South African Journal for Research in Sport, Physical Education and Recreation, 2012, 34(1): 13-28.
Suid-Afrikaanse Tydskrif vir Navorsing in Sport, Liggaamlike Opvoedkunde en Ontspanning, 2012, 34(1): 13-28.
ISBN: 0379-9069

A MULTI-STAGE INTEGER PROGRAMMING APPROACH TO FANTASY TEAM SELECTION: A TWENTY20 CRICKET STUDY

Warren J. BRETTENNY*, David G. FRISKIN*, John W. GONSALVES** & Gary D. SHARP*
**Department of Statistics and **Department of Mathematics and Applied Mathematics, Nelson Mandela Metropolitan University, Port Elizabeth, Republic of South Africa*

ABSTRACT

Team selection is a controversial topic, even more so when a team performs poorly. Many sport fans believe they can perform the selection process better than those tasked with the responsibility. With the developments of fantasy sport games, fans now have a platform to test their claims, albeit in a purely recreational manner. Participants in fantasy league sport games can select their own team and based on the players' performances are awarded points. The participant with the highest points is declared the winner. This study proposes a multi-stage integer optimisation algorithm for a fantasy team selection. The sequential capability of the automated selection algorithm is demonstrated in a Twenty20 cricket tournament.

Key words: Integer programming; Cricket; Statistics; Team selection; Fantasy leagues.

INTRODUCTION

Twenty20 cricket is a new and exciting short form of cricket, which has gained much support since its introduction in 2003 (Weaver, 2008). Since this format of the game is only a few years old, it provides a new focus area for research of many kinds. This paper adds to the emerging research into this topic by providing a team selection strategy for a Twenty20 fantasy league. Fantasy leagues provide a platform for fans of various sports to test their team selecting abilities in competition with each other. The modern fantasy league game involves the selection of a "fantasy" team at the beginning of a season and then managing this team during the course of the season. In 1980, Daniel Okrent (Vichot, 2009) developed this format of the game with the introduction of "Rotisserie League Baseball". Vichot (2009) mentioned that fantasy league participants at this time were presented with many of the team selection problems facing professional coaches. As such, fantasy league participants were required to make changes to their team to account for injury and poor player performance.

The game of cricket has provided a lucrative research area for many researchers. Studies into the mathematics and statistics of the game and of the players have resulted in numerous publications. Arguably, the most prominent of these was the study of Duckworth and Lewis (1998) who provided a mathematical technique to reset targets in rain affected matches. This initial approach, as well as the subsequent adjustments to the approach, has had a large and lasting impact on the limited overs game. The reader who is unfamiliar with the game of cricket is referred to Preston and Thomas (2002) for an overview of the game. Barr and

Kantor (2004) and Lemmer (2004), have conducted research into the performance measures of batsmen. Lemmer (2002) studied the bowling ability of cricketers. The performance measures of both batsmen and bowlers were represented graphically by Van Staden (2009). Simulation techniques have been used by Swartz *et al.* (2006) to determine optimal batting strategies. Swartz *et al.* (2009) then devised a method to simulate entire one-day cricket matches. These simulations are very useful for studying the game. Gerber and Sharp (2006) devised an integer-programming model, which selects a squad of cricketers according to the values of their performance measures. A similar technique was used by Sharp *et al.* (2010) to select an optimal team from the data of the inaugural International Cricket Council (ICC) World Twenty20 held in South Africa in 2007. The current research modifies the integer-programming model of Sharp *et al.* (2010) to select and manage a fantasy league team. This is done in a Twenty20 cricket framework and illustrated using data from the 2008 Indian Premier League (IPL).

FANTASY LEAGUES AND PROBLEM STATEMENT

Fantasy leagues are available for many sports and are becoming a popular pastime for millions of people across the world. Many of these leagues are hosted on the Internet and are usually based on a popular competition or tournament. The host websites manage the league and determine all the rules and regulations for participants. The participants of a fantasy league act as a virtual team manager. Each participant is given a budget of virtual funds and is required to purchase players for their team according to predefined conditions. The players available to be purchased are the sport personalities from the competition on which the fantasy league is based. The cost of the players varies according to the perceived ability of the player. Naturally more skilled players will cost more than less skilled players. The participants then manage their team throughout the course of the competition. The real life performances of the sport personalities are assessed and points are awarded to each player after every match. High scores are given to players who perform well and low scores to those who perform badly.

Typically a fantasy league will be divided into “stages” which consist of a group of matches. Fantasy league participants are thus required to assess recent player performance and make necessary changes to their team in order to maximise the number of points scored at each stage. To perform well in a fantasy league usually requires comprehensive research and planning (Allen *et al.*, 2007). Classically there are strict limitations on the number of changes a fantasy manager is allowed to make. The team with the highest cumulative points at the end of the season is then deemed the winner. Much research has been conducted into fantasy leagues but this research typically focuses on sport popular in the United States of America such as baseball (Allen *et al.*, 2007) and American football. Summers *et al.* (2008) researched optimal drafting techniques in hockey pools, which is another popular North American sport. The current study exposes the game of cricket to similar research.

The Twenty20 competition, which this study used for illustrative purposes of the algorithm, was the inaugural IPL, which was held in India in 2008. This competition was the first of its kind and involved cricketers from all over the world. The current study used a modified version of the rules used for the fantasy league competition hosted on the CricInfo website (CricInfo IPL Fantasy League, 2008). The rules used in this study were:

- no team can cost more than 1 million units (each player costs between 50 000 and 150 000 units),
- each team must consist of 11 players,
- each team must have one of the following combinations:
    - 4 batsmen, 4 bowlers, 2 all-rounders and 1 wicket keeper, or
    - 4 batsmen, 3 bowlers, 3 all-rounders and 1 wicket keeper,
- a maximum of 9 changes can be made to the team,
- the tournament is broken into 4 stages, and
- changes made to a team must be concluded before the commencement of each stage and the team chosen at the start of a stage remains the same until the end of the stage.

The four stages partition the tournament into groups of 16, 14, 14 and 15 games respectively. This results in each team/cricketer participating in approximately four games per stage. The last three rules are unique to this study. These rules are included to partition the competition into a game which allows for a simple illustration of the mathematical programming routine.

The research problem for this study can now be defined. A team of 11 players had to be selected prior to the commencement of the 2008 IPL. The total cost of the players in the team cannot exceed 1 million units at any stage of the tournament. The team must be selected according to the formation restrictions of the game. Once the tournament had commenced, a further nine changes could be made to the team provided none of the game restrictions were violated. The team had to be selected and modified over the four stages in order to maximise the number of points scored by the players. This study proposes a mathematical optimisation procedure that provides a solution to this problem.

METHODOLOGY

The team was selected for each stage of the competition using a binary integer optimisation procedure. This optimisation procedure constructs the team by selecting the players with the highest values of their respective performance measures subject to the constraints imposed by the fantasy league. The performance measures used in this study were similar to those proposed by Croucher (2000) and modified by Barr and Kantor (2004) and again by Sharp *et al.* (2010). These performance measures were selected as they have already been used for team selection and assessment purposes. As such, their effectiveness in this setting has been established. There is on-going research identifying cricketer’s performance measures. The measures used in this study were for ease of application. The primary objective of the paper was the mathematical algorithm for sequential selection.

The measure used for a batsmen was the weighted product of the batting average AVE_{BAT}, and the batting strike rate, SR. This measure can be represented as $BAT = (AVE_{BAT})^{1-\alpha}(SR)^{\alpha}$ where α is the weighting constant. Values of α greater than 0.5 thus imply a greater weighting to the batting strike rate. In Twenty20 cricket the ability to score runs quickly is desirable and thus a higher weighting should be given to the batting strike rate. As such, a value of $\alpha = 0.75$ is used.

The bowling measure used was similar to the batting measure, however, the bowling measure was a weighted product of the bowling average, AVE_{BWL}, and the bowling economy rate, EC. This measure is a variation of the measure proposed by Sharp *et al.* (2010). The bowling measure was thus represented as $BWL = (AVE_{BWL})^{1-\beta}(EC)^{\beta}$ where β was the weighting constant. Values of β greater than 0.5 imply a greater weighting to the bowling economy rate. Owing to the importance of limiting the runs scored in the Twenty20 game, a greater weighting should be allocated to the bowling economy rate. As such, a value of $\beta = 0.75$ was used.

The measure used for all-rounders was a weighted product of the batting and bowling measures (BAT and BWL). Each measure in this instance was given an equal weighting. The measure used for all-rounders was thus represented as $ALR = (BAT)^{0.5}(BWL)^{0.5}$. The measure used to select wicket keepers was identical to that of a batsman. This measure depends solely on the batting performance of a cricketer and was chosen since the bowling statistics (AVE_{BWL} and EC) of a wicket keeper are often not available. Thus, the wicket keeper measure was provided as $WKR = (AVE_{BAT})^{0.25}(SR)^{0.75}$.

The performance measures described above used the career statistics for each cricketer to select the team at the first stage of the competition. The performance measures for each cricketer were then updated as new information became available. The updated performance measures were now calculated using only the data from the most recent stage in the tournament. This allowed the optimisation algorithm to use the recent data to select the cricketers who performed best. The performance measures used in this study were thus calculated sequentially at the beginning of each stage, only using the data available at the time of selection. The team selections were thus made in real time, so as to recreate the fantasy league environment.

To include these performance measures into a single integer program, each measure must be standardised. This was accomplished using the method proposed by Lemmer (2004) and used by Gerber and Sharp (2006) where each measure was divided by its mean value. The resulting measures were then unit less and all had the same mean value of 1. The measure for: batsman *i* for stage *k* thus became

$$F_{ik} = \left(\frac{BAT_{ik}}{\sum_{j=1}^{n_{Fk}} BAT_{jk}} \right) \times n_{Fk}$$

where $BAT_{ik} = (AVE_{BAT_{ik}})^{1-\alpha}(SR_{ik})^{\alpha}$ was the performance measure for batsman *i* using the batting average and strike rate as calculated for stage *k* and n_{Fk} was the number of batsmen stage in *k* for which the measure BAT_{ik} could be determined. The measures for each bowler, all-rounder and wicket keeper for stage *k* (denoted as G_{ik}, H_{ik} and I_{ik}, respectively) were similarly calculated.

This standardisation method, however, did not account for the difference in variance between the categories of cricketers. In order to equate the variances the method proposed by Lemmer

(2004) was employed. This method selects a base standard deviation, σ_b, and then equates all the variances to this base value. This is accomplished using an iterative process which raises each performance measure within a certain player category to the power $\left(\frac{\sigma_i}{\sigma_b}\right)$, where σ_i is the standard deviation of the measures in that category for the previous iteration. This process is continued until the standard deviation for each category of cricketer (batsmen, bowlers, etc.) was equal to σ_b, correct to the third decimal place. This process was conducted for each category of cricketer for each stage of the competition. The base value was arbitrarily chosen to be the value of the variance of G_{i1} that is, the performance measures calculated for bowlers for the first stage.

To model the optimisation problem, the decision variables are defined as

$$x_{ik} = \begin{cases} 1 & \text{if player } i \text{ is selected for stage } k \\ 0 & \text{if player } i \text{ is not selected for stage } k \end{cases}$$

Since all the performance measures have been standardised, they can all be included into a single set *A*. This set will thus contain the performance measures for batsmen, bowlers, all-rounders and wicket keepers. Let a_{ik} be the performance measure of cricketer i prior to stage k. Let the competition consist of *n* registered players and be partitioned into q stages. The objective function for the optimisation procedure at each stage can thus be expressed as $\max\left\{Z = \sum_{i=1}^{n} a_{ik} x_{ik}\right\}$ where $k \in \{1,...,q\}$. Each optimisation procedure is solved under a number of constraints. The constraints must be expressed mathematically in order to be included in the optimisation procedure. Each cricketer taking part in the competition was classified as a batsman, bowler, all-rounder or wicket keeper prior to the competition. These classifications are expressed mathematically using the following methodology.

Define the variable

$$b_{ik} = \begin{cases} 1 & \text{if player } i \text{ is a batsman in stage } k \\ 0 & \text{if player } i \text{ is not a batsman in stage } k \end{cases}$$

This categorisation is constant throughout the competition and thus $b_{ij} = b_{il}$ for all $j, l \in \{1,...,q\}$. Similarly the variables for bowlers, all-rounders and wicket keepers are defined as

$$c_{ik} = \begin{cases} 1 & \text{if player } i \text{ is a bowler in stage } k \\ 0 & \text{if player } i \text{ is not a bowler in stage } k \end{cases},$$

$$d_{ik} = \begin{cases} 1 & \text{if player } i \text{ is an all rounder in stage } k \\ 0 & \text{if player } i \text{ is not an all rounder in stage } k \end{cases}, \text{ and}$$

$$e_{ik} = \begin{cases} 1 & \text{if player } i \text{ is a wicket keeper in stage } k \\ 0 & \text{if player } i \text{ is not a wicket keeper in stage } k \end{cases}$$

respectively. At each stage exactly four batsmen must be chosen, this constraint is expressed as

$$\sum_{i=1}^{n} b_{ik} x_{ik} = 4 \quad \forall\, k \in \{1,\dots,q\}.$$

Similarly the constraint for the number of bowlers selected at each stage is expressed as

$$3 \le \sum_{i=1}^{n} c_{ik} x_{ik} \le 4 \quad \forall\, k \in \{1,\dots,q\}.$$

For all-rounders the constraint is expressed as

$$2 \le \sum_{i=1}^{n} d_{ik} x_{ik} \le 3 \quad \forall\, k \in \{1,\dots,q\},$$

and for wicket keepers as

$$\sum_{i=1}^{n} e_{ik} x_{ik} = 1 \quad \forall\, k \in \{1,\dots,q\}.$$

The total number of cricketers selected for a team must be 11 at each stage of the competition. This constraint is expressed as

$$\sum_{i=1}^{n} \left(b_{ik} x_{ik} + c_{ik} x_{ik} + d_{ik} x_{ik} + e_{ik} x_{ik} \right) = 11 \quad \forall\, k \in \{1,\dots,q\}.$$

The budget of 1 million units sets an upper limit for each stage of the competition. Let the variable p_{ik} be defined as the price of cricketer *i* in stage *k*. The price of a cricketer is constant throughout the competition and thus $p_{ij} = p_{il}$ for all $j, l \in \{1,\dots,q\}$. To ensure that the cost of the selected cricketers never exceeds the budget, the following constraint is included:

$$\sum_{i=1}^{n} p_{ik} x_{ik} \le 1000\,000 \quad \forall\, k \in \{1,\dots,q\}.$$

The final constraint relates to the total number of changes that are allowed to be made to the team. Only changes at the end of each stage are counted. That is, only changes made prior to Stage 2, ... , *q* are counted. Changes made after stage *q* are not included as this coincides with the conclusion of the tournament and changes made at this point are irrelevant. Changes made prior to Stage 1 are unlimited as this coincides with the initial selection of the team. This implies that there are $q-1$ opportunities to make changes. Let the maximum number of changes allowed for the entire tournament be *T*. It is proposed that the maximum number of changes allowed at each of the $q-1$ stages is $t_k = \frac{T}{q-1}$, where $k = 2,3,\dots,q$, if the solution is an integer. If the solution of $\frac{T}{q-1}$ is not an integer then the following formula is used

$$t_k = \begin{cases} \operatorname{int}\left(\dfrac{T}{q-1}\right) & \text{for } 1 < k \le q-1 \text{ and } k \text{ even} \\ \operatorname{int}\left(\dfrac{T}{q-1}\right)+1 & \text{for } 1 < k \le q-1 \text{ and } k \text{ odd} \\ T - \sum_{i=2}^{q-1} t_i & \text{for } k = q. \end{cases}$$

The integer program must thus limit the number of changes made at each stage to a maximum of t_k. To count the number of changes made at each stage the following summation is required, $\sum_{i=1}^{n} x_{ik} x_{i(k-1)}$ where $k \in \{1,...,q\}$. The decision variables x_{ik} and $x_{i(k-1)}$ indicate cricketer i's selection in (or omission from) the team in the k^{th} and the $(k-1)^{\text{th}}$ stage respectively. Since x_{ik} is a binary variable, we have that $x_{ik} x_{i(k-1)}$ is also a binary variable. In fact, $x_{ik} x_{i(k-1)} = 1$ indicates that cricketer i was selected for both stages and $x_{ik} x_{i(k-1)} = 0$ either indicates that a change was made, or that the cricketer was not selected for either stage. The summation $\sum_{i=1}^{n} x_{ik} x_{i(k-1)}$ thus indicates the total number of cricketers selected in the team for both the $(k-1)^{\text{th}}$ and k^{th} stage. Since a maximum of 11 cricketers are selected in the team, the value $m_k = 11 - \sum_{i=1}^{n} x_{ik} x_{i(k-1)}$ indicates the number of changes made to the team between stages $(k-1)$ and k. Since changes are only counted from the end of Stage 1, the changes made to the team prior to the competition are not counted. We thus set $x_{i0} = 1 \ \forall i$, this ensures that the number of changes made prior to Stage 1, $m_1 = 11 - \sum_{i=1}^{n} x_{i1} x_{i0} = 11 - 11 = 0$.

Suppose now that m_k changes are made to the team between stage $(k-1)$ and stage k and that $m_k < t_k$. This implies that an additional $t_k - m_k$ changes can be made later in the competition. The constraint for the number of changes allowed for stage k in the competition where $k \in \{2,...,q\}$ is thus expressed as

$$m_k = 11 - \sum_{i=1}^{n} x_{ik} x_{i(k-1)} \le \left(\sum_{j=2}^{k} t_j - \sum_{j=2}^{k-1} m_j \right)$$

where $\sum_{j=2}^{1} m_j = 0$. The integer optimisation procedure using the above constraints is run at each stage of the competition. The output of the procedure provides the team selected at each stage.

If a cricketer were to get injured or be absent for some reason, it is assumed that this knowledge is available before the stage begins. This is the case for many Australian and New Zealand cricketers who had international commitments during the course of the 2008 IPL. The fantasy league participants would know this information. Knowledge of injured cricketers would also be available to the participants through various media sources. To account for the missing players, cricketers are allocated a performance measure value of $a_{ik} = 0$ for every stage in which they were absent. This was included to encourage a change when a cricketer is absent. The option of setting the decision variable to $x_{ik} = 0$ was disregarded as this may result in an infeasible solution when the constraint on the number of allowed changes is considered.

DATA

Twenty20 career data were collected for each cricketer who took part in the competition. These data were collected from the CricInfo website (CricInfo, 2008). The 2008 IPL data were also collected from the CricInfo website (CricInfo, 2008).

TABLE 1: **BATTING: FANTASY LEAGUE SCORING**

Event	Points	Event	Points
Per run scored	1	$150 \leq SR < 180$	10*
Six hit	2	$180 \leq SR < 200$	20*
Duck	-10	$SR \geq 200$	30*
$0 \leq SR < 50$	-30*	Reaching 25 runs	10
$50 \leq SR < 75$	-20*	Reaching 50 runs	20
$75 \leq SR < 100$	-10*	Reaching 75 runs	40
$100 \leq SR < 150$	0*	Reaching 100 runs	80

*These points are only awarded if the batsman scores at least 20 runs.

Using a fantasy league scoring methodology the overall performance of a cricketer in a match is given a single numerical value. High values of a fantasy league score correspond to good performances and low values to poor performances. The scoring methodology used in this study was developed by CricInfo (CricInfo IPL Fantasy League, 2008) and is given in Tables 1 to 3.

TABLE 2: **BOWLING: FANTASY LEAGUE SCORING**

Event	Points	Event	Points
Dismissing a: batsmen, all-rounder or wicket keeper	25	$7 \leq EC < 9$	0+
		$9 \leq EC < 11$	-10+
		$11 \leq EC < 14$	-20+
		$EC \geq 14$	-30+
Dismissing a bowler	10	Taking 2 wickets	10
Maiden over	40	Taking 3 wickets	20
$0 \leq EC < 3$	30+	Taking 4 wickets	40
$3 \leq EC < 5$	20+	Taking 5 wickets	80
$5 \leq EC < 7$	10+		

+These points are only awarded if the bowler has bowled at least 2 overs.

TABLE 3: **FIELDING: FANTASY LEAGUE SCORING**

Event	**Points**
Catch	15
Stumping	15
Direct run out	30
Indirect run out	10

To illustrate how points are awarded from these tables consider the following example: Suppose a cricketer scored 60 runs off 40 balls hitting 2 sixes. The cricketer also bowled 3 overs conceding 27 runs and taking the wicket of an opposition all-rounder. The points scored by the cricketer are given as:

- 60 points (1 for each run scored),
- 4 points (2 for each six hit),
- 10 points (since $SR = 150$),
- 20 points (for reaching 50 runs),
- 25 points (for dismissing an all-rounder), and
- -10 points (since $EC = 9$).

The total fantasy score for the cricketer for this example is thus $60+4+10+20+25-10=109$ points. This is a typical scoring system used in a fantasy league. A fantasy score thus provides a single numerical value to the overall performance of a cricketer in a given match. This value provides a convenient and informative measure of a cricketer's overall contribution to a match.

RESULTS AND DISCUSSION

The optimisation procedure was run in GNU R 2.10.1 using the package lpSolve version 5.6.4 (Berkelaar, 2008). The fantasy league game discussed in Section 2 is used and thus the optimisation algorithm considers a 4-stage game. The total number of changes allocated to each stage is $t_k = \frac{9}{3} = 3$ for stages $k = 2,3,4$. Unused changes at any stage of the competition would be passed on to the subsequent stage. The team selections considering each category of cricketer are discussed separately.

Batsmen

The batsmen selected for the fantasy team for each round are presented in Table 4. Before the start of the competition the batsmen with the best performance measures were M. Hayden, R. Ponting, R. Sharma and R. Taylor. This selection provided a total of 790 fantasy league points. As the tournament progressed and data from each player's performance in the first stage of the competition became available, R. Ponting was replaced by V. Sehwag. In Stage 1, V. Sehwag scored 254 points, almost 5 times more than R. Ponting.

TABLE 4: **BATSMEN CHOSEN FOR THE FANTASY LEAGUE TEAM**

Stage 1		Stage 2		Stage 3		Stage 4	
Name	**Pts**	**Name**	**Pts**	**Name**	**Pts**	**Name**	**Pts**
M. Hayden	331	M. Hayden	0	G. Gambhir	300	G. Gambhir	32
R. Ponting	51	V. Sehwag	364	V. Sehwag	222	V. Sehwag	68
R. Sharma	182	R. Sharma	298	R. Sharma	152	R. Sharma	165
R. Taylor	226	R. Taylor	93	S. Marsh	261	S. Marsh	678

Since a fantasy league score, in itself, is a fair reflection of the performance of a cricketer, one can conclude that V. Sehwag's performance in Stage 1 was far superior to that of R. Ponting. This resulted in V. Sehwag replacing R. Ponting in Stage 2. This was the only change in batsmen for this stage. The batsmen selected in Stage 2 provided an aggregate fantasy score of 755 points. The decrease owes itself to the selection of M. Hayden in this stage, even though he was no longer participating in the tournament. The optimisation procedure, in this case, determined that it would be more beneficial to make changes to other members of the team. The poor performance by R. Taylor in Stage 2 (only 93 points) and M. Hayden's absence from the competition resulted in their replacement by S. Marsh and G. Gambhir. These batsmen scored 324 and 343 points respectively in Stage 2. These performances increased the performance measures sufficiently for S. Marsh and G. Gambhir's inclusion in the team for Stage 3. The aggregate fantasy score for Stage 3 for the batsmen was 935. This increase lends support to the benefits of using the model. No changes were made to the batsmen for Stage 4. Although G. Gambhir and V. Sehwag did not perform particularly well, the inclusion of S. Marsh resulted in an aggregate score for the batsmen in Stage 4 of 943. This is an improvement on the result of the previous stage indicating useful selections were made.

Bowlers

The bowlers selected for the fantasy team for each round are presented in Table 5.

TABLE 5: **BOWLERS CHOSEN FOR THE FANTASY LEAGUE TEAM**

Stage 1		Stage 2		Stage 3		Stage 4	
Name	**Pts**	**Name**	**Pts**	**Name**	**Pts**	**Name**	**Pts**
A. Kumble	25	A. Dinda	36	A. Dinda	161	A. Dinda	-10
D. Zoysa	25	D. Zoysa	41	D. Zoysa	-10	L. Balaji	45
M. Mural'ran	95	M. Mural'ran	95	M. Mural'ran	5	A. Mishra	87
D. Vettori	25	D. Vettori	45	–	-	S. Tanvir	374

The optimisation algorithm selected A. Kumble, D. Zoysa, M. Muralitharan and D. Vettori for the first stage. This suggests that these cricketers performed well in competitions prior to the 2008 IPL. This selection provided a total of 170 fantasy league points. This is not an ideal result, and provides some evidence that the bowling performance measures do not provide a good reflection of a bowler's ability in a fantasy league setting.

The bowlers selected in Stage 2 are similar to Stage 1 except for a single change where A. Dinda replaced A. Kumble. A. Dinda scored 170 points in the first stage, resulting in his selection for Stage 2. A. Dinda then performed poorly in Stage 2, only scoring 36 points. The selection of D .Zoysa and D. Vettori in Stage 2 despite poor performances in Stage 1 is likely a result of the constraint on the number of changes allowed between stages, as well as both cricketers having good initial performance measures. The total number of fantasy points scored in this stage is 217, which, although poor, is an improvement on the selection for Stage 1.

D Vettori is replaced by an all-rounder in Stage 3.This is a consequence of his absence from the remainder of the tournament. The rest of the bowlers remain unchanged. The resulting fantasy score for this stage is 156. Although this is a reduction from the previous stage, it must be noted that only three bowlers are considered. M Muralitharan and D Zoysa performed badly in Stage 3, resulting in their replacement by L Balaji and A Mishra in the fourth and final stage.

A. Dinda's inclusion in Stage 4 is a result of a good performance in Stage 3. The most notable change in Stage 4 is the inclusion of S. Tanvir, who had been performing well throughout the competition. S. Tanvir's score of 374 in this stage pushed the total points scored by bowlers in the final stage to 496. S. Tanvir's inclusion identifies both a limitation and an advantage of the approach used in this study. The advantage is that he was included and the algorithm noticed his performances. The limitation is that he had been performing well throughout the tournament and was only included in the final stage, despite scoring 389 and 215 points in Stage 2 and 3 respectively. When compared with batsmen, the selected bowlers under performed. This is observed by the total fantasy scored of the batsmen (3 423) who score 2 384 points more than those of the bowlers (1 039). This possibly indicates that the measure used to determine the performance of bowlers was less suitable for fantasy league selection purposes.

All-rounders

The all-rounders selected for the fantasy team for each round are presented in Table 6.

TABLE 6: **ALL-ROUNDERS CHOSEN FOR THE FANTASY LEAGUE TEAM**

Stage 1		**Stage 2**		**Stage 3**		**Stage 4**	
Name	**Pts**	**Name**	**Pts**	**Name**	**Pts**	**Name**	**Pts**
Y. Pathan	271	Y. Pathan	312	Y. Pathan	162	Y. Pathan	475
P. Kumar	128	P. Kumar	303	P. Kumar	2	–	–
–	–	–	–	S. Watson	341	S. Watson	373

Y. Pathan and P. Kumar were selected for the first Stage of the tournament and both performed well, scoring 271 and 128 fantasy points respectively. These good performances resulted in their inclusion in the fantasy team for the second stage. Both all-rounders scored over 300 points in this stage. The continued inclusion of P. Kumar in Stage 3 was a result of his performance in Stage 2. The interesting inclusion was that of S. Watson in Stage 3. He scored 466 and 275 points in Stage 1 and 2 respectively, resulting in him replacing a bowler (D. Vettori) in Stage 3. P. Kumar's poor performance in Stage 3 (only scoring 2 points) resulted in his exclusion from the team in the final round.

The all-rounders selected for the final round were S Watson and Y Pathan, who both scored over 1 200 points in the competition. This result provides some indication that the performance measure defined for all-rounders provided useful insight into the selection of a fantasy cricket team. The total number of fantasy points scored by the selected all-rounders was 2 367. This provides further evidence that the measure used for bowlers, which only resulted in 1 039 points, was not ideal for the selection of the fantasy league team.

Wicket keepers

The wicket keepers selected for the fantasy team for each round are presented in Table 7.

TABLE 7: **WICKET KEEPERS CHOSEN FOR THE FANTASY LEAGUE TEAM**

Stage 1		**Stage 2**		**Stage 3**		**Stage 4**	
Name	**Pts**	**Name**	**Pts**	**Name**	**Pts**	**Name**	**Pts**
L. Ronchi	101	A. Gilchrist	237	A. Gilchrist	111	A. Gilchrist	167

L. Ronchi, an Australian wicket keeper, was selected for the first stage. This was a result of good domestic Twenty20 performances prior to the start of the 2008 IPL. A. Gilchrist (also Australian) replaced him for the remainder of the competition. This was a result of A. Gilchrist scoring 384 points in Stage 1, as well as L. Ronchi taking no further part in the competition. As a result, the selected wicket keepers score was a total of 616 fantasy league points in the competition.

Fantasy league team

The team selected for the fantasy league and as each player's price, are presented in Table 8.

The budgeting constraint does not seem to influence the selections noticeably, as the budget amount of 1 million spent at each stage of the competition is 930 000, 965 000, 935 000 and 880 000. This indicates that the budget and pricing restrictions established by the organisers may make the fantasy league selection process too easy. This is illustrated by showing that if the total budget (1 million) is split evenly between the 11 selected players, then approximately 90 000 is available for each cricketer. This means that each cricketer in an average fantasy league team would be of the quality of M. Hayden, A. Kumble and Y. Pathan. All of these are very good cricketers. Ideally, the budget should allocate between 60 000 and 80 000 per cricketer. This would make the selection process more difficult, and thus lend support to an algorithmic selection method.

TABLE 8: **SELECTED FANTASY LEAGUE TEAM**

Stage 1		**Stage 2**		**Stage 3**		**Stage 4**	
Name	**Price**	**Name**	**Price**	**Name**	**Price**	**Name**	**Price**
M. Hayden	90 000	M. Hayden	90 000	G. Gambhir	105 000	G. Gambhir	105 000
R. Ponting	90 000	V. Sehwag	105 000	V. Sehwag	105 000	V. Sehwag	105 000
R. Sharma	100 000	R. Sharma	100 000	R. Sharma	100 000	R. Sharma	100 000
R. Taylor	75 000	R. Taylor	75 000	S. Marsh	55 000	S. Marsh	55 000
A. Kumble	85 000	A. Dinda	50 000	A. Dinda	50 000	A. Dinda	50 000
D. Zoysa	75 000	D. Zoysa	75 000	D. Zoysa	75 000	L. Balaji	60 000
M. Mural'ran	100 000	M. Mural'ran	100 000	M. Mural'ran	100 000	A. Mishra	100 000
D. Vettori	100 000	D. Vettori	100 000	S. Watson	75 000	S. Watsonr	75 000
Y. Pathan	85 000	Y. Pathan	85 000	Y. Pathan	85 000	Y. Pathan	85 000
P. Kumar	75 000	P. Kumar	75 000	P. Kumar	75 000	S. Tanvir	75 000
L. Ronchi	55 000	A. Gilchrist	110 000	A. Gilchrist	110 000	A. Gilchrist	110 000

Performance of the algorithm

The proposed algorithm resulted in a total score of 7 445 for the tournament. To determine the usefulness of the algorithm, it must be compared to alternative selection strategies. Since the four-stage fantasy league scenario used is unique to this study, a simulation approach must be used to generate other possible team selections. The simulation was conducted using two approaches. The first approach randomly selected a team for the initial stage of each simulation while the second approach used the initial team selected in this study for each simulation. Based on the rules described in this paper, 50 000 fantasy teams were generated for each of these approaches.

After each team was selected for the initial stage, changes were made to the team at the subsequent stages. The number of changes allocated to each stage was 3, and any unused changes were passed on to the following stage. The number of changes made at each stage was randomly selected from the number of allowed changes at the stage, where the possibility of making zero changes was included. The worst performing players at each stage were dropped. These players were randomly replaced by top performing players in each category. For a player to be considered a “top performer” they must achieve a fantasy league score in the top 10% of their category in the previous stage. Any unused changes in a particular stage were carried over to the subsequent stages. The total fantasy league score of each simulated team was then recorded.

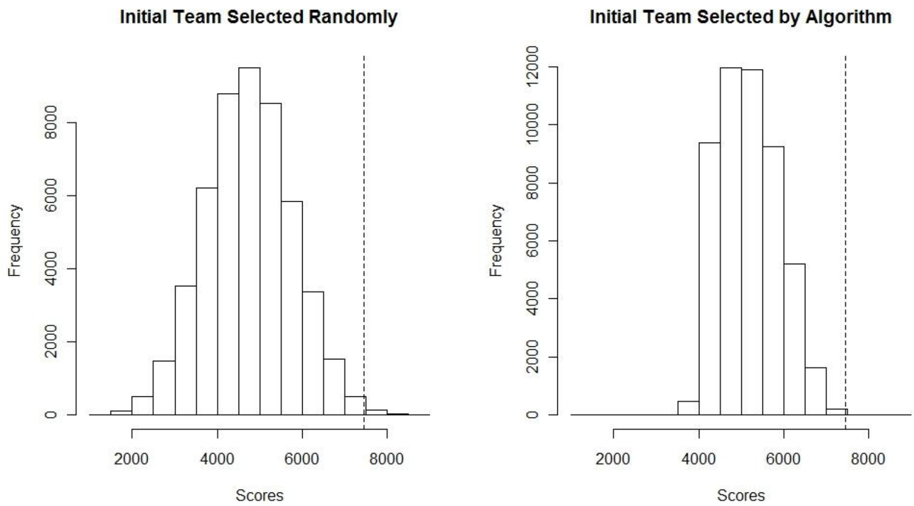

FIGURE 1: **DISTRIBUTION OF FANTASY LEAGUE SCORES***
(*Scores from 50 000 randomly simulated fantasy league teams using a randomly selected initial team and a fixed initial team [selected by the algorithm])

The distribution of these scores is given in Figure 1. For the first approach (randomly selected initial team), the distribution of fantasy league scores is fairly symmetric. The dotted vertical lines represent the score (7 445) obtained using the selection algorithm described in this paper. It is clear that this score lies near the maximum value from the simulations. In fact, this score places itself in the top 0.5% of the simulated scores. This result provides evidence that the selection method documented in this paper is competitive when compared to typical fantasy league selection strategies.

The results of the second approach (fixed initial team) indicate that the distribution in this case is more positively skewed and shifted to the right. Using this approach, the score of 7 445 places itself in the top 0.02% of the simulated results. Once again, there is evidence that the selection algorithm of this paper provides a highly competitive selection procedure when the initial starting teams are identical. The results of the simulated teams, in both cases, indicate that the algorithm proposed in this study provides useful and competitive results when selecting a fantasy league team.

CONCLUSION AND RECOMMENDATIONS

Fantasy leagues are growing in popularity, so too is Twenty20 cricket. This study combines these two emerging interests and provided an algorithm, which facilitates the fantasy league selection process. The total number of points scored using the proposed integer optimisation approach is 7 445.

This result (7 445) placed in the top 0.5% of the total fantasy league scores of 50 000 randomly simulated fantasy league teams. This indicates that the integer optimisation procedure provided useful and even competitive results in a fantasy league setting. Furthermore, no intuition is used in this study, only inputting data as it becomes available. This makes the procedure described in this study useful to participants with limited knowledge of cricket.

There are a few areas that require further investigation. The performance of the bowling measure indicates that possible improvements are required. This introduces an area for further research opportunities. Restricting the number of changes allowed per stage is a limitation of the algorithm. Adjusting the number of changes allowed per stage according to the magnitude of the increase of the objective function might provide a better solution to this problem. This, too, requires further investigation. Lastly, analysing the fixture list of a tournament is an important part of selecting a fantasy league team. Ideally if a strong team were to play a weak team, selecting cricketers from the stronger team might provide better results.

In conclusion, this study provides a sequential team selection procedure in a fantasy league setting. The study also provides a simple illustration of the implementation of the optimisation technique, and illustrates its effectiveness using computer simulation.

Acknowledgements

The authors thank the following for their financial support: the National Research Foundation (South Africa); the Dormehl-Cunningham trust and the NMMU Research Capacity and Development Department. A further word of thanks goes to Prof. Tim Swartz for his useful comments as an examiner of the original research. Lastly, thanks go to the anonymous referees for their valued input.

REFERENCES

ALLEN, I.E.; KUSTOV, K. & RECCK, G. (2007). Building a better fantasy baseball team. *Quality Progress*, 40(4): 24–29.

BARR, G.D.I. & KANTOR, B.S. (2004). A criterion for comparing and selecting batsmen in limited overs cricket. *Journal of the Operational Research Society*, 55(12): 1266–1274.

BERKELAAR, M. (2008). “LpSolve: Interface to Lpsolve v. 5.5 to solve linear/integer programs”. R package version 5.6.4. URL. [http://cran.ru.ac.za]. Retrieved on 11 March 2010.

DUCKWORTH, F.C. & LEWIS, A.J. (1998). A fair method for resetting the target in interrupted one-day international cricket matches. *Journal of the Operational Research Society*, 49(7): 220–227.

CRICINFO IPL FANTASY LEAGUE (2008). [http://fantasy.cricinfo.com/fantasy/fantasyleague/user/game_home.html?game_id=22]. Retrieved on 21 April 2008.

CRICINFO (2008). [www.cricinfo.com]. Retrieved from 17 April 2008 to 7 September 2008.

CROUCHER, J.S. (2000). Player ratings in one-day cricket. In G. Cohen & T. Langtry (Eds.), *Proceedings of the Fifth Australian Conference on Mathematics and Computers in Sport* (95-106). Sydney, NSW: Sydney University of Technology.

GERBER, H. & SHARP, G.D. (2006). Selecting a limited overs cricket squad using an integer-programming model. *South African Journal for Research in Sport, Physical Education and Recreation*, 28(2): 81–90.

LEMMER, H.H. (2002). The combined bowling rate as a measure of bowling performance in cricket. *South African Journal for Research in Sport, Physical Education and Recreation*, 24(2): 37–44.

LEMMER, H.H. (2004). A measure for the batting performance of cricket players. *South African Journal for Research in Sport, Physical Education and Recreation*, 26(1): 55–64.

PRESTON, I. & THOMAS, J. (2002). Rain rules for limited overs cricket and probabilities of victory. *The Statistician*, 51(2): 189–202.

SHARP, G.D.; BRETTENNY, W.J.; GONSALVES, J.W.; LOURENS, M. & STRETCH, R.A. (2010). Integer optimisation for the selection of a Twenty20 cricket team. *Journal of the Operational Research Society*, advance online publication, August 25, doi:10.1057/jors.2010.122.

SUMMERS, A.E.; SWARTZ, T.B. & LOCKHART, R.A. (2008). Optimal drafting in hockey pools. In J. Albert & R.H. Koning (Eds.), *Statistical thinking in sports* (263–276). Boca Raton, FL: Chapman and Hall/CRC, Taylor and Francis Group.

SWARTZ, T.B.; GILL, P.S.; BEAUDOIN, D. & DE SILVA, B.M. (2006). Optimal batting orders in one-day cricket. *Computers & Operations Research*, 33(7): 1939–1950.

SWARTZ, T.B.; GILL, P.S. & MUTHUKUMARANA, S. (2009). Modelling and simulation in one-day cricket. *Canadian Journal of Statistics*, 37(2): 143–160.

VAN STADEN, P.J. (2009). Comparison of cricketers' bowling and batting performances using graphical displays: Scientific correspondence. *Current Science*, 96(6): 764-766.

VICHOT, R. (2009). “A history of fantasy sports and its adoption by sports journalists.” [http://jag.lcc.gatech.edu/blog/2009/01/history-of-fantasy-sports-and-its-adoption-by-sports-journalists.html]. Retrieved on 19 November 2009.

WEAVER, P. (2008). “The Twenty20 revolution that turned innocent sideshow into a hunt for millions.” The Guardian, 26 July 2008. [http://www.guardian.co.uk/sport/2008/jul/26/twenty20 domestic.cricket1]. Retrieved on 27 February 2009.

Mr Warren J. BRETTENNY: Nelson Mandela Metropolitan University, Department of Statistics, P.O. Box 77000, Port Elizabeth 6031, Republic of South Africa. Tel.: +27 41 504 2895, Fax.: +27 41 504 2659, E-mail: warren.brettenny@nmmu.ac.za

(Subject Editor: Prof. Hoffie Lemmer)