

## Duckworth–Lewis run out?

J. Piotto<sup>1</sup>J. Turner<sup>2</sup>A. V. Metcalfe<sup>3</sup>

(Received 16 January 2016; revised 5 June 2018)

### Abstract

In one-day cricket matches the two competing teams have the same number of overs, typically 50, and the highest scoring team wins, with ties possible but unusual. If the match is interrupted, then the Duckworth–Lewis method is accepted by the International Cricket Council as the algorithm for deciding the result. In 2007, John Turner suggested an alternative method, named the Player Specific Method, that models the performance of batters against different types of bowlers and the remaining time in an innings, and then uses computer simulation to generate an ensemble of possible scores. Then, either the mean value of these scores is used to determine the winning team, or a probability is assigned to one of the teams winning. This method is illustrated for the one-day match between Australia and Sri Lanka at the MCG on 22nd February 2008, and its feasibility as an alternative to the Duckworth–Lewis method is discussed.

---

[DOI:10.21914/anziamj.v57i0.10477](https://doi.org/10.21914/anziamj.v57i0.10477) gives this article, © Austral. Mathematical Soc. 2018. Published August 15, 2018, as part of the Proceedings of the 12th Biennial Engineering Mathematics and Applications Conference. ISSN 1445-8810. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to the DOI for this article.

# Contents

<b>1</b>	<b>Introduction</b>	<b>C414</b>
<b>2</b>	<b>Calculating a simulated score</b>	<b>C416</b>
2.1	What is the PSM? . . . . .	C416
2.1.1	Scoring with PSM . . . . .	C416
2.1.2	Ordinal logistic regression . . . . .	C418
2.2	PSM in Practice . . . . .	C422
2.2.1	PSM Example . . . . .	C422
<b>3</b>	<b>Comparison between PSM and Duckworth–Lewis</b>	<b>C425</b>
<b>4</b>	<b>Conclusion</b>	<b>C426</b>
	<b>References</b>	<b>C427</b>
<b>A</b>	<b>Batting record data</b>	<b>C428</b>

## 1 Introduction

Test cricket, which dates back to 1877, is an unlimited overs bat and ball game played between two teams of 11 players over a period of up to five days, typically with four innings. One team, determined by the toss of a coin, bats during the first innings. The team bats until either ten players are dismissed or the captain decides to declare, so as to leave time to dismiss the other team. The other team now bats during the second innings, and the teams continue to swap roles for the third and fourth innings. Results take one of the following forms.

- A win: the winning team scores more runs than the opposition and dismisses the whole opposing team twice.

- A draw: an inconclusive result that occurs when the team with more runs fails to dismiss the opposing team twice. This is a common result and a weaker team may play defensively in the hope of achieving a draw.
- A tie: when a match is finished and all four innings are completed and the combined total runs scored for both teams are the same, this has only occurred twice.

One-day cricket is an alternative form of international cricket which was first played in 1971. It involves only two innings, each of which ends after 50 overs or ten dismissals. The match is completed within a day and the result is either a win (team which scored more runs) or a tie. The one-day cricket match has proved popular with cricket enthusiasts, with the style of play and strategies adapted due to the limited time to achieve a result.

Interruptions to one-day cricket, such as rain or poor lighting, can affect the number of overs a team is entitled to. To combat this, Frank Duckworth and Tony Lewis devised a system to allow for a fair match, this is called the Duckworth–Lewis (D/L) method. The essence of this method is that it models the proportion of total run scoring resources lost as a function of overs received and wickets lost [1]. This method was formally adopted by the International Cricket Council (ICC) in 2001 as a standard means of calculating target scores in interrupted one-day matches. Consequently, the D/L method is used to provide both teams with an opportunity to win the game if all 50 overs in an innings are not played. The D/L method was first used during a one-day international match between Zimbabwe and England in Harare on the 1st January 1997. Zimbabwe received 50 overs and made 200 runs before rain interrupted play; England were restricted to 42 overs and their target to win, using the D/L method, was 186 runs. In 42 overs England only scored 179, therefore losing the match by 6 runs [1]. The latest revision of this method is known as the Duckworth–Lewis–Stern Method [2].

The D/L method is based on average scoring rates of teams playing one-day matches, and in 2007, after scoping Julia Piotto's honours thesis [3], John

Turner proposed an alternative that takes account of the specific players in a side and their current form. The idea behind this alternative method, which is named the Player Specific Method (PSM), is to model each batter's performance against different types of bowlers, and then use a computer simulation to model the results of the match. The PSM, its application and an example are described below.

## 2 Calculating a simulated score

### 2.1 What is the PSM?

The Player Specific Method (PSM) is a method which determines the winner of an interrupted one-day cricket match by a statistical analysis of data from the remaining players. From open source cricket statistics, it estimates the probabilities of the remaining players getting out or scoring between zero and six runs, given the bowler type and the number of overs remaining. These probabilities are then used in a computer simulation.

#### 2.1.1 Scoring with PSM

To develop PSM, matches in which the first innings was completed and the second was terminated were considered. Following the D/L method the objective is to calculate a score and a winning/losing margin for the team that had an incomplete innings. The next step is to determine how well each of the remaining batters (from the team whose innings was terminated) would perform against fast, medium and slow type bowlers. This data, for one-day matches, was sourced from the ESPN Cricinfo website [4]. The information obtained included the number of balls the remaining batters had faced in their international career, the over they received those balls, the type of bowler they faced, and the score they made from that ball. This data was then input to

Table 1: Silva and Dilshan against medium pace bowlers.

Ball	Silva	Dilshan
out	0.014	0.047
0	0.521	0.511
1	0.366	0.302
2	0.070	0.070
3		
4	0.028	0.070
5		
6		

the open source software R [5] to calculate the probability of the player being ‘out’ or scoring (0, 1, . . . , 6) runs against each bowler type. This principle is straightforward, although the precise coding required some care [3].

The result of a single ball is simulated by an independent pseudo-random draw from a uniform [0, 1) distribution. The uniform deviate defines one of the eight outcomes (out, 0, 1, . . . , 6) from the estimated cumulative probabilities for the specific player against bowler type. To demonstrate this, the second column (Silva) in Table 1 is interpreted, for medium paced bowlers, as:

- generated number  $\leq 0.014$ , then Silva is out;
- generated number  $> 0.014 \cap \leq 0.535$ , Silva doesn’t score any runs;
- generated number  $> 0.535 \cap \leq 0.901$ , Silva scores 1 run;
- generated number  $> 0.901 \cap \leq 0.971$ , Silva scores 2 runs;
- generated number  $> 0.971$ , Silva scores 4 runs.

The same logic can be applied to Dilshan’s results in the third column of Table 1.

Subsequently, the results of the different batters are tallied to provide a

simulated total score for the match using

$$\text{expected runs} = \frac{f'}{t} \times \bar{f} + \frac{m'}{t} \times \bar{m} + \frac{s'}{t} \times \bar{s}, \quad (1)$$

where

- $f'$  number of overs fast bowlers completed;
- $\bar{f}$  mean runs against fast bowlers in remaining overs;
- $m'$  number of overs medium bowlers completed;
- $\bar{m}$  mean runs against medium bowlers in remaining overs;
- $s'$  number of overs slow bowlers completed;
- $\bar{s}$  mean runs against slow bowlers in remaining overs;
- $t$  total overs before interruption.

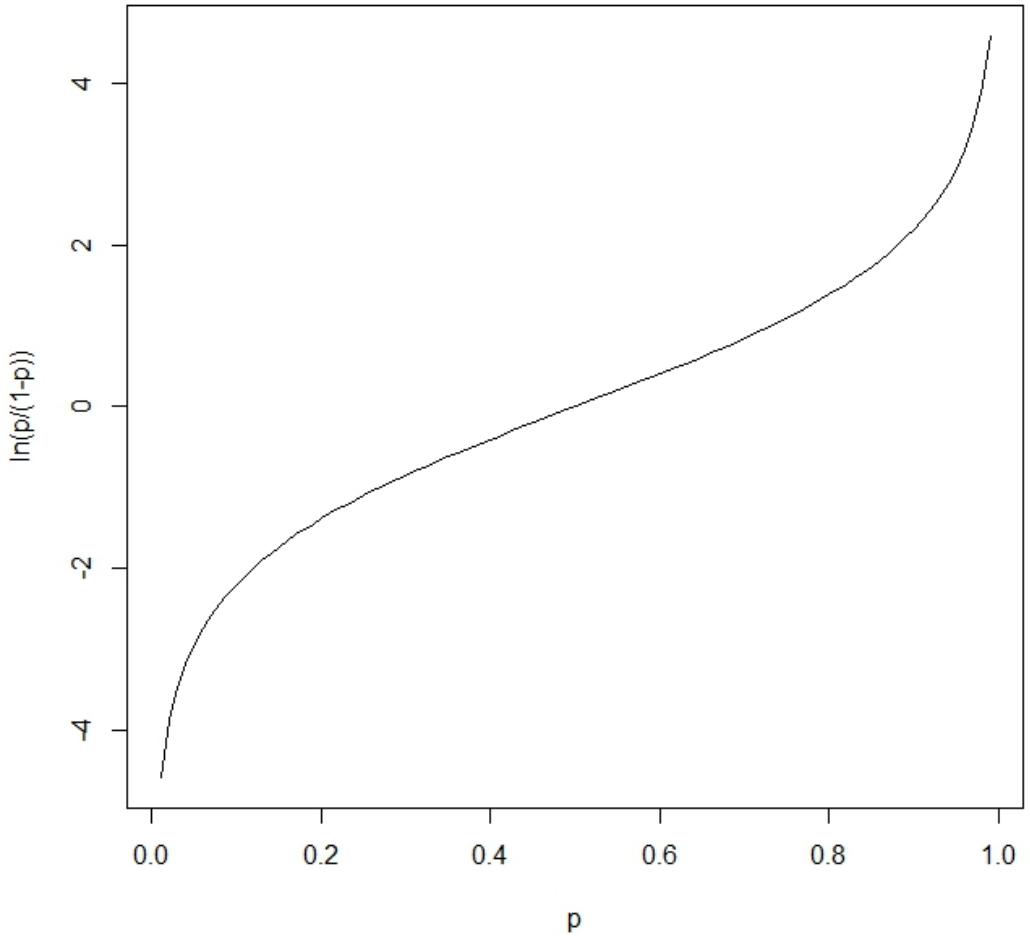
This simulation is repeated 1000 times to give an empirical probability distribution of scores. The mean of the distribution can be used to give a single result (win, loss or tie) or the entire distribution can be used to give an estimated probability for each result.

### 2.1.2 Ordinal logistic regression

The simple simulation doesn't take account of the number of overs that have been played or the fact that batsmen will typically tend to play aggressively in the final few overs if few wickets have been lost [1]. An ordinal logistic regression provides a neat solution to this problem and can accommodate many other features of a cricket match, such as state of the wicket, provided there is sufficient data. The logistic function transforms a probability measured on the interval  $[0, 1]$  to the infinite interval  $(-\infty, \infty)$  by taking the natural logarithm of the odds,

$$\text{logit}(p) = \log \left( \frac{p}{1-p} \right).$$

Figure 1: Logit function.



The logit function is plotted in Figure 1. The inverse function defined by:

$$\text{logit}(p) = \theta \Leftrightarrow p = \frac{e^\theta}{1 + e^\theta}.$$

A logistic regression model for the probability of being declared out that

Table 2: Indicator variables for medium and slow bowlers relative to fast bowlers.

Bowler	$x_2$	$x_3$
Fast Pace	0	0
Medium Pace	1	0
Slow Pace	0	1

allows for both bowler type and number of overs played is

$$\text{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

where  $x_1$  is the number of the overs played, and  $x_2$  and  $x_3$  are indicator variables that model medium and slow bowlers relative to fast bowlers (Table 2).

The data, provided in Appendix A, for Sri Lankan batter Farveez Maharooof is now used to demonstrate this logistic regression principle. The data is limited as only ten games are analysed. In the data there are only two instances of Maharooof being declared out in the short record of 39 balls; and it is not feasible to estimate the  $\beta_j$  coefficients, other than the intercept  $\beta_0$ , to any acceptable precision. The estimate of Maharooof being declared out on any ball is  $\frac{2}{39} = 0.0513$ .

It is feasible to fit an ordinal logistic regression for the number of runs scored given that a batter is not out. For Maharooof the ordered categories ( $j = 1, 2, 3$ ) for this record are:  $j = 1$  representing no runs scored off the ball;  $j = 2$  representing a single run scored off the ball; and  $j = 3$  representing four runs scored off the ball. The model is

$$\text{logit}(p_j) = \beta_{0,j} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

where  $j = 1, 2$ ,  $p_1$  is the probability of 0 runs, and  $p_2$  is the probability of 0 or 1 run. It follows that the probabilities of 0, 1 or 4 runs are  $p_1$ ,  $p_2 - p_1$ , and  $1 - p_2$ , respectively. The model is fitted conveniently in R using the VGAM package [6]. An excerpt from the results is given in Table 3.



Table 3: Fitted ordinal logistic regression.

Factor	Coeff.	Estimate	Std. error	z-value	p-value
1	$\beta_{0,1}$	5.988	3.760	1.59	0.111
1	$\beta_{0,2}$	8.193	3.870	2.12	0.034
$x_1$ (overs)	$\beta_1$	-0.150	0.083	-1.81	0.071
$x_2$ (med. rel. fast)	$\beta_2$	0.179	1.276	0.14	0.888
$x_3$ (slow rel. fast)	$\beta_3$	0.651	0.751	0.87	0.386

In Table 3, the z-value is the ratio of the estimate to its standard error and the p-value is  $\Pr(|Z| > z\text{-value})$ . The estimated coefficient of  $x_1$  for the number of overs played is  $\beta_1 = -0.150$  and this is statistically significantly different from 0 at a 10% level (p-value= 0.07). The interpretation is that as the number of overs played increases the logit of the probability, and hence the probability, of being in categories  $j = 1$  or 2 decreases. Consequently the probability of being in category 3 increases. Category  $j = 3$  is scoring four runs off a ball, so the model is emulating batters hitting out as the number of remaining overs decreases. The estimated coefficient of  $x_2$ , which takes the value of 1 for a medium paced bowler, is  $\beta_2 = 0.179$ . Although this estimate is not statistically significantly different from 0 at the 10% level (p-value= 0.89), its positive sign is plausible. The interpretation is that the probability of being in category 3 is less for a medium paced bowler than for a fast bowler, in agreement with the generally held view that it is easier to score four runs off fast bowlers. A similar argument applies for slow bowlers.

The probability of 1 run is obtained by subtracting the probability of 0 runs from the probability of 0 or 1 run. For example, suppose Maharooof is facing a slow bowler in the 40th over, then

$$\begin{aligned} \text{logit}(p_1) &= 5.988 - 0.150 \times 40 + 0.651 = 0.645 \\ &\Rightarrow p_1 = 0.656, \\ \text{logit}(p_2) &= 8.193 - 0.150 \times 40 + 0.651 = 2.850 \\ &\Rightarrow p_2 = 0.945, \end{aligned}$$

and it follows that the probability he scores 0, 1 or 4 runs, provided he is not out, are 0.656, 0.289 and 0.055, respectively.

The results of the logistic regression analysis can be used in place of the relative frequencies obtained directly from summary tables in the simulation code.

## 2.2 PSM in Practice

To demonstrate how PSM works, the terminated match between Australia and Sri Lanka at the MCG on 22nd February 2008 is considered. Australia batted for 50 overs and scored 184 runs while Sri Lanka batted for 29 overs and 3 of the 6 balls in the 30th over (29.3 overs) before rain stopped play. At this point Sri Lanka had scored 77 runs for 4 wickets. The game was terminated some time later due to the continuing rain, resulting in the implementation of the D/L Method. This concluded that Sri Lanka required 102 runs from the 29 overs and 3 balls that they faced. Since Sri Lanka only scored 77 runs Australia was declared the winner by 24 runs <sup>1</sup>.

### 2.2.1 PSM Example

The PSM takes a different approach to the D/L Method and simulates the remainder of Sri Lanka's innings, rather than setting a revised target. The simulated runs scored are added to the 77 runs that Sri Lanka actually scored to give a simulated total for the match. The simulation involves random numbers, so the total will vary from one simulation to the next. This variation is effectively removed by averaging the results of many simulations.

The PSM takes account of the remaining batsmen: Chamara Silva and Tillakaratne Dilshan, who were not out; and Farveez Maharoo, Chaminda

---

<sup>1</sup>This can be calculated by either looking at the D/L graph, or table [1], where the resources remaining for 20.3 overs and 4 wickets is 45.2%.

Vaas, Lasith Malinga, Ishara Amerasinghe and Muttiah Muralitharan, who were yet to bat. As mentioned previously, the simplest model is to simulate the performance of the above remaining batters against different bowler type and then weight the runs scored in proportion to the mix of bowlers that Australia would likely use. For example, assume Australia uses equal numbers of medium and slow paced bowlers to defend their lead. Statistics for all the players in one-day international matches are available from the ESPN Cricinfo website [4], and Table 1 gives an example of the output for Silva and Dilshan facing medium paced bowlers.

A simulation then generates a uniform pseudo-random number in the interval  $[0, 1)$ , as described in Section 2.1.1. This continues until either 20.3 overs are completed or 6 wickets are taken, and then the total runs scored against medium bowlers  $m_k$ , for simulation number  $k = 1, \dots, K$ , is stored. The simulation is repeated in a similar fashion for fast bowlers and for slow bowlers, and the scores,  $f_k$  and  $s_k$ , respectively, are stored. The results for  $K = 1000$  simulations are shown in Figure 2 for the remaining Sri Lankan batters against the three types of bowlers. Before the game was stopped, Sri Lanka required 107 runs to win (the difference between Australia's total and their total before the innings was interrupted). Figure 2 shows that Sri Lanka are unlikely to score the required runs against slow and medium paced bowlers; however, when faced with fast paced bowlers their chance of winning is increased.

A weighted score  $W_k$  can be calculated for each simulation  $k$  where the weights are the proportions of overs faced against fast, medium, and slow bowlers. In this example the proportions are 0, 0.5 and 0.5, respectively. So, the weighted score is

$$W_k = 0 \times f_k + 0.5 \times m_k + 0.5 \times s_k .$$

In the lower right frame of Figure 2 the weighted scores  $W_k$  are shown, with the vertical line indicating the 107 runs. Using the weighted mean runs calculated in Table 4, the losing margin for Sri Lanka is

$$81.82 - 107 = -25.18 .$$

Figure 2: Histograms of runs scored by Sri Lanka from remaining the 20 overs and 3 balls in 1000 simulations, against type of bowler. Scores to the right of the vertical lines are a win for Sri Lanka.

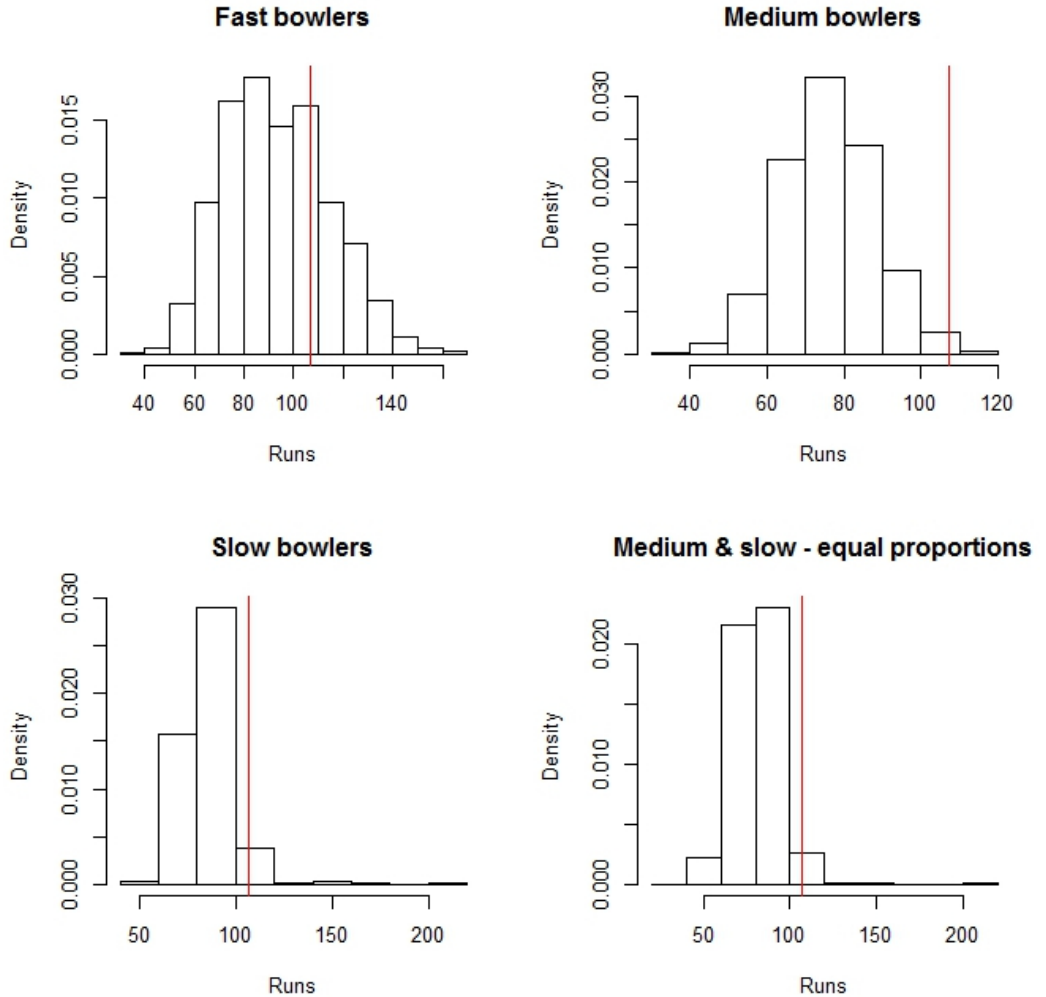


Table 4: Mean of ensemble of simulated total runs scored.

Bowler	Mean Total Runs
$\bar{f}$	94.33
$\bar{m}$	76.78
$\bar{s}$	86.66
Weighted: 50% medium 50% slow	81.82

Although the PSM will not be used as a replacement for the D/L method, the result produced is quite similar: a 25 run loss versus a 24 run loss, respectively.

### 3 Comparison between PSM and Duckworth–Lewis

The PSM can be used for any scenario of interruptions. For example, suppose the innings of the first team (Team A) is interrupted by rain after 30 overs. The rain persists and the match continues with the second team (Team B) being sent in to bat for 30 overs. A PSM solution is to simulate the remaining 20 overs of Team A and to tell Team B that they have 30 overs and the 20 remaining overs will be simulated. In this way both teams bat as if they have 50 overs. The solution does require teams to accept computer simulation as part of their score. Since the simulation is based on past performance, it can be argued that PSM will disadvantage a team that is performing particularly well before the interruption. In principle, this limitation could be removed by applying a match specific factor to the simulated score. The basis for calculating such a factor would have to be agreed in advance and teams would have to accept that the factor could be less than 1 rather than greater than 1.

The data requirement for serious implementation of PSM is considerable, but Ty Stanford [personal communication] has demonstrated that the text

processing capability of R, in particular its handling of regular expressions, can be linked with ESPN Cricinfo [4] commentaries to automatically produce tables of statistics for the simulation. An expert system for implementing PSM is quite feasible.

PSM has four advantages over the D/L method. The first is that it does not rely on an average use of resources and can make fair allowances for a team that has eleven strong batsmen or a strategy of playing the stronger batsmen lower in the order. A second advantage is that PSM can provide a probability of Team A winning, rather than a result defined as win, lose or draw. A third advantage is that PSM can be used by selectors and team captains to investigate the likely performance of different teams and different playing strategies. The final advantage is that PSM is not restricted to one-day cricket and can be applied to 20/20 cricket matches.

## 4 Conclusion

We have described, and compared with the D/L method, the simulation-based PSM for deciding the result of an interrupted one-day cricket match. In the match considered, Sri Lanka had a reduced number of overs because of rain. The D/L method is based on the proportion of resources, relative to 50 overs and 10 wickets, that Sri Lanka had used when the match was terminated. In this case Sri Lanka had used 29 overs and 3 balls and lost 4 wickets. In contrast, PSM simulates the score that Sri Lanka would have achieved had it continued to bat until either it lost 10 wickets or had faced all 50 overs. The results obtained by the two methods were close.

Turning to the question posed in the title, "Duckworth Lewis run out?", the answer is certainly, not yet. The D/L method is established as the official method for deciding the result of interrupted one-day matches and has generally provided acceptable results, despite a few contentious matches. A criticism that the D/L method is based on outdated statistics that do not

take account of the increasingly high match scores has been answered by the Duckworth–Lewis–Stern updating. Nevertheless, PSM has potential for investigating the consequences of different players and strategies in matches.

**Acknowledgements** We thank Frank Duckworth for providing a transcript and slides from his talk to Leicester Branch of the Institute of Mathematics and its Applications (IMA) in January 2008.

## References

- [1] F. Duckworth and T. Lewis. *Duckworth Lewis: the method and the men behind it*. SportsBooks Ltd, 2011. C415, C418, C422
- [2] S. E. Stern. The Duckworth–Lewis–Stern method: extending the Duckworth–Lewis methodology to deal with modern scoring rates. *J. Oper. Res. Soc.* 67(12):1469–1480, 2016. doi:10.1057/jors.2016.30 C415
- [3] J. Piotto *Duckworth Lewis, Run Out?* Honours thesis, University of Adelaide, 2008. C415, C417
- [4] Sambit Bai. *ESPN Cricinfo*. <http://www.espnricinfo.com/> C416, C423, C426
- [5] The R Project for Statistical Computing. Vienna, Austria, 2013. <http://www.R-project.org/> C417
- [6] T. W. Yee. The VGAM package for categorical data analysis. *J. Stat. Softw.* 32:1–34, 2010. doi:10.18637/jss.v032.i10 C420

## A Batting record data

Table 5: Farveez Maharooof's batting record ( $-1$  represents declared out) against fast (F), medium (M) and slow (S) bowlers.

Over	Ball	Bowler type	Runs
38	4	F	0
35	5	F	1
42	3	F	1
44	1	F	0
45	1	F	0
45	2	F	1
45	5	F	0
45	6	F	0
46	6	F	1
48	1	F	4
48	2	F	4
48	3	F	1
48	5	F	1
49	1	F	4
49	2	F	0
49	3	F	4
49	4	F	-1
32	6	M	0
34	3	M	0
34	4	M	4
34	5	M	0
34	6	M	0
36	3	M	0
36	4	M	0
36	5	M	1



35	3	S	1
35	5	S	0
35	6	S	0
37	5	S	0
39	1	S	-1
41	1	S	0
41	2	S	1
47	1	S	1
47	5	S	0
47	6	S	1
49	1	S	0
49	2	S	1
49	4	S	1
49	6	S	1

## Author addresses

1. **J. Piotto**, Defence Science and Technology Group, Edinburgh, South Australia 5111, Australia.  
<mailto:Julia.Piotto@dsto.defence.gov.au>
2. **J. Turner**, Norwood, South Australia  
<mailto:jognturn@gmail.com>
3. **A. V. Metcalfe**, School of Mathematical Sciences, University of Adelaide, South Australia 5005, Australia.  
<mailto:andrew.metcalfe@adelaide.edu.au>