Lancaster University Management School
Working Paper
2009/034

OR/MS applied to cricket

Mike Wright

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

OR/MS applied to cricket

There has been a steady trickle of academic papers over the past twenty years relating to the application of OR/MS to cricket. There are several reasons for this, including:

- cricket's widespread popularity
- the vast array of strategic and tactical decisions involved in playing cricket
- cricket's discrete nature which lends itself easily to being modelled
- the availability of enormous quantities of statistical data relating to professional cricket
- the large amounts of money involved in professional cricket
- the growing popularity of betting on cricket
- the complexity of managing cricket competitions
- the amenability of cricket timetabling and scheduling problems to OR/MS formulations and solutions

Moreover, among professional practitioners the game has recently become much more analytical. It is now commonplace to see not only administrators and managers but also players consulting their laptops in the pavilion during play. At the moment this is probably mainly used for video analysis of their own play and that of their opponents, but it opens the way for the more widespread practical use of OR/MS implementations before, during and after a match.

This paper starts by discussing cricket's popularity and outlining how a game of cricket works. It then moves on to discuss some of the way in which OR/MS can help decision-makers at various levels – players, managers and administrators.

1. Cricket's range and popularity

Cricket is played across the world, especially in English-speaking countries. Its popularity is highest in the Indian subcontinent, where it is commonly referred to as a religion. In India, "the country comes to a stop when a cricket match is being played – the roads are deserted, parties and weddings are postponed, operations in hospitals are rescheduled, parliament goes in for early closing" [1]. In Sri Lanka, it was reported that the imminence of a new world record temporarily stopped a war [2].

Cricket is generally regarded as the world's second most popular sport, though popularity is hard to define and even harder to measure. The number of players in the UK has been reliably estimated as 3.5 million [3]. With the population of the Indian subcontinent being about 30 times that of the UK and, it can be safely assumed, cricket being far more popular there than in the UK, this suggests that the number of players worldwide must be well over 100 million. The number of professional players in England is about 450 [4] – the worldwide total is probably around 2,000.

The number of cricket fans has been reported as between 2 and 3 billion [5], though it is not clear where this figure comes from or what the definition is of a fan. Also unreliable are the estimates for the largest TV audience, for the 2007 Twenty20 World Cup Final between India and Pakistan, which has been reported variously as 400 million [6] and 1.4 billion [7]. Perhaps most surprisingly, the North American TV figures have been reported as 37 million [7] – and North America is traditionally indifferent to cricket.

However much credence one may give to such unofficial and speculative figures, there is no doubt that cricket is of major importance to an enormous number of people and is thus a highly proper area for the serious application of the tools and approaches of OR/MS.

2. A brief description of a cricket match

Cricket is a sport played between two teams of eleven players on a large field. At any given time, one team is batting with two batsmen on the field, and the other is fielding with all eleven players on the field. One batsman is the "striker", who stands at one end of the "pitch" (a strip of well-rolled grass, matting or other material 22 yards long) and receives a delivery (or "ball") bowled by the "bowler" (one member of the fielding side) from the "non-striker's" end of the pitch; the other batsman stands at the non-striker's end. A "wicket-keeper" stands behind the striker's stumps and the other nine members of the fielding side are "fielders", placed anywhere on the field (though there may be some restrictions) by the team captain.

The batsman wields a bat in order to hit the ball and thus score "runs". When the ball is hit (or even if it is not) the batsmen may choose to attempt to score one or more "runs", by running from one end of the pitch to the other, as a result of which they may change ends. "Boundaries" may also be scored by hitting the ball out of play – this counts for four runs if the ball hits the ground before going over the boundary, six otherwise.

The fielding side aims to get batsmen "out" as well as trying to restrict the scoring of runs. There are ten distinct ways of getting a batsmen out, including "bowled" (where the ball hits one of three vertical pieces of wood known as "stumps"), caught (where the batsman hits the ball and a fielder catches it before it bounces) and "run out" (where the fielder hits the stumps with the ball before the batsman has completed his run). The dismissal of a batsman is known as a "wicket", or the fall of a wicket. (Cricket terminology can be very confusing to the layman – for example, the pitch is often referred to as the wicket, and the three stumps are also often called wickets.)

Balls are bowled in batches of six ("overs") by the same bowler, after which a different bowler must be used from the opposite end of the pitch. There may be limits to the number of overs any individual bowler may bowl.

Usually each team just has one "innings" in which they bat, but in some competitions each team has two innings. An innings ends when ten wickets have fallen, but possibly also after a given number of overs (typically 20, 40 or 50), or after a time limit. In addition, it is allowable and sometimes wise for a team captain to declare his team's innings "closed" early (see later). The team that scores the most runs wins, except that in some forms of the game a time limit may be overstepped before a result has been reached, in which case the game is a "draw". If both sides score the same number of runs, the match is a "tie" – this is quite unusual. A typical number of runs scored in an innings of a professional match may be as low as 100, but is frequently over 500 if there is no overs limit; the record low and high totals are 12 and 1,107.

One-innings matches (i.e. where each team has one innings) may take anywhere between two and seven hours. Two-innings matches usually last between three and five days (for "Test" matches between national teams).

For the full laws of cricket, which cover every game of cricket anywhere in the world, see [8]. Every cricket competition also has its own "rules" in addition to the laws. For non-competitive

matches the team captains agree the rules to be applied concerning, for example, the time or overs limit and any limits applying to individual batsmen or bowlers.

3. Decision-making, tactics and strategy

Cricket is full of decision-making at many levels. Every ball a host of real-time decisions are made.

The captain decides where to place the fielders, normally in consultation with the bowler. His decisions are influenced by the prowess of the batsman, the bowler and the fielders, by the state of the pitch, the weather and the state of the game. Professional captains have become more inventive with their field placings recently, making use of video analysis portraying the opposing batsman's strengths and weaknesses. In tense situations a captain may make changes every ball, but often the positions will remain the same throughout an over.

The bowler decides what kind of ball to attempt to bowl. Variables include ball speed, the location where the ball will bounce (if it does bounce), whether to apply swing, whether to induce lateral movement off the seam of the ball, whether to spin the ball and if so what type of spin – leg-spin, off-spin, googly, top-spin, "doosra" and possibly others. The more ambitious the intention, the less likely even a top-class bowler is to achieve exactly the desired effect, and most bowlers specialise in particular types of delivery (e.g. fast swing, slow spin).

The batsman has a split second to react to the ball that is hurtling towards him at up to 100 miles per hour. Not only does he have to decide whether to hit the ball or leave it, the types of shot he can play include the glance, glide, square cut, late cut, upper cut, pull, hook, cow-shot, slog, straight drive, on-drive, off-drive, cover drive, square drive, sweep, reverse sweep, slog sweep, paddle sweep, switch-hit, forward defensive, backward defensive, prod, nudge, nurdle, flick, scoop and possibly others – every now and then an unexpected new shot is invented. He must also decide how hard to hit it and with what degree of elevation. His decision will depend upon the pace, line (direction), length (where it bounces), height of bounce, spin, swing etc. of the delivery, upon his own strengths and weaknesses and upon the state of the game – whether he needs to play aggressively or defensively. When he has hit it, or even if he hasn't, he and his partner need collectively to decide whether to attempt one or more runs.

While there is no time for rational real-time analysis and batsmen react using a mix of reflexes, skill and experience, there are plentiful opportunities for rational OR/MS interventions in terms of the approach adopted by batsmen. Two examples are discussed below.

3.1 The balance between attack and defence

One of the most fundamental decisions for a batting team concerns the balance between attack and defence. An aggressive approach will score runs more quickly, but at a greater risk of dismissal. Conversely, with a defensive approach a batsman may well stay in longer, but his runs will come more slowly.

The optimal balance depends upon the state of the match – the number of overs remaining, the number of batsmen out and, for the team batting second, the number of runs still needed for victory. (This assumes that the match is of one innings and that there is a fixed limit to the number of overs available. The situation is a little more complex where there is no such limit, especially for two-innings matches, but the same general principles apply.)

This issue has been approached [9] by first creating a model of a limited-overs match, based on stochastic dynamic programming (DP – 1.6). A state of the system is defined by the number of runs required (for the team batting second), the number of wickets lost and the number of balls remaining to be bowled. Using data from actual matches, probabilities are assumed for the following events: one wicket and no run; neither wicket nor run; one run; two runs; three runs; four runs; and six runs. (This does not quite cover all eventualities: for example, it is possible for runs to be scored and a wicket to fall at the same ball, usually by means of a run out; it is possible to score five or even seven runs; and some deliveries may be "wides" or "no balls", resulting in an extra run but not counting towards the total number of balls to be bowled. However, it is adequate for the study in question.)

These probabilities vary depending upon the approach taken; under a very aggressive approach the probability of a wicket is high, the probability of scoring runs is high and the probability of neither run nor wicket (known in cricket parlance as a "dot ball") is very low; conversely, with a very defensive approach the probability of a wicket is low, of scoring runs is low and of a dot ball is high.

The study calculated the optimal scoring rate to aim for, from any given state of the system, in order to maximise the number of runs scored for the team batting first, or the probability of victory for the team batting second. The results showed that the prevalent strategy of starting slowly and then speeding up was sub-optimal, and that a more aggressive attitude should be adopted right from the start. Subsequent matches have supported this conclusion, and teams now usually show some aggression right from the start.

This analysis has been extended [10] using larger data sets and with variable probabilities – reasonable enough since not all batsmen or bowlers are of equal ability, and probabilities can also vary significantly with pitch and weather conditions. This can be used as a simulator (2.4) to answer all manner of strategic and tactical questions.

3.2 A strong batsman protecting a weaker one

An intriguing tactical battle often occurs when nine wickets have fallen, if a specialist batsman is accompanied by a much less proficient batsman (probably selected in the team for his bowling). It is up to the good batsman to score runs, since the other batsman is unlikely to score many, and if the weak batsman faces the bowling he has a high probability of being out, leaving the good batsman stranded; even though he is not out himself, the team will have lost ten wickets and thus the innings ends. So he will aim to face most of the balls bowled.

Since the bowling changes ends every over, his ideal is to score an even number of runs from each of the first five balls and an odd number on the sixth ball – generally a single run, since it is very hard to score exactly three runs on purpose. Thus he would face every ball. However, he cannot guarantee that the runs will come exactly as he wishes; and if he fails to score a single on the final ball the weak batsman may then have to face all six balls of the next over.

DP analysis [11] has concluded that the strong batsman should normally refuse to take an odd number of runs on the first four balls, but should accept a single run on the fifth or sixth ball, if nine wickets have fallen; otherwise both batsmen should take every run on offer.

However, this is also a tactical issue for the fielding side. If the strong batsman is on strike at the start of an over, the captain may place the fielders towards the boundary edge, such that it is easy for the batsman to score a single, but very hard for him to run two safely or to pierce the field for

a boundary four. He could try to hit the ball over the top of the fielders for a six, but this is often very risky – he could easily be caught out.

This creates a game of "cat and mouse", where for the first four balls the batsman may hit the ball to a faraway fielder and refuse to run, and for the fifth and sixth balls the captain will bring all the fielders in much closer to minimise the probability of a single. Batsmen may react by accepting a single off the fourth ball, or by trying to hit the ball over the fielders on the fifth or sixth, perhaps happy to accept a boundary four or six as compensation for not being the facing batsman at the start of the next over. The situation could be approached using game theory analysis (3.3).

4. Strategic decisions

The decisions above concern what to do with an individual delivery. Other decisions are more strategic. Those addressed by OR/MS analysts include decisions on team selection, batting order, whether or not to bat first, choosing a "night watchman" and whether and when to declare an innings closed. However there are plenty of other important decisions which have not yet been studied.

4.1 Team selection

This is an issue about which every fan of any sport has strong views, but very little headway has been made by OR/MS. There have been attempts [12] to devise measures of performance which measure a player's actual or potential contribution more accurately than those prevalently used, but there is much more work to be done. For example, highly sophisticated techniques have been developed to select football teams [13] and these are widely used among top clubs in the Netherlands. This is a target that OR/MS analysts could aim at for cricket.

4.2 Batting order

Normally the best batsmen go in early in a team's innings and the less able batsmen come in later. This is partly because it is wasteful to have a good batsman stranded if all ten wickets have fallen, and also because not all batsmen may be needed – this is especially important in games with a low over limit. One study showed how, making assumptions for transition probabilities for each batsman separately, any potential order can be simulated [14]; permutations of possible orders were generated using simulated annealing in order to find an order close to optimal.

A batting order does not have to be decided upon at the start of an innings, and a DP study [15] has shown that captains should make dynamic decisions depending on the state of the game. Another study [16] has considered situations where a day's play is nearly over when a wicket falls, for a team whose innings will continue the following day. In such circumstances some captains like to send in a "night watchman", who is good at defence but poor at scoring runs, to see out the day's play. This may be useful because any batsman, even a good one, is especially vulnerable at the start of his innings or at the start of a day's play, and such a batsman would then have to make two "starts". Some captains don't like to risk this for their good batsmen, whereas other captains think this unimportant and will send in their better batsmen ahead of their weaker batsmen in all circumstances. The study's results came down between these positions, concluding that in some circumstances a night watchman should be used and in other circumstances he should not.

4.3 Whether to bat or field first

Before the start of every match the captains toss a coin; the winner of the toss decides whether to bat or field first. The captain makes his decision based on factors including the current and predicted future condition of the weather and the pitch, plus perhaps factors relating to particular players. The toss is regarded as an important part of the game and pundits often put forward opposing views as to the right decision. Where it is not clear-cut, a toss can be regarded as a "good toss to lose", on the grounds that the captain cannot later be blamed for making the wrong decision.

Regression models [17] [18] have been used to analyse captains' choices. Results appear to indicate that for one-day matches it is generally better to field first unless the match is due to finish under floodlights, in which case it is preferable to bat first. However, in Test Matches it appears to be usually advantageous to field first. These results are surprising since most toss-winners choose to bat first in all forms of the game.

4.4 When to declare

Where the length of an innings is not limited by a number of overs, but instead there is an overall time limit, a captain may be faced with the decision whether and when to terminate his team's innings early, or "declare". The timing of a declaration is often very difficult. Too early and he may be giving the opposition too good a chance of winning. Too late and a draw may be almost inevitable – there will be insufficient incentive for the opposition to take risks, and not enough time for his bowlers to dismiss them.

This has been analysed using a multinomial logistic regression model [19] and a decision tool has been create with the potential to help second innings declaration decisions in Test matches. Declarations are also possible in a team's first innings – this again is an important decision but rather more difficult to analyse. Further OR/MS work would be very helpful here.

5 Decisions for administrators

5.1 How to determine a winner for a shortened match

Often play is deemed impractical or unsafe, usually because of heavy rain, a wet ground or poor light. This situation is improving as more of the top grounds install improved covers and devices to remove surface water, and more install floodlights, but it still frequently impossible to play a match to its natural conclusion.

Often this is unproblematic: the match is declared a draw. However, in some circumstances it is important that a winner be declared – this is especially the case in knock-out matches, where the winner proceeds to the next round. If there has been a reasonable amount of play then it may be possible to determine a winner even though the normal winning criteria may not have been met.

Several different methods were originally tried for this, but all led to very unjust outcomes at times. However, two OR analysts devised a system [20] which is now used throughout the world and accepted as fair by virtually all professionals, though probably few of them understand its intricacies. Given professional cricket's international following, it seems certain that the names of Duckworth and Lewis are far better known than those of any other OR/MS analyst in any field.

The precise details are quite complex, and the model requires calibration using real data, but the main feature of the Duckworth/Lewis method involves the consideration of wickets left and overs remaining as resources; thus, for example, when the number of overs remaining for the side batting second changes because of an interruption, the revised number of runs needed to win depends upon the overall resources available before and after the interruption.

A useful potential by-product of this system is that it can also determine putative winning margins [21] to be used as tie-breakers for teams with identical win records in a league.

Other methods have been devised [22] which aim to preserve win probabilities. While it is arguable that these may be superior methods, the cricket authorities are very unlikely to consider discarding Duckworth/Lewis.

5.2 Fixture scheduling

For some competitions, scheduling fixtures is easy, using fixed patterns. However, often a more analytical approach will be needed using computer algorithms. It may sometimes be appropriate to use theoretical models [23], but often the situation is too complex for anything other than a custom-built solution.

Sometimes it must be decided which teams are to play against each other, where, and how often, which can involve issues of fairness [24]. However, it is usually preordained which matches are to be scheduled; just the dates must be decided.

The most extreme case concerns English professional cricket. Eighteen teams take part in four different competitions – three one-day, one four-day – with different formats. The competitions overlap throughout the season; a team may frequently finish a match in one competition one day and start another match in a different competition the next day.

There are many constraints and objectives; also the teams have strong preferences which are far from identical for all counties. Travel considerations are important, especially late evening travel, as is the overall pattern of home matches. Teams don't like long gaps between matches; there are matches against touring international teams to fit in; some pairs of teams don't like to have home matches simultaneously; many teams have traditional "festivals", often at holiday resorts, for which the dates possible are usually very restrictive; often there are also dates to avoid if possible, such as when other sporting fixtures are involved, where the ground is required for a concert, etc.

This problem was successfully approached [25] using a variant of Tabu Search (1.7.1.3) which takes advantage of information derived from each separate objective as well as the overall objective.

Other successful implementations have been carried out for the World Cup [26] using a mixture of human decision-making and Integer Programming (1.4), for Australian professional cricket [27] using Simulated Annealing (1.7.1.1), for New Zealand professional cricket [28] using a mixture of Simulated Annealing (1.7.1.1) and tree search (1.4.2.3), and for an amateur league in England [29] using a partly manual heuristic (1.7).

5.3 Umpire scheduling

Every match requires two umpires out in the field. For televised matches there is usually another umpire who analyses replays of incidents to help the two in the field with difficult decisions. Moreover, for Test matches there is a fourth umpire mainly acting as a reserve in case another umpire falls ill.

Scheduling these umpires can be difficult. A successful implementation for English professional cricket [30] considered a variety of constraints and objectives and solved the problem using local search, though more recently an improved simulated annealing procedure (1.7.1.1) has been used. Objectives and constraints included journey time, balancing various types of match, limiting consecutive day's work and spreading umpires out between the teams and each other.

This can also be difficult in amateur leagues; in one implementation [31] an important consideration was the value of one umpire giving a lift to the other, so as to save money on petrol.

6. Forecasting and betting

This is a growing area for study and it is probable that much OR/MS analysis has been undertaken confidentially by betting companies. However, an interesting study [32] has been published using not only regression modelling but also the Duckworth/Lewis analysis. This considered betting during the course of a match, concluding that punters tend to overreact to important events (such as wickets falling).

7. Conclusion

Cricket is a very fruitful area for OR/MS study, with a plethora of decisions to be analysed, and a fast-growing data bank to help the analysis. Nowadays, for professional matches, a substantial amount of data is recorded for every ball, including field placements, the type of delivery, the type of shot played and the outcome. Moreover, issues for administrators get ever more complex as new types of tournament are created. We can look forward to many more papers on the subject.

References

(all web sites as accessed on August 20th 2009)

[1] http://news.bbc.co.uk/1/hi/programmes/3734038.stm
[2] http://www.cricinfo.com/columns/content/story/255212.html
[3] http://timesonline.typepad.com/line_and_length/2009/03/here-come-the-g.html
[4] http://www.independent.co.uk/sport/cricket/henry-blofeld-surfeit-of-counties-leaves-england-chasing-their-tail-579197.html
[5] http://ezinearticles.com/?Most-Popular-Sports-Around-The-World&id=551180
[6] http://blog.livecurrent.com/archives/102-T20,-IPL-among-biggest-sports-stories-of-2008.html
[7] http://news.bbc.co.uk/sport1/hi/cricket/7340331.stm
[8] http://www.lords.org/laws-and-spirit/laws-of-cricket/
[9] Clarke SR. Dynamic programming in one-day cricket – optimal scoring rates. Journal of the Operational Research Society 1988; 39(4):331-337.
[10] Swartz TB, Gill PS, Muthukumarana S. Modelling and simulation for one-day cricket. The Canadian Journal of Statistics 2009; 37(2):143-160.

[11] Clarke SR, Norman JM. To run or not to run? Some dynamic programming models in cricket. Journal of the Operational Research Society 1999; 50(5):536-545.
[12] Barr GDI, Kantor BS. A criterion for comparing and selecting batsmen in limited overs cricket. Journal of the Operational Research Society 2004; 55(12):1266-1274.
[13] Sierksma G. Computer support for coaching and scouting in football. Sports Engineering 2006; 9(4):229-249.
[14] Swartz TB, Gill PS, Beaudoin D, deSilva BM. Optimal batting orders in one-day cricket. Computers and Operations Research 2006; 33(7):1939-1950.
[15] Norman JM, Clarke SR. Optimal batting orders in cricket. Journal of the Operational Research Society online doi:10.1057/jors.2009.54
[16] Clarke SR, Norman JM. Dynamic programming in cricket: choosing a night watchman. Journal of the Operational Research Society 2003; 54(8):838-845.
[17] Allsopp PE, Clarke SR. Rating teams and analysing outcomes in one-day and Test cricket. Journal of the Royal Statistical society A 2004; 167(4):657-667.
[18] Dawson P, Morley B, Paton D, Thomas D. To bat or not to bat: an examination of match outcomes in day-night limited overs cricket. Journal of the Operational Research Society online doi:10.1057/jors.2008.135.
[19] Scarf P, Shi X. Modelling match outcomes and decision support for setting a final innings target in Test cricket. IMA Journal of Management Mathematics 2005; 16:161-178.
[20] Duckworth FC, Lewis AJ. A fair method for resetting the target in interrupted one-day cricket matches. Journal of the Operational Research Society 1998; 49(3):220-227.
[21] de Silva BM, Pond GR, Swartz TB. Estimation of the magnitude of victory in one-day cricket. Australian and New Zealand Journal of Statistics 2002; 43(3):259-268.
[22] Carter M, Guthrie G. Cricket interruptus: fairness and incentive in limited overs cricket matches. Journal of the Operational Research Society 2004; 55(8):822-829.
[23] Easton K, Nemhauser GL, Trick M. Pp 52.1–52.19 **in** Leung JT, ed., Handbook of scheduling: algorithms, models and performance analysis. 2004. CRC Press, Boca Raton.
[24] Wright MB. A fair allocation of county cricket opponents. Journal of the Operational Research Society 1992; 43(3):195-201.
[25] Wright MB. Timetabling county cricket fixtures using a form of tabu search. Journal of the Operational Research Society 1994: 45(7):758-770.
[26] Armstrong J, Willis RJ. Scheduling the cricket World Cup – a case study. Journal of the Operational Research Society 1993; 44(11):1067-1072.
[27] Willis RJ, Terrill BJ. Scheduling the Australian state cricket season using simulated annealing. Journal of the Operational Research Society 1994; 45(3):276-280.
[28] Wright MB. Scheduling fixtures for New Zealand Cricket. IMA Journal of Management Mathematics 2005; 16(2):99-112.
[29] Johns S. Complexity in amateur sports scheduling. Presented at the Operational Research Society Conference, Bath, 2001.
[30] Wright MB. Scheduling English cricket umpires. Journal of the Operational Research Society 1991; 42(6):447-452.
[31] Wright MB. Case study: problem formulation and solution for a real-world sports scheduling problem. Journal of the Operational Research Society 1997; 58(4):439–445.
[32] Bailey M and Clarke SR. Predicting the match outcome in one day international cricket matches while the game is in progress. Journal of Sports Science and Medicine 2006; 5:480-487.