Integer optimisation for the selection of a Twenty20 cricket team

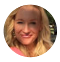
Article *in* Journal of the Operational Research Society · September 2011

DOI: 10.1057/jors.2010.122 · Source: OAI


Gary Sharp
Nelson Mandela University


Warren J. Brettenny
Nelson Mandela University


John Gonsalves
Nelson Mandela University


Michelle E Lourens
Texas Health Resources

Journal of the Operational Research Society (2011) 62, 1688–1694

www.palgrave-journals.com/jors/

Integer optimisation for the selection of a Twenty20 cricket team

GD Sharp*, WJ Brettenny, JW Gonsalves, M Lourens and RA Stretch

Nelson Mandela Metropolitan University, Port Elizabeth, South Africa

The attraction of Twenty20 cricket for spectators has sparked huge commercial interest in the marketing of the concept. The International Cricket Council has sanctioned world tournaments, the inaugural tournament in South Africa and the recent 2009 event, while the 2008 and 2009 Indian Premier League tournaments have generated profitable returns for both players and franchise owners. This paper describes methods for quantifying a cricket player's performance based on his ability to score runs and take wickets. These performance measures are then used to determine the optimal team using an integer programme. An illustration of the method is provided using data from the inaugural Twenty20 World Cup held in South Africa in 2007.

Journal of the Operational Research Society (2011) **62,** 1688–1694. doi:10.1057/jors.2010.122

Published online 25 August 2010

Keywords: sports; integer programming; team selection; statistics

1. Introduction

The statistics of professional sports players and teams provides numerous opportunities for research (De Silva and Swartz, 1997; Durbach and Thiart, 2007). Researchers continue to evaluate the performances of teams and players in many sporting disciplines. There are several journals that provide publication options for researchers. Albert and Koning (2008) summarise a few of the more well-known publications and emphasise the interest of the Sports Statistics Committee of the International Statistical Institute. They conclude that there are several publishing opportunities for statisticians who undertake serious research on sports data.

Cricket, as a research area, has continued to evolve as the game itself has evolved. Recent developments in the game have, to a large extent, been driven by commercial factors. The traditional game is a two innings per team contest that is played over several days. Spectator interest in the traditional game has waned as the end result is often inconclusive and considered by many to be boring. This brought about a shortened version of the game, a one innings per team contest with a limited number of overs, commonly known as a one-day game. This format was commercially successful, spectators enjoyed the shortened version of the game and ground attendances increased. A more recent format of the game, a Twenty20 contest, is an even shorter version in which each side is allowed to bat and bowl for no more than 20 overs. The commercial success of this format has seen the introduction of a World Tournament with the inaugural event held in South Africa in 2007, the second held in England in 2009 and the next to be held in the West Indies in 2010.

To those unfamiliar with the game of cricket, Preston and Thomas (2000) summarise the important concepts of limited overs cricket. Others (Lemmer 2002, 2004 and 2006; Beaudoin and Swartz, 2003; Gerber and Sharp, 2006; Barr *et al*, 2008; Bracewell and Ruggiero, 2009; Van Staden, 2009) introduce the game of cricket while specifically considering statistical methods for determining a player's performance. Both Van Staden (2009) and Bracewell and Ruggiero (2009) use interesting graphical measures to illustrate player performances. Simulation exercises have been attempted (Swartz *et al*, 2009), teaching related graphical displays have been published (Kimber, 1993) and applications to dynamic programming have been conducted (Clarke, 1988; Preston and Thomas, 2000). These are just a few of the mathematically inclined publications analysing cricket data, undoubtedly there will be many more to follow.

This study considers the most recent adaptation of the game, the Twenty20 format. In these contests, each 11-person team is expected to both bat and bowl. When batting, each team is allowed to bat for a maximum of 20 overs. Assuming there are no additional deliveries this equates to a maximum of 120 balls. When bowling, each team is allowed to bowl for a maximum of 20 overs. Each bowler is restricted to bowl a maximum of four overs. The batting and bowling requirements restrict the composition of a team to ensure that there are players capable of performing both these tasks.

Correspondence: GD Sharp, Department of Statistics, Nelson Mandela Metropolitan University, Port Elizabeth, Eastern Cape 6031, South Africa.

E-mail: gary.sharp@nmmu.ac.za

In this study, integer programming is used for team selection. Individuals are selected using binary response decision variables. The decision variables are defined to determine whether or not an individual is good enough for selection based on the optimisation of a linear function for pre-defined criteria. This method is described in Taha (2003) as an illustrative example for work-scheduling and capital budgeting but is relatively new to sports team selection (Gerber and Sharp, 2006). A possible reason for this is that quantifying and equating different skills for individuals in a team is difficult. Researchers have traditionally considered ranking methods for one cricket skill. Lemmer (2004), Ovens and Bukiet (2006) and Swartz *et al* (2006) propose ranking systems for batting, while Lemmer (2002, 2006) proposes ranking systems for bowling. In particular, Swartz *et al* (2006) propose a simulation model for selecting a first innings batting sequence. The simulation model allows for real-time batting order changes subject to the conditions of the match. These methods are useful for selecting the best-ranked batsmen, but the integer programming method proposed here is more general in that the optimal solution selects an entire team.

The method applies the index modelling proposal of Gerber and Sharp (2006) while using a variation of the weighted product measure of performance as given in Barr *et al* (2008). Gerber and Sharp (2006) provided a description of the need to convert performance abilities, batting and bowling, to index measures rather than common cricket statistical terms. The motivation is that to run an integer programming model one requires that the objective function coefficients are of equivalent value.

This study's contributions include the defining of an integer programming model for selection of an 11-person cricket team. The model adds to the research of Gerber and Sharp (2006) by extending the modelling framework to Twenty20 cricket. In addition, the coefficients in the model are determined by defining ability indices that are both scale and location adjusted. Ability measures for Twenty20 cricket are defined by incorporating the methods of Lemmer (2004), Barr *et al* (2008) and Lourens (2009). The methodology is discussed in Section 2. In particular, the cricket statistics are defined and coefficient indices justified. An illustrative example, using current data, is given in Section 3 while the integer programme for team selection is defined in Section 4. Section 5 discusses the results and implications of modelling, while Section 6 concludes with limitations and contributions of the study.

2. Methodology

2.1. Background

The emphasis in cricket is simple; a team should score as many runs as possible while restricting the opposing side to as few runs as possible. Scoring runs requires that the teams have players capable of batting well; two common measures of batting performance are batting average and batting strike rate. Restricting the opposition's score requires that the teams have players capable of bowling well; three common measures of bowling performance are bowling economy rate, bowling average and bowling strike rate. These statistics, defined below as Y_{ij}, are used to obtain performance measures as indices that can then be used as coefficients for an integer programming problem.

2.2. Requirements of the integer programming model

The theory of integer programming has been well established (Taha, 2003) and the application to team selection was shown in Gerber and Sharp (2006). A summary of the Gerber and Sharp (2006) approach follows, although for a more comprehensive description it is best to refer to the original paper. Integer optimisation for team selection requires the optimisation of a linear objective function subject to linear constraints and integer restrictions on the decision variables. The integer restrictions are constrained to be binary where a zero indicates non-selection and one indicates selection.

The programming model is defined as the maximisation (or minimisation) of an objective function $Z = \sum_{i=1}^{m}\sum_{j=1}^{n_i} c_{ij}x_{ij}$ and $x_{ij} \in (0,1)$ subject to team selection constraints. The decision variables, x_{ij}, are the binary identifiers for ability i and player j, where $(i = 1,2, \ldots, m)$ and $(j = 1,2, \ldots, n_i)$. The coefficients, c_{ij}, are the numerical quantities derived from the performance measure indices for ability i and player j.

The linear constraints for the programming model are used to ensure an adequate number of players are selected for each ability. The team selection constraints can include a minimum number of batsmen, a minimum number of bowlers, a minimum number of all-rounders, ensure a wicketkeeper is selected and even ensure the selection of a captain if deemed necessary.

Constraints are included only if they make cricketing sense, a simple example would be to include the restriction that exactly 11 players are needed for a team. This restriction is defined using two constraints. To ensure 11 players are selected, it is necessary to include the constraint $\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} = 11$. This constraint ensures that exactly 11 decision variables equate to unity. To ensure no player is selected more than once the constraint $\sum_{i=1}^{m} x_{ij} \leqslant 1$ for all j is included. Together these two constraints ensure exactly 11 players are selected for a team.

2.3. Batting statistics

This study considers two batting statistics often used as a measure of a player's performance, the player's batting average denoted as Y_{1j} and the player's batting strike rate denoted as Y_{2j}. These statistics are defined as follows:

$$\text{Batting average: } Y_{1j} = \frac{\text{Total number of runs scored by player } j}{\text{Total number of dismissals for player } j}$$

and

$$\text{Batting strike rate: } Y_{2j} = \frac{\text{Total number of runs scored by player } j}{\text{Total number of balls received by player } j} \times 100.$$

A batsman with high values for Y_{1j} and Y_{2j} is considered to be a good player as increasing values for these statistics improves the team's chances of winning. Barr *et al* (2008) use a product measure of these statistics as a criterion for batting ability defined as $U_{1j} = (Y_{1j}^{\alpha})(Y_{2j}^{1-\alpha})$ where $0 \leqslant \alpha \leqslant 1$. Weighting factor α determines the importance of Y_{1j} and Y_{2j} in terms of the contest format.

This measure was initially proposed by Croucher (2000) without the power restriction and referred to as a batting index. In this study the batting index is defined as $c_{1j} = \left(\frac{U_{1j}}{\sum_{j=1}^{n_1} U_{1j}}\right) \times n_1$, where U_{1j} is the batting ability for player j and n_1 is the number of batsmen for whom batting abilities are available. This index is used as a quantitative score of the player's batting ability relative to the other batsmen. The higher the value, the better the batsman.

2.4. Bowling statistics

This study considers three bowling statistics often used as a measure of a player's performance, the bowler's economy rate (Y_{3j}), the bowler's bowling average (Y_{4j}) and the bowler's strike rate (Y_{5j}). These statistics are defined as follows:

$$\text{Bowling economy rate: } Y_{3j} = \frac{\text{Total number of runs conceded by bowler } j}{\text{Total number of overs bowled by bowler } j},$$

$$\text{Bowling average: } Y_{4j} = \frac{\text{Total number of runs conceded by bowler } j}{\text{Total number of wickets taken by bowler } j}$$

and

$$\text{Bowling strike rate: } Y_{5j} = \frac{\text{Total number of balls bowled by bowler } j}{\text{Total number of wickets taken by bowler } j}.$$

A bowler with low values for Y_{3j}, Y_{4j} and Y_{5j} is considered to be a good player as decreasing values for these statistics improves the team's chances of winning. Adapting the Barr *et al's* (2008) batting product measure for bowling allows for a single criterion for bowling ability defined as $U_{2j} = (Y_{3j}^{\alpha_1})(Y_{4j}^{\alpha_2})(Y_{5j}^{1-\alpha_1-\alpha_2})$, where $0 \leqslant \alpha_1 \leqslant 1$, $0 \leqslant \alpha_2 \leqslant 1$ and $0 \leqslant \alpha_1 + \alpha_2 \leqslant 1$. The weighting factors α_1 and α_2 determine the importance of Y_{3j}, Y_{4j} and Y_{5j} in terms of the contest format.

Given that low values indicate good performance, a programming model would seek to minimise some objective function value, as opposed to the batting index which would require maximising the value. This dilemma is addressed by transforming the performance measure using the method proposed by Gerber and Sharp (2006). Multiplying the bowling ability by -1 allows that the ability be modelled with a maximisation objective function. This multiplication yields negative coefficient values, and hence the definition by Gerber and Sharp (2006) gives positive coefficients which ensure that the higher the index, the better the player.

The bowling performance is defined as $V_{2j} = \left(k - \left(\frac{U_{2j}}{\sum_{j=1}^{n_2} U_{2j}}\right)\right)$, where k is a constant chosen as the smallest positive value such that $\left(k - \left(\frac{U_{2j}}{\sum_{j=1}^{n_2} U_{2j}}\right)\right) > 0$ ensures that all coefficients (c_{2j}) are positive. The bowling ability indices are defined as $c_{2j}^1 = \left(\frac{V_{2j}}{\sum_{j=1}^{n_2} V_{2j}}\right) \times n_2$ to ensure location distribution of the coefficient is equivalent to that of the batting coefficient. This transformation means that the higher the bowling index, the better the player, a requirement for the (maximisation) objective function.

To ensure that the spread of the batting and bowling indices are equivalent the index scales are adjusted. This is achieved using the scale adjustment in Lemmer (2004) and utilised by Lourens (2009). The scale adjustment of the bowling index is obtained iteratively by calculating a new index. This index is a function of the previous index relative to the ratio of the standard deviations of the batting index to the bowling index. This approach was proposed in Lemmer (2004) and used in Lourens (2009). The function is defined for the $(p+1)$th bowling index as $c_{2j}^{p+1} = (c_{2j}^{p})^{\left(\sigma_{c1}/\sigma_{c_2^p}\right)}$, where σ_{c1} is the standard deviation of the batting index and $\sigma_{c_2^p}$ is the standard deviation of the bowling index for the pth iteration. No alternative scale adjustment measure was sourced in the literature hence the use of the Lemmer (2004) measure. Once the iterative procedure converges to an acceptable lower limit, the $(p+1)$th bowling index is used as the coefficient index in the objective function, that is $c_{2j}^{p+1} = c_{2j}$. The indices for batting and bowling can now be used as coefficients in the objective function of a programming optimisation model.

2.5. All-rounder

In cricket, some players possess both batting and bowling skills. These players are commonly referred to as all-rounders. The all-rounder index is defined as a product function of batting and bowling indices, that is $c_{3j}^1 = (c_{1j})^{\beta}(c_{2j})^{1-\beta}$, where $0 \leqslant \beta \leqslant 1$. The parameter β is the weighting factor that determines the importance of the batting and bowling indices. The scale adjustment method is also used to ensure that all-rounder indices are equivalent to batting and bowling. The iterative process follows the same mechanism as the scale adjusted-bowling index.

Similar indices are possible for other abilities that are deemed essential for team selection. In Gerber and Sharp (2006) an attempt to justify the use of a fielding index was suggested. This study makes no such attempt as there is insufficient data available to quantify the index. An alternative approach, used in this study, is to include a wicketkeeping constraint based on the wicketkeeper's batting index.

3. An example: the inaugural Twenty20 World Cup 2007

To illustrate the proposed model, data from the 2007 inaugural Twenty20 World Cup were captured and the statistics for each player obtained. These statistics were used to select the optimal 11-person team for the tournament. Given that the statistics for each player have been determined from a small number of games, a few simple rules were used to reduce the chance of selecting an outlier. An outlier was considered to be an unusual occurrence, for example a tail-end bowler who, because of match circumstances, had a good batting average or batting strike rate. The rules for this tournament were motivated by the batting, bowling and all-rounder abilities. To reduce the likelihood of a non-specialist batsman being selected, only players who had played at least two games and had scored at least 20 runs were considered. Similarly, to reduce the likelihood of non-specialist bowlers being selected, each player had to have played at least two games, taken at least one wicket and bowled a minimum of eight overs. All-rounders were restricted to having played at least two games, scored at least 15 runs, bowled at least six overs and taken at least one wicket. These rules were used to exclude the outlier cases. These game rules are subjective and may be revised by the person applying the programming model.

3.1. The data

The data for the 186 players who took part in the tournament were captured and an Excel 2003 spreadsheet established. Tables 1–3 summarise the player statistics for the 10 best batting, bowling and all-rounder indices, respectively. Table 4 summarises the player statistics for the five best batting indices for players classified as wicketkeepers.

3.2. Batting top 10 indices

These indices were calculated for the batting performance function $U_{1j}=(Y_{1j}^{\alpha})(Y_{2j}^{1-\alpha})$ by setting $\alpha=1/3$; alternative choices for α may provide alternative rank orders. Barr *et al* (2008) provide illustrations to this effect with rank positions changing by four positions or less for their data. The choice of α is dependent on the user; it can be argued that for Twenty20 cricket, for teams to set a competitive score, the strike rate is more important than batting average, and hence the 2:1 weighting in favour of strike rate to batting average. Arguments for the choice of α can be given based on the format of the game. In the two innings per team format, with an unlimited number of overs allowed, batting average is considered more important than strike rate. However, as the number of overs decreases, strike rate becomes more important, thus requiring that the weighting between the two statistics changes. Research on the choice of α has been undertaken by Brettenny (2010) in a fantasy league team selection setting. Brettenny (2010) optimised the correlation coefficient between fantasy league scores and performance measures. The results reported showed little support for a definitive approach to selecting the optimal α and provide a clear area for further research.

The results in Table 1 show that of the top 10 batsmen, two players were considered to be all-rounders while one player was considered to be a wicketkeeper. All 10 players had strike rates exceeding 100, implying that they scored their runs quickly. In addition, nine of the 10 players had batting averages exceeding 40. The exception was the middle order batsman, Yuvraj Singh, with a batting average of 29.6. This case is worth mentioning as Singh tends to bat late in the innings limiting the number of balls that he has to strike. Despite this restriction, his strike rate was

Table 1 The batting statistics and indices of the 10 best-ranked batsmen

Rank no.	*Surname and initial*	*Games*	*Runs*	Y_{1j}	Y_{2j}	U_{1j}	c_{1j}
1	Gayle, CH	2	117	58.5	195.0	130.5	1.711
2	Hayden, ML	6	265	88.3	144.8	122.8	1.609
3	Sharma, RG	4	88	88.0	144.3	122.3	1.603
4	Taylor, BRM	2	107	107.0	127.4	120.2	1.575
5	Kemp, JM	5	173	86.5	139.5	119.0	1.559
6	Mubarak, J	5	105	52.5	169.4	114.6	1.502
7	McMillan, CD	6	163	40.8	181.1	110.2	1.444
8	Singh, Y	5	148	29.6	194.7	103.9	1.362
9	Gibbs, HH	3	110	55.0	142.9	103.9	1.362
10	ul-Haq, M	7	218	54.5	139.7	102.1	1.338

Table 2 The bowling statistics and indices of the 10 best-ranked bowlers

Rank	*Surname and initial*	*Games*	*Over*	*Wick*	Y_{3j}	Y_{5j}	Y_{4j}	U_{2j}	V_{2j}	c_{2j}^{p+1}
1	Vettori, DL	6	24	11	5.33	13.09	11.64	8.11	0.04802	1.4191
2	Gul, U	7	27.4	13	5.60	12.77	11.92	8.31	0.04773	1.4049
3	Clark, SR	6	20	11	6.55	10.91	11.91	8.64	0.04725	1.3819
4	Singh, RP	6	24	12	6.33	12.00	12.67	8.84	0.04696	1.3682
5	Morkel, M	5	20	9	6.00	13.33	13.33	8.94	0.04680	1.3607
6	Malinga, SL	5	14	7	7.14	12.00	14.29	9.67	0.04573	1.3103
7	Afridi, S	7	28	12	6.71	14.00	15.67	9.97	0.04528	1.2897
8	Fernando, C	5	17	6	6.12	17.00	17.33	10.25	0.04487	1.2709
9	Pathan, I	6	22	9	6.77	14.67	16.56	10.27	0.04484	1.2692
10	Razzak, A	5	19	7	6.37	16.29	17.29	10.34	0.04474	1.2648

Table 3 The statistics and indices of the 10 best-ranked all-rounders

Rank	*Surname and initial*	*Games*	*Runs*	*Overs*	c_{1j}	c_{2j}	c_{3j}^{p+1}
1	Jayasuriya, ST	5	154	12.5	1.346	1.0879	1.351
2	Morkel, JA	5	120	10	1.338	1.0896	1.346
3	Afridi, S	7	91	28	1.222	1.1799	1.335
4	Sreesanth, S	6	20	23	1.133	1.0519	1.149
5	Vettori, DL	6	36	24	0.964	1.2300	1.144
6	Schofield, C	4	24	12.5	1.020	1.1159	1.108
7	Flintoff, A	5	70	18	0.979	1.1366	1.088
8	Arafat, Y	3	44	9	1.315	0.8077	1.049
9	Pathan, I	6	34	22	0.876	1.1716	1.020
10	Vaas, C	5	33	18	0.869	1.1587	1.006

Table 4 The statistics and indices of five wicketkeepers according to batting index ranking

Rank	*Surname and initial*	*Games*	*Runs*	Y_{1j}	Y_{2j}	c_{1j}
1	Taylor, BRM	2	107	107.0	127.4	1.575
2	Gilchrist, A	6	169	33.8	150.9	1.201
3	Dhoni, MS	6	154	30.8	128.3	1.045
4	McCullum, BB	6	139	27.8	121.9	0.976
5	Prior, M	3	69	23.0	111.3	0.862

194.7; this means that on average for the tournament, he scored just less than two runs off every ball he faced, an exceptional scoring rate.

3.3. Bowling top 10 indices

These indices were calculated for the bowling function $U_{2j} = (Y_{3j}^{\alpha_1})(Y_{4j}^{\alpha_2})(Y_{5j}^{1-\alpha_1-\alpha_2})$ by setting $\alpha_1 = 1/2$ and $\alpha_2 = 1/4$; alternative choices for α_1 and α_2 may provide alternative rank orders. These weight adjustments determine which statistics are more valuable. It is argued that economy rate is more important than bowling average and strike rate, hence the ratio weighting of 2:1:1, respectively. Thereafter, the bowling index scales are spread adjusted using the previously defined function $c_{2j}^{p+1} = (c_{2j}^{p})^{\left(\sigma_{c_1}/\sigma_{c_2^p}\right)}$.

The results in Table 2 show that of the top 10 bowlers, two players were considered to be all-rounders. All but one player had an economy rate of less than seven, a good statistic in this format of the game. Interestingly, bowlers ranked one to six were all specialist bowlers lending credence to the argument that specialist bowlers are required for this format of the game.

3.4. All-rounder top 10 indices

These indices were calculated for the all-rounder index $c_{3j} = (c_{1j})^{\beta}(c_{2j})^{1-\beta}$ by setting $\beta = 1/2$. Alternative choices for β may provide alternative rank orders. In the context of this format of the game an even weighting to batting and bowling abilities is assigned. This cautious approach is motivated by the lack of information available to provide an alternative choice. Twenty20 cricket is in the early stages of development and as the game develops more informed choices should be possible. Thereafter, the all-rounder index scales are spread adjusted using the function $c_{3j}^{p+1} = (c_{3j}^{p})^{\left(\sigma_{c_1}/\sigma_{c_3^p}\right)}$, where $\sigma_{c_3^p}$ is the standard deviation of the all-rounder index for the pth iteration.

The results in Table 3 show that of the top 10 all-rounders, five players were considered to be all-rounders and five players were considered to be bowlers. This result is interesting as the empirical evidence emphasises the importance of selecting bowlers who are capable of batting.

3.5. Five best batting indices for wicketkeepers

The norm for most teams' selection policies is to select a specialist wicketkeeper. The integer programming model allows for this by including a wicketkeeping selection constraint. In addition, the model illustrated chooses the specialist wicketkeeper based on their batting index. This method adds depth to the teams, batting line-up while having no adverse implication for the programming model.

Table 4 summarises the batting performances of the best five wicketkeepers based on batting index. The results in Table 4 indicate that the Zimbabwean wicketkeeper BRM Taylor was the best wicketkeeper based on his batting performance. Unfortunately, Zimbabwe were eliminated after just two games and Taylor's statistics were based on a small sample. Notwithstanding, the criteria for batting eligibility allows for the qualification of Taylor, perhaps similar to that of the West Indian C Gayle; their omission from the remainder of the tournament was a loss for the spectators.

4. Integer programming model

To illustrate the method, the integer programming model is defined for the following decision variables:

$$\text{Let } x_{ij} = \begin{cases} 1, & \text{if player } j \text{ is selected for discipline } i \\ 0, & \text{otherwise} \end{cases}$$

$$\text{where discipline } i = \begin{cases} 1, & \text{if batting ability} \\ 2, & \text{if bowling ability} \\ 3, & \text{if all} - \text{rounder ability} \\ 4, & \text{if wicketkeeping ability} \end{cases}$$

$j = 1, \ldots, n_i$, the number of players with ability i, and c_{ij} = the index value for the ith ability of player j.

The objective function is defined as the maximization of $Z = \sum_{i=1}^{4} \sum_{j=1}^{n_i} c_{ij} x_{ij}$, subject to the following constraints:

The batting constraint $\sum_j (x_{1j} + x_{3j}) \geqslant 6$ ensures that at least six batsmen (ie either specialist batsmen or batting all-rounders) are selected. This restriction and the selection of a batting wicketkeeper provides batting strength for the optimal team.

The bowling constraint $\sum_j (x_{2j} + x_{3j}) \geqslant 5$ ensures that at least five bowlers (ie either specialist bowlers or bowling all-rounders) are selected. This restriction ensures that there are at least five players collectively capable of completing the 20 overs allotted for the match.

Although most teams have players with all-round ability, the restriction of a minimum requirement for all-rounders is unnecessary. The selection of all-rounders will be automatic if the number of bowlers, batsmen and wicketkeepers exceeds the allowable limits of the number of players in the team. Therefore, we have the all-rounder constraint $\sum_{j=1}^{n_3} x_{3j} \geqslant 0$.

The wicketkeeper constraint $\sum_{j=1}^{n_4} x_{4j} = 1$ ensures that only one specialist wicketkeeper is selected. The team restriction constraint $\sum_{i=1}^{4} \sum_{j=1}^{n_i} x_{ij} = 11$ ensures that exactly 11 players are selected for the team. Given that this is an illustration of the method, no consideration for a 12th man was included. This addition can be accommodated in the programming model by amending the constraint to include an additional player. To ensure no player is selected more than once the constraint $\sum_{i=1}^{4} x_{ij} \leqslant 1$ for all j is included.

5. Results and discussion

The results of this programming model were obtained using Solver, an optimisation routine included as an Excel add-in. Table 5 lists the team selection for this programming model using the data in Tables 1–4.

The optimal team selected for the Twenty20 World Cup in 2007 would have been the batsmen Hayden, Sharma and Mubarak, supported by the all-rounders Gayle and Kemp with wicketkeeping duties allocated to Taylor. The bowling attack would have consisted of Vettori, Gul, Clark, R Singh and M Morkel with additional support from the all-rounders.

The results are interesting when compared to the one-day (50-over format) World XI of Barr *et al* (2008). Their analysis supports the selection of Styris, Hayden, Clarke, Ponting, Gibbs, Gilchrist, McGrath, Bracken, Nel, Malinga and Tait. With the exception of Hayden there is very little consensus between the sides. There are several reasons for this; the player pool was different, for example Nel only played one Twenty20 game and the pre-defined rules excluded him from selection. The format of the game is different, which undoubtedly requires different skills and

Table 5 The optimal team for the defined constraints and game rules set

No.	*Surname*	*Category*
1	Gayle, CH	All-rounder
2	Hayden, ML	Batsman
3	Sharma, RG	Batsman
4	Taylor, BRM	Wicketkeeper
5	Kemp, JM	All-rounder
6	Mubarak, J	Batsman
7	Vettori, DL	Bowler
8	Gul, U	Bowler
9	Clark, SR	Bowler
10	Singh, RP	Bowler
11	Morkel, M	Bowler

team strategies. In Twenty20, a player who bats more aggressively is likely to get more batting opportunities, dependent on match circumstances. One observation is clear, Mathew Hayden is a batsman who can adapt his style to suit the format of the game, a very useful person to have in a cricket team.

6. Conclusion

The programming model provides a useful method for optimising team selection. The requirements that coefficients for different skills be equivalent are addressed using an indexing method, while constraints are dependent on the requirements of the match and team circumstances. This paper contributes to the ever-increasing literature on cricket performance evaluations by combining the useful techniques of Lemmer (2004), Gerber and Sharp (2006) and Barr *et al* (2008). The integer programming method proposed is more general than traditional ranking methods as it provides an opportunity to select an entire team. The approach followed in this study provides a method for team selection, which can be extended to a multistage team selection game. Using an integer programme for team selection has subsequently been used in a fantasy league game that requires updating the team as more information is made available (Brettenny, 2010). The limitations to this study are the lack of empirical justifications for the power coefficients when obtaining the performance measures U_{ij}. Limited theoretical arguments are stated, but this provides an avenue for future research. An additional area for further work includes the evaluation of other performance measures while a comparative study of what is done in practice with the theoretical outcomes of the method will be of use to practioners.

Acknowledgements—The authors wish to extend their appreciation to the anonymous referees, the Journal Editor and Dr Catherine Logie whose comments and suggestions improved significantly on an earlier version of this paper. This material is based upon work supported financially by the National Research Foundation, Grant numbers TTK2006061500023 and TTK2007041600001. Any opinion, findings and conclusions or recommendations expressed in this material are those of the author(s) and therefore the NRF does not accept any liability in regard thereto.

References

Albert J and Koning RH (2008). *Statistical Thinking in Sport*. Chapman and Hall/CRC, Taylor and Francis Group: Boca Raton, Florida.

Barr GDI, Holdsworth CG and Kantor BS (2008). Evaluating performances at the 2007 cricket world cup. *S Afr Stat J* **42**: 125–142.

Beaudoin D and Swartz TB (2003). The best batsmen and bowlers in one day cricket. *S Afr Stat J* **37**: 203–222.

Bracewell PJ and Ruggiero K (2009). A parametric control chart for monitoring individual batting performances in cricket. *J Quant Anal Sports* **5**(3): 1–19.

Brettenny WJ (2010). *Integer optimisation for the selection of a fantasy league cricket team*. Unpublished masters dissertation, NMMU, Port Elizabeth, South Africa.

Clarke SR (1988). Dynamic programming in one-day cricket—Optimal scoring rates. *J Opl Res Soc* **39**: 331–337.

Croucher JS (2000). Player ratings in one-day cricket. In: Cohen G and Langtry T (eds). *Proceedings of the Fifth Australian Conference on Mathematics and Computers in Sport*. Sydney University of Technology; NSW: Australia, pp 95–106.

De Silva MB and Swartz TB (1997). Winning the coin toss and the home team advantage in one-day international cricket matches. *N Z Stat* **32**(2): 16–22.

Durbach IN and Thiart J (2007). On a common perception of a random sequence in cricket. *S Afr Stat J* **41**: 161–187.

Gerber H and Sharp GD (2006). Selecting a limited overs cricket squad using an integer programming model. *S Afr J Res Sport Phys Educ Recreation* **28**(2): 81–90.

Kimber AC (1993). A graphical display for comparing bowlers in cricket. *Teach Stat* **15**(3): 84–86.

Lemmer HH (2002). The combined bowling rate as a measure of bowling performance in cricket. *S Afr J Res Sport Phys Educ Recreation* **24**(2): 37–44.

Lemmer HH (2004). A measure for the batting performance in cricket players. *S Afr J Res Sport Phys Educ Recreation* **26**(1): 55–64.

Lemmer HH (2006). A measure of the current bowling performance in cricket. *S Afr J Res Sport Phys Educ Recreation* **28**(2): 91–103.

Lourens M (2009). *Integer optimisation for the selection of a Twenty20 cricket team*. Unpublished masters dissertation, NMMU, Port Elizabeth, South Africa.

Ovens M and Bukiet B (2006). A mathematical modelling approach to one-day cricket batting orders. *J Sports Sci Med* **5**: 495–502.

Preston I and Thomas J (2000). Batting strategy in limited overs cricket. *Statistician* **49**(1): 95–106.

Swartz TB, Gill PS, Beaudoin D and de Silva BM (2006). Optimal batting orders in one-day cricket. *Comput Opns Res* **33**: 1939–1950.

Swartz TB, Gill PS and Muthukumarana S (2009). Modelling and simulation for one-day cricket. *Can J Stat* **37**: 143–160.

Taha HA (2003). *Operations Research, An Introduction*. 7th edn, Prentice Hall: Upper Saddle River, New Jersey.

Van Staden PJ (2009). Comparison of cricketers' bowling and batting performances using graphical displays. *Curr Sci* **96**: 764–766.

Received November 2009;
accepted May 2010 after two revisions