BARÇA SPORTS ANALYTICS SUMMIT

BARÇA INNOVATION HUB
Reinventing Sport

Estimating Locomotor Demands During Team Play from Broadcast-Derived Tracking Data

Jacob Mortensen, Luke Bornn / Simon Fraser University / jmortens@sfu.ca

Introduction

In order to improve performance, reduce fatigue, and prevent injury, sports scientists monitor measures of external load, such as total distance and total acceleration. The gold standard for obtaining such measurements is to have athletes wear a device, such as an accelerometer or GPS, but this is not always possible. In these cases, multicamera optical tracking data is often a good alternative but is only available at the most elite levels and, because it has only become widespread in about the past decade, allows for only limited historical comparison. Broadcast-derived tracking data allows x-y coordinate data to be generated from a regular broadcast feed (i.e., what you would see watching a sporting event on television), and has the potential to overcome both of these limitations. However, the obvious shortcoming of broadcast-derived tracking data is that tracks are available for a player only as long as they are on-screen, limiting its usefulness for external load metric estimation. In this work, we build models using available on-screen information to make predictions about the censored external load metrics. We obtain accurate predictions and show that broadcast-derived tracking data is viable for load metric estimation.

Data

- 18 home games played by Chelsea FC in the 2014-15 English Premier League
- Train on first 13 games, test on last 5
- We establish ground truth by applying camera window to full multicamera tracking data

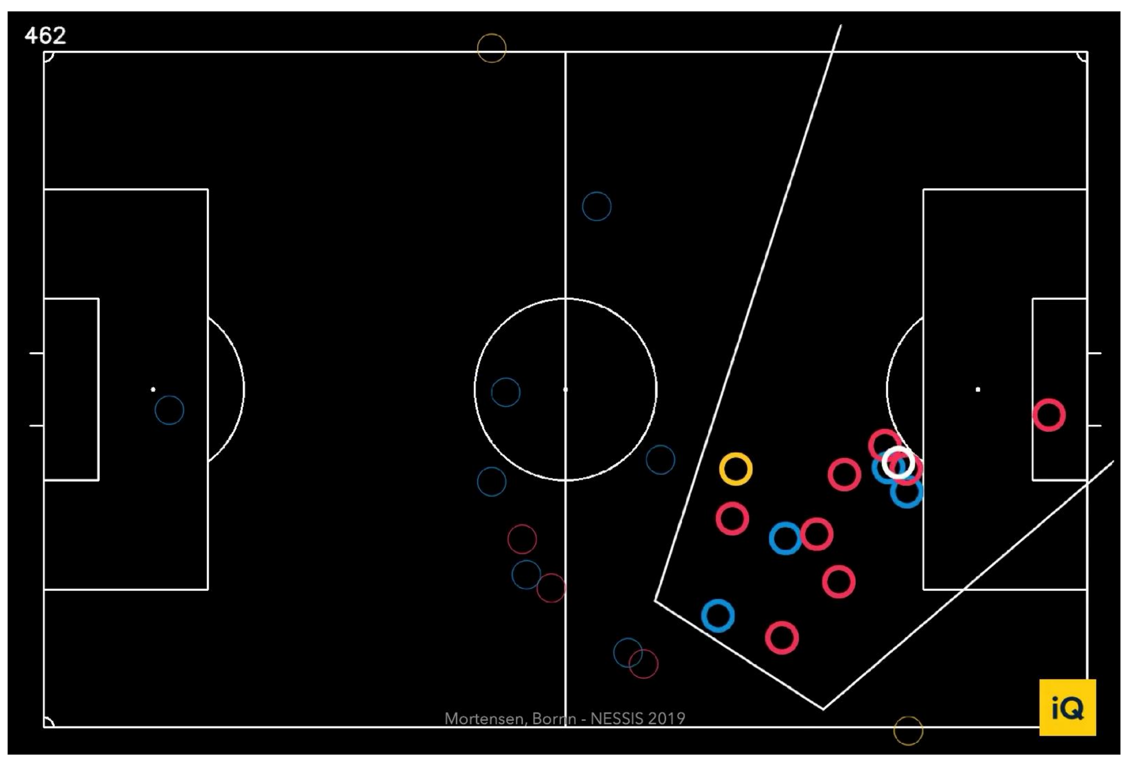

Figure 1. One frame of broadcast-derived tracking data. The camera window is represented by the white trapezoid. Players on-screen are brightly colored while off-screen players are muted. Image courtesy of Sportlogiq.

Category	Metric	Definition
Distance	Total distance	Sum of the distance travelled by an athlete
	High speed distance	Sum of the distance travelled by an athlete at a velocity between 3.5 and 5.7 m/s
	Very high speed distance	Sum of the distance travelled by an athlete at a velocity greater than 5.7 m/s
Velocity	Time spent in velocity band [x,y)	Number of seconds spent with velocity (m/s), v, in the interval x ≤ v < y. Intervals considered are [0,3.5), [3.5, 5.7), and [5.7, ∞), based on the work of Dwyer and Gabbett (2012)
	Peak x-second velocity	Maximum velocity of average velocities calculated over x = 1, 3, 5, and 10 second rolling windows
Acceleration	Total acceleration	Sum of the absolute acceleration at 0.1 second intervals
	Acceleration density	Mean absolute acceleration
	Time spent in acceleration band [x,y)	Number of seconds spent with acceleration (m/s^2), a, in the interval x ≤ a < y. Intervals considered are [0.65,1.46), [1.46,2.77), and [2.77,∞), based on the work of Johnston et al. (2014)

Table 1. External load metrics estimated in this work.

Methods

- Fit all models using extreme gradient boosting
- Three types of models:
    - Linear
    - Linear with all two-way interactions
    - Nonlinear (random forest)
- Two levels
    - Subtrack
    - Game
- We compare these 6 models to a base linear model that consists of just an intercept, amount of censored time, and the observed metric value

Predictor	Included at subtrack level	Included at game level
Player position	X	X
Offscreen time	X	
Total censored time		X
Offscreen distance	X	
Observed total distance		X
Average velocity in previous and next 2 seconds	X	
Average absolute acceleration in previous and next 2 seconds	X	
Observed values for all the player load metrics	X	X

Table 2. Covariates considered at the subtrack and game levels.

Results

- Models are compared using root mean square predictive error (RMSPE) and the coefficient of variation (CV)
- Subtrack level models outperform game level models for all metrics
- The linear model with interactions performs best for most of the external load metrics
- As one example, the RMSPE for time in velocity band [0, 3.5) is just 29.1 seconds, which is a very small value in the context of a 90 minute game

	Base model		Linear model		Linear model w/ int		Random forest	
Metric	RMSPE	CV	RMSPE	CV	RMSPE	CV	RMSPE	CV
Total distance (m)	288.2	0.08	202.0	0.06	188.3	0.05	**183.0**	**0.05**
High speed distance (m)	164.5	0.22	113.8	0.15	**106.0**	**0.14**	113.4	0.15
Very high speed distance (m)	60.4	0.44	53.4	0.39	53.3	0.39	**42.8**	**0.31**
Time in velocity band [0, 3.5) (s)	49.9	0.03	30.4	0.02	**29.1**	**0.02**	29.8	0.02
Time in velocity band [3.5, 5.7) (s)	37.8	0.22	26.4	0.15	**24.5**	**0.14**	26.4	0.15
Time in velocity band [5.7, ∞) (s)	8.8	0.43	7.9	0.38	7.9	0.38	**6.4**	**0.31**
Total acceleration (m/s^2)	2473	0.11	1448	0.07	**1365**	**0.06**	1366	0.06
Acceleration density (m/s^2)	0.140	0.12	**0.113**	**0.10**	0.129	0.11	0.119	0.11
Time in acceleration band [0.65, 1.46) (s)	34.9	0.06	25.8	0.05	**25.7**	**0.05**	29.5	0.05
Time in acceleration band [1.46, 2.77) (s)	45.3	0.13	30.6	0.09	**27.8**	**0.08**	31.5	0.09
Time in acceleration band [2.77, ∞) (s)	36.5	0.22	21.5	0.13	**20.9**	**0.13**	22.3	0.13

Table 3. Root mean square predictive error and coefficient of variation for the subtrack level models compared to the base model. The subtrack level models outperform the game level models in all cases, so results at the game level are omitted.

Conclusion

Using out-of-the-box statistical methods, we can predict censored external load metrics with a high degree of accuracy