi.org/10.1214/09-AOAS265>
2. GUDMUNDSSON, J., & HORTON, M. (2016). Spatio-Temporal Analysis of Team Sports - A Survey. *arXiv Prepr. arXiv1602.06994*. Retrieved from <http://arxiv.org/abs/1602.06994>
3. HAWS, D. C., RISH, I., TEYSSÉDRE, S., HE, D., LOZANO, A. C., KAMBADUR, P., ... PARIDA, L. (2015). Variable-selection emerges on top in empirical comparison of whole-genome complex-trait prediction methods. *PLoS One*, 10(10), 1–22. <http://doi.org/10.1371/journal.pone.0138903>
4. KARCHER, C., & BUCHHEIT, M. (2014). On-court demands of elite handball, with special reference to playing positions. *Sport. Med.*, 44, 797–814. <http://doi.org/10.1007/s40279-014-0164-z>
5. LANZONI, I. M., DI MICHELE, R., & MERNI, F. (2011). Performance indicators in table tennis : a review of the literature. *Int. J. Table Tennis Sci.*, 7, 71–75.
6. NG, V. W., & BREIMAN, L. (2005). Bivariate variable selection for classification problem.pdf, 1–22. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.62.3321>
7. PENG, H. C., LONG, F. H., & DING, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(8), 1226–1238. <http://doi.org/10.1109/TPAMI.2005.159>
8. PODGÓRSKI, T., & PAWLAK, M. (2011). A Half Century of Scientific Research in Field Hockey. *Hum. Mov.*, 12(December 2010), 108–123. <http://doi.org/10.2478/v10038-011-0008-8>
9. READ, B., & EDWARDS, P. (1997). *Teaching children to play: A resource for primary teachers 5-11*. Leeds: White Line Publishing.
10. TIBSHIRANI, R. (1996). Regression Selection and Shrinkage via the Lasso. *J. R. Stat. Soc. B*. <http://doi.org/10.2307/2346178>