Tactical sugarcane harvest scheduling

BJ Stray\textsuperscript{1}  CN Bezuidenhout\textsuperscript{2}  JH van Vuuren\textsuperscript{3}

June 23, 2010

Abstract

An ongoing sugarcane decision support research project in South Africa is aimed at developing a decision support system capable of assisting in deciding the seasonal harvesting sequence for sugarcane fields. Commercial growers have provided data suitable for regression modelling of the parameters that govern the values and costs involved, and have participated in two consecutive preliminary system evaluation and development efforts conducted during the 2009 and 2010 harvesting seasons. The optimisation models underlying the decision support system are based on a time-dependent travelling salesman problem formulation and are solved approximately by means of a tabu search. According to several of the industrial participants, the decision support system is useful to large-scale sugarcane producers.

Keywords: agriculture, travelling salesman problem, metaheuristics, harvest scheduling, sugarcane, decision support systems.

1 Introduction

Sugarcane harvesting in South Africa is governed by law as well as economic, political, environmental and contractual factors. Harvesting should occur at a consistent rate—so called \textit{daily rateable delivery}—throughout an approximately nine month long milling season providing a steady flow of feed-stock to the sugar mills. A sugarcane grower may produce several thousand to several hundred thousand tonnes of sugarcane over an entire season. Fields vary in size, and are usually larger for larger farms. Farms may contain as
few as 10 or as many as 100 fields or more. The fields should be harvested in a sequence that maximises the grower’s profit. Traditionally, this sequence is determined by first prioritising fields that must be harvested due to physical constraints or external events, such as harvesting fields lying on steep slopes outside of the wet season. The remaining fields are then sorted from oldest to youngest and ties are typically broken by choosing fields with higher sucrose content. Another consideration is to attempt to maintain a significant age difference between adjacent fields, both to combat fire risks and to reduce the spread of disease. Rönnqvist (2003) points to several similar adjacency factors which arise in forestry optimisation.

The Eston Mill region in the South African province of KwaZulu-Natal was selected as a representative case study area and four growers (who co-own a harvesting syndicate) agreed to supply the necessary data and to contribute towards schedule evaluation during the 2009 harvesting season, later also continuing during 2010. Grower groups are becoming more common in South Africa, and usually share both equipment and the responsibility to honour their mill contracts. These groups schedule their fields into a single or, perhaps, two parallel harvesting fronts in a manner that ensures equitable treatment of all. Since most grower groups are rather informally constituted, redesigning farm layouts in order to reduce the total number of fields is unlikely to occur. This situation gives rise to difficult scheduling problems due to the large number of fields involved. Furthermore, many extraneous events such as frost, diseases, insect invasions or drought may severely affect the growth rate or quality of sugarcane. Affected fields must be analysed and in many cases harvested sooner than originally planned. The aim of the research project described in this paper was therefore to build a flexible decision support system (DSS) capable of aiding growers in determining a suitable harvesting schedule in a group context, as well as for single farms.

The paper is organised as follows. The literature relevant to the project is briefly reviewed in the next section. This is followed by a description of the optimisation models forming the core of the DSS and a description of the parameter estimation methodology employed. Finally, the validation of the DSS is described, followed by a conclusion.
2 Literature Review

The complex nature of sugarcane production has led to several interesting operations research papers. In Australia, Higgins et al. (1998) and Muchow et al. (1998) attempted to optimise the harvest schedule of an entire mill region. The model parameters were supported by one-way analyses of variance of sugar yield which were used to find significant differences due to geographical location, variety, crop class and other factors. Muchow et al. (1998) concluded that there was potential to increase the profitability of the investigated mill region.

Higgins and Postma (2004) worked on the optimisation of siding rosters (the schedules used by Mackay Sugar to distribute time-slots for cane deliveries to railway sidings throughout the mill region) and discovered a solution to a staff roster problem, later implemented by the participating mill with cost savings estimated to approximately AUD 150 000 per annum. Higgins and Postma (2004, p. 237) pointed out the importance of allowing each grower fair harvesting time windows throughout the season, which “...prevents any grower from being unfairly exposed to wet weather risks towards the start or the end of the harvest season, as well as other seasonal effects.”

Higgins and Davies (2005) built a stochastic simulation model for the purpose of long term transport system capacity planning and considered several scenarios developed together with representatives from growers, harvesters and the miller in the Mourilyan mill region. One of the scenarios was deemed promising and was later implemented.

Higgins (2006) adopted a mixed integer linear programming approach towards modelling the transportation of sugarcane in the Maryborough mill region, which was later used by transport planners. He emphasised the importance of using a participatory research approach based on the work of Martin and Sherington (1997).

Grunow et al. (2007) considered the South American sugar supply network, where most farms are owned by the mill. The main objective of their mixed integer linear programming models were to ensure a minimum cane supply to the mill. Their hierarchical modelling procedure begins by solving a strategic model which assigns a cultivation date to each farm. A tactical level model is then solved in order to determine optimal ha-
ciendas (fields) to start harvesting for every day for two consecutive weeks. Finally, data updating is accommodated in a bid to address the occurrences of environmental events such as “criminal fires”, before an operational model generates dispatching schedules for crew and equipment.

In Thailand, the daily cane delivery quantities are not determined by contract as in South Africa, which leads to problems such as under-utilisation of the mill and sugar losses. Piewthongngam et al. (2009) developed a framework for field level cultivation planning using crop growth simulation models for a sugar yield estimation module and mathematical programming for a planning module.

3 Sequence Optimisation Models

Our DSS is designed to sequence the fields of an entire harvesting operation for an entire season, which approximately extends from April to December. The sequence is subject to monthly revision or updated whenever conditions have changed significantly. The formulation of the underlying tactical harvest sequencing problem (THSP) has to be valid in the situation where some fields in the sequence have been harvested earlier and some have been skipped. In other words, it must not become worthless just because it is not perfect.

The asymmetric travelling salesman problem with time-dependent costs (ATSPTDC) is a well-known variation of the celebrated asymmetric travelling salesman problem in which inter-city costs depend on how much time has passed. According to Albiach et al. (2008), the ATSPTDC is an extension of the asymmetric travelling salesman problem with time windows and it is capable of modelling time-dependency, both in terms of travel time and travel cost. This rendered the ATSPTDC very attractive in terms of a model for the THSP, since the cane yield and quality of a sugarcane crop changes with time. The ATSPTDC has indeed provided inspiration for model formulations in this section.
3.1 Integer Programming Model

Let $I = \{0, 1, 2, \ldots, n, n + 1\}$ be the set of fields that have to be harvested during the remainder of the harvesting season. Here the fields 0 and $n + 1$ are a dummy starting field and a dummy ending field, respectively, for a feasible harvesting sequence. The combined time required to harvest field $u \in I$ and to physically travel to any other field is denoted by $t_u$. Let $J = \{1, 2, \ldots, b_0\}$ be the set of time instants (days) into which the season is divided, where $b_0 = \sum_{u \in I} t_u$. Denote the profit from harvesting field $u \in I$ beginning at time instant $j \in J$ by $P_{uj}$. The parameters $P_{0j}$, $P_{n+1,j}$, $t_0$ and $t_{n+1}$ are assumed to be 0 for all $j \in J$. The decision variable $x_{uvj}$ is defined to take the value 1 if the harvesting operation leaves field $u \in I$ for field $v \in I$ at time instant $j \in J$, or the value 0 otherwise. Furthermore, let $y_u$ denote the time instant at which the harvesting operation leaves field $u$ and let $\mathbb{Z}^+$ denote the set of positive integers. The travelling times between fields are assumed to be negligible (this assumption is case-specific) and the time required to harvest a field is assumed to be independent of the time of harvest (case-specific). The objective of the THSP is then to

$$\text{minimise} \quad z = - \sum_{u \in I} \sum_{v \in I} \sum_{j \in J} P_{uj} x_{uvj}$$

subject to the constraints

$$\sum_{v \in I} \sum_{j \in J} x_{uvj} = 1, \quad u \in I \setminus \{n + 1\},$$

$$\sum_{u \in I} \sum_{j \in J} x_{uvj} = 1, \quad v \in I \setminus \{0\},$$

$$\sum_{v \in I} \sum_{j \in J} x_{n+1,v,j} = 0,$$

$$\sum_{u \in I} \sum_{j \in J} x_{u0j} = 0,$$

$$x_{uvj} = 0, \quad u = v \in I, \quad j \in J,$$

$$x_{0,n+1,j} = 0, \quad j \in J,$$

$$y_0 = 1,$$
\[ y_{n+1} - 1 = b_0, \]  
\[ y_u + t_u - M (1 - x_{uvj}) \leq y_v, \quad u \in I \setminus \{n + 1\}, \quad v \in I \setminus \{0\}, \quad j \in J, \]  
\[ \sum_{v \in I} \sum_{j \in J} j x_{uvj} = y_u, \quad u \in I \setminus \{n + 1\}, \]  
\[ x_{uvj} \in \{0, 1\}, \quad u, v \in I, \quad j \in J, \]  
\[ y_u \in \mathbb{Z}^+, \quad u \in I, \]  

where (1) computes the total harvesting operational profit, (2) ensures that all fields except the dummy ending field are “exited” exactly once and (3) ensures that all fields except the dummy starting field are “entered” exactly once. Constraint sets (4) and (5) ensure that the dummy ending field is not “exited” and that the dummy starting field is not “entered”, (6) forbids looping from a field to itself, while (7) ensures that the ending field is not “visited” immediately after the starting field. Constraint sets (8), (9), (10) and (11) ensure that the time of harvest is advanced by \( t_u \) time instants when the harvesting operation moves from field \( u \) to field \( v \). Finally, (12) ensures that the decision variable \( x_{uvj} \) is binary and (13) ensures that the decision variable \( y_u \) is a positive integer.

### 3.2 Integer Programming Solution

The THSP was formulated in Lindo System’s LINGO 11.0 (LINDO Systems, 2010) and several pseudo-randomly generated instances were solved on an Intel Core2 vPro 3 GHz PC processor with 4 Gb RAM. LINGO mainly uses branch-and-bound but also a relaxation induced neighbourhood search, a metaheuristic procedure designed to bring the advantages of local search to the realm of integer programming and mixed integer programming solution methodology (see Danna et al. (2005) for a detailed description). The sizes of these problem instances and their solution times are shown in Table 1. In solving instance D, LINGO found the optimal solution within two hours and the last six hours were spent improving the lower bound so as to establish optimality. No feasible solutions were found within 15 hours for instance E. As a practical size problem instance may involve more than 50 fields and 200 time instants, it became apparent that an alternative
<table>
<thead>
<tr>
<th>Instance</th>
<th>Fields</th>
<th>Instants</th>
<th>CPU time limit [s]</th>
<th>Feasible solution</th>
<th>Optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>14</td>
<td>3</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>28</td>
<td>15</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>C</td>
<td>7</td>
<td>21</td>
<td>30</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>D</td>
<td>8</td>
<td>21</td>
<td>29100</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>21</td>
<td>54000</td>
<td>N</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1: Various THSP instances were solved using LINGO 11.0. The time unit is seconds.

3.3 A Tabu Search Approach

The tabu search (TS) developed to solve the model (1)–(13) is primarily based on the work by Glover (1989) and inspired further by the paper of Yagiura et al. (2004). A feasible solution to the THSP is represented by a permutation vector \( \omega \) on the set \( \{1, 2, \ldots, |I|\} \) whose \( j \)-th component \( \omega(j) = i \) corresponds to field \( i \in I \) being assigned harvesting sequence position \( j \in I \). Let \( t_i \) denote the time required to harvest field \( i \in I \) and define the indexed set function \( U_i(\omega) \) to contain the set of fields in \( \omega \) that have a sequence value less than that of field \( i \) \( (i.e. \ U_i(\omega) = \{ \omega(u) \in I : u < j, \ \omega(j) = i \} ) \). Consecutive time instants for which profit values remain approximately constant are amalgamated as a time period in a bid to reduce the problem instance size. The set of time periods which thus arises is denoted by \( K \). Let the parameter \( d_k \) denote the first time instant of period \( k \in K \) and let \( P_{ik} \) denote the resulting profit from harvesting field \( i \in I \) during period \( k \).

Then the function

\[
\tau_i(\omega) = \sum_{\omega(u) \in U_i(\omega)} t_{\omega(u)}, \quad i \in I \tag{14}
\]

gives the time instant at which the harvesting of field \( i \) begins and the function

\[
\theta_i(\omega) = \min_{k \in K} \{ k : \tau_i(\omega) - d_k \geq 0 \}, \quad i \in I \tag{15}
\]
expresses the period during which field $i$ is harvested. Finally, if the indicator function \( \chi_{ik}(\omega) \) is defined to take the value 1 if \( \theta_i(\omega) = k \), or 0 otherwise, the objective function in (1) may be expressed as

\[
V(\omega) = \sum_{i \in I} \sum_{k \in K} P_{ik}\chi_{ik}(\omega). \tag{16}
\]

The set $K$ may be constructed by dividing the season into weeks instead of days, without significant loss of resolution.

The TS adopted to find a good solution \((\omega^*)\) to the THSP comprises several components: a random restart procedure (RR), a shift neighbourhood move (SN), an ejection chain compound move (EC), as well as procedures for handling tabu restrictions and aspiration criteria, all described below. The TS is mainly guided by two objective function values encountered during the course of its execution; these are the best objective function value in (16) encountered so far, denoted by $V^*$, and the best objective function value encountered since the last EC or RR application, denoted by $V'$. The TS begins by applying the SN to the current solution, which returns a new current solution $\omega$. If the objective function value of the new current solution is better than $V'$, an iteration counter denoted by $no_{\text{local\_impr}}$ is set to 0, otherwise this counter is increased by 1. If the objective function value of the solution is better than $V^*$, an iteration counter denoted by $no_{\text{global\_impr}}$ is set to 0. If $no_{\text{local\_impr}}$ is greater than a certain value, the EC is applied to the current solution and the iteration counter $no_{\text{global\_impr}}$ is increased by 1. If the counter $no_{\text{global\_impr}}$ is larger than a specified threshold, the RR is applied.

The random restart procedure (RR) generates a pseudo-random permutation of a range of integers from 1 to the total number of fields. The RR thus provides a new starting solution for the TS.

The SN finds the best admissible solution in a shift neighbourhood of the current solution. This shift neighbourhood is formed by considering a shift of any field to another position in the harvesting sequence, generating a trial solution, denoted by $\omega'$.

Define a neighbouring pair, denoted by $e = (i, j)$, where $i, j \in I$ and $i \neq j$, as any
two neighbouring fields (in terms of their position in the harvesting sequence) and let them together constitute a set denoted by $E$. Furthermore, assume that $\omega$ is “closed” by “connecting” its last field with its first field (i.e. by wrapping the solution encoding). A sequence $\omega$ of fields may then be partially described by a subset of $E$, constructed by including those neighbouring pairs which appear in $\omega$. For example, if $\omega = [2, 3, 1]$, the neighbouring pairs are $(2, 3)$, $(3, 1)$ and $(1, 2)$, the rightmost neighbouring pair $(1, 2)$ constituting the closing (or wrapping around) of the sequence.

The move from a current solution to a trial solution may now be described in terms of deleted and added neighbouring pairs. For example, if moving from $\omega = [2, 3, 1]$ to $\omega' = [3, 2, 1]$ (by shifting field 3 into position 1) none of the three neighbouring pairs $(2, 3), (3, 1)$ or $(1, 2)$ remain neighbouring pairs in $\omega'$, hence these neighbouring pairs are said to have been deleted. The three added neighbouring pairs are $(3, 2), (2, 1)$ and $(1, 3)$.

There are six so-called tabu lists, denoted by $T_1, T_2, T_3, T_4, T_5$ and $T_6$, respectively, which list neighbouring pairs that either may not be deleted or may not be added. When considering a move from $\omega$ to $\omega'$, the deleted neighbouring pairs are tested for membership in the tabu lists which contain neighbouring pairs that may not be deleted and the corresponding test is also performed for the added neighbouring pairs. During these membership tests, if any neighbouring pair $e$ is found in any of its corresponding tabu lists, an aspiration function, denoted by $A(e)$, is applied. If the aspiration criterion $A(e) < V(\omega')$ is met, the move receives a passing status with respect to the particular neighbouring pair. However, if the aspiration criterion is not met, the move is given a non-passing status. The function $A(e)$ is updated so as to contain the best objective function value obtainable by allowing the addition or deletion of neighbouring pair $e$. The trial solution $\omega'$ with the best objective function value that has achieved passing status for all associated neighbouring pairs is selected to be returned to the TS as the new current solution $\omega$. Then the neighbouring pairs that were added are added to the tabu lists that list neighbouring pairs that may not be deleted, while the neighbouring pairs that were deleted are added to the tabu lists that contain neighbouring pairs that may not be added. The tabu lists have specified lengths (tenures) and each time a neighbouring pair
is added to any of the lists, the oldest neighbouring pair is removed.

The reason for employing six tabu lists is that it provides for the possibility of enforcing more or less restrictiveness on moves based on the type of deleted or added neighbouring pair. If desired, the length of the list that retains the middle added neighbouring pair, for example, could be made longer or shorter as a means of disfavouring or favouring certain kinds of algorithmic behaviour. Alternatively, a passing status may be awarded to a move as long as the associated neighbouring pairs are not present in more than some fraction of the lists.

The EC rearranges the field harvesting sequence, employing rules that facilitate the preservation of parts of the quality of the sequence while seeking to diversify the TS. The EC starts by computing a parameter value for each field that measures the relative performance of the field during each period. This is achieved by means of the performance parameter

\[ \Upsilon_{ij} = \frac{P_{ij}}{\max_{k \in K} \{P_{ik}\}}, \quad i \in I, \quad j \in K. \]

Given a particular harvesting sequence \( \omega \), each field is assigned a current relative performance, as a result of its time of harvest, by means of (15). The field with the lowest current relative performance is deemed to have the largest potential for improvement. This field is ejected from the sequence, “opening” its position for another field. A sequence with such an “opening” is by convention called a reference structure (Yagiura et al.; 2004). The ejected field is reserved until the end of the application of the EC, when it is assigned to the then “open” position. The reference structure has an improvement potential \( g \) defined as \( g = 1 - \Upsilon_{ij} \). All other fields \( k \in I, \ k \neq i \) are now compared and the field with the largest value of \( \Upsilon_{kj} - \Upsilon_{k\ell} \), is selected to move from position \( k \) to position \( i \) in the reference structure (corresponding to moving from period \( \ell \) to period \( j \)), where \( \ell \) is the current period of the selected field and \( j \) is the period associated with the “open” position \( i \). This is called a reference structure move. The reference structure move only occurs if the condition \( \Upsilon_{kj} - \Upsilon_{k\ell} < -g \) holds. If the condition does not hold, the EC stops performing reference structure moves. However, if a reference structure move does indeed occur, \( g \) takes the value \( g + \Upsilon_{kj} - \Upsilon_{k\ell} \) and the EC continues
Table 2: THSP instances solved by means of the TS. The time limits are in seconds. Note that instances D and E did not require an increase in the time limit. Optimality was computed as what fraction of the lower bound remained as a gap between the lower bound and the objective function value of the best solution found. A practical size instance was run twice (F1 and F2) using a lower bound computed by summing the minimum cost during any period for each field, since there was no LP-relaxation solution available.

<table>
<thead>
<tr>
<th>Instance</th>
<th>Fields</th>
<th>Instants</th>
<th>CPU time limit [s]</th>
<th>Feasible solution</th>
<th>Optimality gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>14</td>
<td>10</td>
<td>Y</td>
<td>0.001100</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>28</td>
<td>15</td>
<td>Y</td>
<td>0.002480</td>
</tr>
<tr>
<td>C</td>
<td>7</td>
<td>21</td>
<td>30</td>
<td>Y</td>
<td>0.000788</td>
</tr>
<tr>
<td>D</td>
<td>8</td>
<td>21</td>
<td>30</td>
<td>Y</td>
<td>0.000100</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>21</td>
<td>30</td>
<td>Y</td>
<td>0.010351</td>
</tr>
<tr>
<td>F1</td>
<td>60</td>
<td>210</td>
<td>300</td>
<td>Y</td>
<td>≤ 0.0955</td>
</tr>
<tr>
<td>F2</td>
<td>60</td>
<td>210</td>
<td>900</td>
<td>Y</td>
<td>≤ 0.0954</td>
</tr>
</tbody>
</table>

The TS was coded in Microsoft’s Visual Basic for Applications for Excel, and the investigative computational results shown in Table 2 were obtained on the same machine as for the integer programming instances mentioned earlier. The results indicate that the solution times necessary to obtain acceptable solutions by means of the TS are stable for instances A–E. Instance F was solved twice (the two runs named F1 and F2, respectively) using different time limits, achieving results approximately 0.01 unit percentage apart in terms of optimality. Solutions to practical size instances were obtained on a regular basis during the case study reported later in this paper, uncovering satisfactory objective function values.
4 Optimisation Model Parameters

The model parameters $t_i$ and $P_{ij}$ ($i \in I, j \in K$) are case-specific. The harvesting times $t_i$ were estimated by assuming that the total area of the fields harvested is distributed consistently across the season. For example, if the season is 200 days long, the total area to be harvested is 200 hectares and a certain field has an area of 5 hectares, then the field will require 5 days to harvest. The computations required for determining the profit parameters $P_{ij}$ require the estimation of the value of a particular field for a set of future periods as well as the cost of harvesting the field during those periods. A sugarcane field contains a certain mass of sugarcane (cane yield) at a certain recoverable value percentage (RV %). A multiplication of the cane yield and the RV %, subsequently multiplied by the RV price (South African Canegrowers, 2010) gives the value of the sugarcane in South African Rands. The definition of RV is given in detail by Murray (2000), but may in simple terms be described as an estimated sugar potential from delivered cane.

4.1 Cane Yield

The prediction model used for cane yield is based on a regression approach and there is one model per variety. The data spanned the time period 2002 to 2009, came from four farms in the case study area and were sufficient to model the cane varieties N12, N16, N35 and N37. The response variable is in tonnes per hectare and the regressor is called effective growth time (EGT), which is the age of the cane in months less the time spent during months of slow or halted growth. In the case study area, June, July and August are generally classified as such “no-growth” months. These so-called base yield models are indexed over the field set $I$, and the time period set $K$. Let $B_{ij}^M$ ($i \in I, j \in K$) denote the response variable of a base yield model. The input needed to compute a $B_{ij}^M$-value for field $i$ during period $j$ are the variety present on the field and the future EGT of the field at the midpoint of period $j$. The EGT of field $i$ at the midpoint of period $j$ is denoted by $a_{ij}$. For example, a field containing the N12 cane variety, with $i = 4$, $j = 5$ and an
age of $a_{45} = 21.2$ months has the $B_{ij}^M$-value

$$B_{45}^M = \beta_1 a_{45} + \beta_{11} a^2_{45} = 11.7 \times 21.2 - 0.29 \times 21.2^2 = 117.7 \text{ [t.ha}^{-1}] ,$$

associated with it, where $\beta_1$ and $\beta_{11}$ are regression coefficients.

During the case study, it was found that other factors than the variety and the regressor are crucial in order to estimate the values of the model parameters realistically. These factors were partly incorporated by means of so-called event yield models. These event yield models are simple first-order polynomials where cane yield is a function of the number of days elapsed since some time marker. Denote the set of possible events by $L$. If an event $\ell_i(k) \in L$ takes place on field $i$, the time between its occurrence and the midpoint of period $j$ is first computed; denote this time by $q_{ij\ell_i(k)}$, where $k \in \{1, 2, 3\}$ is the order number of the event (only as many as three different events for any field are kept in record, assuming that any serious events will simply replace less serious ones, should a fourth or further event occur). Let $E_{ij}^M$ denote an event yield model value that adjusts cane yield in t.ha$^{-1}$ during period $j$ to account for events that have occurred on field $i$, let $\delta_{0,ij\ell_i(k)}^M$ denote the step cane yield decrease due to event $\ell_i(k)$ and let $\delta_{1,ij\ell_i(k)}^M$ denote the cane yield decrease in t.ha$^{-1}$d$^{-1}$ (tonnes per hectare per day after the day of the event) due to event $\ell_i(k)$. Then $E_{ij}^M$ may be expressed as

$$E_{ij}^M = \sum_{k=1}^{3} \left( \delta_{0,ij\ell_i(k)}^M + q_{ij\ell_i(k)} \delta_{1,ij\ell_i(k)}^M \right), \quad i \in I, \quad j \in K. \tag{17}$$

To compute the resulting cane yield $M_{ij}$ on field $i$ during time period $j$, the base yield model and the event yield model are added together to yield

$$M_{ij} = B_{ij}^M + E_{ij}^M, \quad i \in I, \quad j \in K. \tag{18}$$

4.2 Recoverable Value Percentage

In order to model RV %, different varieties present in the case study area were grouped according to similar characteristics. This grouping was only performed for varieties for
which data were insufficient. Secondly, a second-order polynomial regression model for predicted RV % was fitted using the number of the day of the year as the regressor. In this case study, there were not sufficient data to prove statistical significance of other regressors. Higgins et al. (1998) as well as Lawes and Lawn (2005) used similar approaches in Australia. Given the cane variety, the predicted RV % of a field may be computed for each period using the variety’s regression model. This predicted RV % was termed the base RV % model and denoted by $B_{ij}^R$.

In order to account for the effect on RV % caused by the occurrence of environmental and other events, the use of event RV % models was introduced, similar to the event yield models ($E_{ij}^M$). The notation used in the event RV % models mirror that of the event yield models, where $q_{ij\ell_i(k)}$ is the time in days between the occurrence of event $\ell_i(k) \in L$ on field $i$ at the midpoint of period $j$. The order number $k \in \{1, 2, 3\}$ retains the same significance as previously. Let $E_{ij}^R$ denote an event RV % model value that adjusts RV % during period $j$ to account for events that have occurred on field $i$, let $\delta_{0,ij\ell_i(k)}^R$ denote the step decrease due to the $k$th event and let $\delta_{1,ij\ell_i(k)}^R$ denote the RV % decrease per day after the day of the event due to the $k$th event. Then $E_{ij}^R$ may be expressed as

$$E_{ij}^R = \sum_{k=1}^{3} \left( \delta_{0,ij\ell_i(k)}^R + q_{ij\ell_i(k)} \delta_{1,ij\ell_i(k)}^R \right) \quad i \in I, \quad j \in K. \quad (19)$$

To compute the resulting RV % value $R_{ij}$ on field $i$ during time period $j$, the base RV % model and the event RV % model are summed together to yield

$$R_{ij} = B_{ij}^R + E_{ij}^R, \quad i \in I, \quad j \in K. \quad (20)$$

If $P$ is the season’s price per ton of recoverable value, $\bar{y}$ is the seasonal average percentage of recoverable value for the mill region and $\bar{w}_j$ is the average percentage of recoverable value for the mill region during period $j$, then the remuneration for harvesting field $i$ during period $j$ is given by

$$V_{ij} = PM_{ij}(R_{ij} + \bar{y} - \bar{w}_j), \quad i \in I, \quad j \in K. \quad (21)$$
In order to predict mill region RV %, data recorded by the South African Sugarcane Research Institute for the Eston Mill was employed. Each year from 1996 to 2008 was modelled using regression, forming a set of 13 possible models to use for prediction purposes. In order to predict a future season, the regression model is selected which belongs to the year with the most similar climatic history to the current year. This model is subsequently used to compute $\bar{y}$ and $\bar{w}_j$.

4.3 Costs

Harvesting sugarcane field $i$ during period $j$ was assumed to cost a certain basic amount $c_{ij}$ in South African Rands per hectare. Some fields are, however, less costly to harvest during dry weather, and this was modelled using a risk of wet conditions model based on queueing theory. Let $\rho_{ij}$ be the risk of wet conditions on field $i$ on a day during period $j$ and let $Q_i$ be the penalty, i.e. an estimate of the cost to the business, encountered if field $i$ is scheduled to be harvested on a wet day. The cost to the business mainly consists of the harvesting operation being forced to resort to loading the cane onto small tractor-trailer combinations that subsequently unload the cane onto the ground at so-called loading zones. It is not until this stage that the cane may be loaded onto long-range, large vehicles, designed to transport the cane to the mill. The zone-loading cost component is given by

$$c^d_{ij} = Q_i \rho_{ij}, \quad i \in I, \quad j \in K. \quad (22)$$

The value $\rho_{ij}$ may be computed if records have been kept of field conditions, or may be approximated using historical rainfall data combined with the estimation of dry-out rates of various field types. The penalty $Q_i$ depends only on the accessibility of field $i$, which for fields that are accessible by long-range, large vehicles when dry is equal to the cost of zone-loading and for fields which may never be entered by large vehicles is equal to zero. Such fields are always zone-loaded, so their $c^d_{ij}$s are fixed instead.

Our decision support system imposes penalties in the form of costs as a means of
discouraging schedules that are deemed beforehand to be undesirable. This was achieved
by introducing the concepts harvesting time window (HTW) and harvesting time deadline
(HTD) which specify the number of days to wait at most before harvesting some affected
field and the date before which to harvest some affected field, respectively. An HTD
arises, for example, if a grower decides to replant a field, in which case the farming
practice is to harvest and plough it early in the season in order to optimise total cane
yield. The field is in such a case associated with an HTD of 181, encouraging harvesting
to be scheduled before the end of June. An HTW arises, for example, if a field is invaded
by the destructive stalk borer *Eldana saccharina*. The field is in such a case associated
with an HTW of 20 due to the life-cycle of the insect. The HTD cost penalty is expressed
by $c_{ij}^{HTD} = P_L q_{ij} f_i(k)$, where $P_L$ is a per-day-per-tonne penalty coefficient and the HTW
cost penalty is expressed by $c_{ij}^{HTW} = P_L q_{ij} f_i(k)$. The value of $P_L$ reflects the importance
of adhering to the time windows and deadline dates. In summary, the cost component
$C_{ij}$ is given by

$$C_{ij} = c_{ij} + c_{ij}^{HTW} + c_{ij}^{HTD} + c_{ij}^d, \quad i \in I, \quad j \in K. \tag{23}$$

Given that $M_{ij}$, $R_{ij}$, $C_{ij}$, $\bar{y}$, $\bar{w}_j$ and the RV-price $P$ have been computed according to
the above procedures, the profit $P_{ij}$ should be computed as

$$P_{ij} = V_{ij} - C_{ij}, \quad i \in I, \quad j \in K. \tag{24}$$

5 Validation and Case Study

The DSS was used on a biweekly basis to generate actual harvesting schedules during the
2009 milling season and twice during the early 2010 season for the four growers participat-
ing in the case study. For every schedule, the manager of the harvesting operation
returned a form capturing general comments, harvested fields, fields which had suffered
from one of the twenty-three possible environmental or other events as well as an appraisal
of the suggested harvesting schedules.
The harvesting operation was divided into two *harvesting fronts*, called HF A and HF B, respectively. A harvesting front, in the sense intended here, is a set of fields that are harvested one after the other, and as a rule, no concurrent harvesting of any fields within the harvesting front is allowed. Each harvesting front was treated separately.

A measure of desirability was introduced, which may be called the *scheduling prediction accuracy* (SPA). The SPA is defined as the percentage of the fields scheduled for period 1 and period 2 combined (the first four weeks of the schedule), that were actually harvested within 6 weeks’ time from the first day of the schedule.

Geographically, HF A is situated a few kilometres north of Richmond, KwaZulu-Natal. The *area under cane* for HF A is approximately 150 hectares spread across 37 fields varying in size between 0.56 and 9.18 hectares. The varieties present in this harvesting front at the beginning of the 2009 milling season were N12, N16, N35, N37 and N41. HF A comprised two adjacent growers and all of their cane was delivered to Eston Mill, located approximately 35 kilometres away from the centre of the harvesting front.

HF B is situated approximately 5 kilometres north of HF A, the area under cane being approximately 300 hectares distributed across 79 fields ranging between 0.70 and 10.5 hectares. The varieties present were N12, N16, N23, N29, N35, N37, N40 and N41.

The SPA values for the first nine schedules suggested for HF A and the first ten schedules suggested for HF B during the 2009 season are shown in Table 3. The appraisal grade from 1 to 5, returned by the harvesting front manager, is also shown. SPA values were fairly consistent throughout the season, as were the appraisal grades. Most schedules received a grade of 3 or more, suggesting that schedules put out by the DSS were at least “acceptable”.

The grower response to the first schedule of the 2009 season was clear in that the schedule did not reflect the difference in maturity between varieties properly. The growers argued that varieties N35 and N37 were relatively mature compared to N12 and N16, and should be prioritised. Furthermore, the fifth set of schedules were generated after a number of adverse environmental events having occurred on all four farms. For example, five fields in HF A had lodged (blown down) and six had been ripened. The lodging was
Table 3: SPA values for HF A and HF B during the first five months of the 2009 season and during the three first months of the 2010 season, displayed by period. The appraisal grades are on a scale of 1 to 5 where 1 means “very poor”, 2 means “poor”, 3 means “acceptable”, 4 means “good” and 5 means “very good.”

6 Conclusion

The development of a tactical sugarcane harvest scheduling decision support system was described in this paper. The system incorporates practically all relevant issues that may arise when scheduling sugarcane harvest, and was validated and developed during two harvesting seasons in the Eston Mill area of South Africa. The system was found to provide sensible schedules under normal conditions and has proven some capability with respect to handling external events or decisions. In future work, validation and further model development may be conducted for cases with larger harvesting operations as well as with a larger number of case study participants and regions.

7 Acknowledgements

Work towards this project by the first author was funded by the School of Engineering, Industrial Engineering Department, University of Borås, Sweden, and is part of his PhD
dissertation at Stellenbosch University, while the third author was funded by the South African National Research Foundation in the form of a research grant (GUN 2072999). The project has received support from Clive Coulthard, Edgar Bruggeman, Eric Lewis, Illovo Sugar’s Noodsberg Mill, Malcolm Thompson, the Noodsberg Mill Group Board, Noodsberg Canegrowers, Roger O’Neill, Shawn Kyle and the South African Sugarcane Research Institute.

References


