DELIVERING THE CAPABILITY TO EVALUATE ALTERNATIVE CANE SUPPLY ARRANGEMENTS ACROSS THE SUGAR INDUSTRY USING A WHOLE INDUSTRY SYSTEMS APPROACH

A.J. Higgins[1,2], R.C. Muchow[2], M.A. Haynes[1,2], S.L. Peel[1,2], D.B. Prestwidge[2] and M.A. Williamson[2]

[1]CRC for Sustainable Sugar Production and [2]CSIRO Sustainable Ecosystems
Long Pocket Laboratories, 120 Meiers Road, Indooroopilly, Q 4068

Sugar Research and Development Corporation

FINAL REPORT
PROJECT Nº CTA044

November 2002

The Research Organisation is not a partner, joint venturer, employee or agent of SRDC and has no authority to legally bind SRDC, in any publication of substantive details or results of this Project

Sugar Research and Development Corporation

FINAL REPORT

Project Reference N°. CTA044

Project Title: Delivering the capability to evaluate cane supply arrangements across the sugar industry using a whole industry systems approach

SRDC Program: 8. Whole-of-Industry Competitiveness

Organisations: CRC for Sustainable Sugar Production (CRC-Sugar) and CSIRO Sustainable Ecosystems

Address: Long Pocket Laboratories
120 Meiers Road
Indooroopilly Q 4068

Project Supervisor: Dr R.C. Muchow (to March 2001)
Dr A.J. Higgins (from March 2001)

Location: Long Pocket Laboratories
120 Meiers Road
Indooroopilly Q 4068
Telephone: 07 3214 2340
Facsimile: 07 3214 2325
E-mail: Andrew.Higgins@ csiro.au

Commencement Date: 1 July 1999 **Completion Date:** 30 June 2002

ABSTRACT

Increasing cost/price pressure has forced the Australian sugar industry to seek innovative avenues for increasing profitability. To address this, the industry saw opportunities for increasing productivity and hence profitability through optimising the harvest date of sugarcane. Optimising the harvest date of sugarcane accounts for geographical and crop differences in cane yield and the sugar content of cane for different harvest dates throughout the harvesting season. Whole-of-system research within CRC-Sugar produced the statistical and optimisation models needed to conduct options analysis for these alternative cane supply arrangements using case studies in Mackay and Mossman. Options analysis using data from these case studies showed average potential gains in profitability of $119/ha for Mackay and $79/ha for Mossman, at a sugar price of $350/t. To provide the capacity for evaluating and implementing alternative cane supply options in other mill regions, computer tools, user interfaces and standardised databases were developed. The use of this software and documented processes was applied not only to the case study regions but also in partnership with other mill regions, namely Maryborough and the Burdekin. Key outcomes were the reduction in time from start-up to implementation reduced from four years down to one year, an increased uptake from growers and harvesting groups, ease of interpretation of schedules, and increased interest from other mill regions. A computer tool was also developed for the mathematical cane yield re-estimation model, and was used successfully by the Mossman cane inspector during the 2001 harvest season.

SUMMARY

The harvest date of sugarcane is a key determinant of sugar yield and hence net revenue. Limited mill crushing and cane transport capacity dictate that the harvesting of cane be carried out over a harvest season of several months. Opportunities exist for increasing industry profitability through alternative cane supply schedules that exploit differences in CCS and cane yield at harvest date, found across farms and across individual paddocks on a farm. Developing the capability to capture this opportunity required groundbreaking research combining four key components: operations research and statistical modelling for integration across the industry value chain; novel implementation pathways; industry cooperation; and grass roots participation.

The research to construct the combined statistical and optimisation model used for the cane supply options analysis, was carried out within CRC-Sugar Subprogram 3.2 and through participatory research with the growing and milling sectors in Mackay and Mossman. The model accounted for CCS and cane yield differences of farm, variety, crop class and harvest date. Costs of growing, harvesting, transport and milling were also incorporated. Capacities of harvesting, transport and milling were integrated into the optimisation model. These constraints were modelled using the expertise from the milling and growing sectors across the mill regions. The optimisation model gave rise to an enormous number of decision variables and required the development of novel heuristic mathematical solution techniques. This component of the work provided the capability to assess options exploiting the geographical and crop differences that occur in cane yield and CCS across harvest dates.

Through a broad-based industry workshop (March 1999, Townsville), with 145 representatives from across the Australian sugar industry, there was a general agreement that the gains in profitability were significant; other regions were keen to explore options as well and implementation should commence soon. To conduct options analysis in other sugar mill regions, the capacity needed to be developed through the construction of computer tools and standard procedures for new mill regions to follow. Implementation of alternative cane supplies was not a

simple plug-and-play process, since it requires significant change across the sugar industry sectors and the options produced by the model did not provide proof of concept alone. This is unlike a plant-breeding program, for example.

Increasing the capacity to deliver cane supply options to new mill regions (other than Mackay and Mossman) was initially achieved by standardising the model data and developing a model input data manual. This was used to efficiently collate model data when partnerships were developed with Maryborough, Tully and the Burdekin mill regions and allowed preliminary options analysis for alternative cane supplies to be delivered within six months of the start of the partnerships in 1999. The capacity was also increased through the development of "Database for Whole-of-Industry Models" (DWIM) which stored all model data (including data used for implementation) in a standard format that is easily accessible for other project staff and researchers. DWIM helped link all of the computer tools developed for the cane supply options and cane yield estimation. It provided a capability to efficiently update model inputs from raw industry data (for all mill regions) for implementation in 2000, 2001 and 2002, as well as the capability to meet the needs of a large-scale piloting process.

One of the key needs for piloting alternative cane supply arrangements in Mackay and Mossman, was to produce harvest schedules in a format that is easily interpreted by growers and harvester contractors for implementation but able to account for on-farm constraints (*e.g.* wet paddocks, grub damage) that are impractical for the model to capture. The software "HarvSched" was developed with industry partners. HarvSched is linked with DWIM and the optimisation model for efficient transfer of alternative cane supply arrangements from the optimisation model to final schedules that are easily interpreted by growers. This software was used by the local industry in Mackay, Mossman and Maryborough for producing harvest schedules with growers and harvesting groups since the 2000 harvest season.

Using the knowledge gained from producing DWIM, a user-friendly computer tool was developed with the Mossman cane inspector to generate cane yield re-estimates as the harvest season progresses. The software, which directly links with the Mossman Agricultural Services databases and is written in Visual Basic and Fortran, was installed and used at Mossman during the 2001 harvest season. The Mossman cane inspector found the software easy to use and it was able to add value to his work. Ultimately, it improved cane yield estimate over and above other available methods. After five weeks of the 2001 harvest season, the user-friendly model produced a mill level estimate for Mossman of 840,000 tonnes. The mill eventually cut 828,000 tonnes, and this small margin of error highlights the value of the software to the Mossman Mill Board.

BACKGROUND

Moving into the 21st century, increasing cost/price pressure has forced the Australian sugar industry to seek innovative avenues for increasing profitability. In response, the milling, harvesting and growing sectors have sought ways to make more profitable use of their capital, infrastructure and land resources. Past advances have come from improvements in technical and cost efficiency of different industry sectors, but the challenge is to achieve quantum gains in competitiveness by exploiting interdependencies across the value chain. This involves modelling the key drivers and constraints across the value chain from cane growing on-farm, to harvesting and transport systems to milling cane to produce raw sugar. No previous technology is available to the industry to address the complexity of value chain optimisation, nor pathways for implementation that accounted for the differences in the growing, harvesting and transport and milling sectors. Whole-of-industry research has been minimal and yet it offers scope for the delivery of even greater benefits.

One aspect of the whole-of-industry approach is concerned with cane supply management – an aspect that has implications for all three industry sectors, production, milling, and marketing. Deeply entrenched practices - such as operational cultures, use of capital and infrastructure, and equity arrangements for all cane growers, may have to be questioned if greater industry profitability in a mill region can be demonstrated. Research in this area required new tools and new skills for integration and analysis of information in order to evaluate alternative cane supply practices. However, it also drew heavily on the knowledge and understanding built up over the years for the individual components of the production chain through the activities of researchers in CRC-Sugar, other CRC Programs and other organisations.

Assessing opportunities for alternative cane supply arrangements for enhancing profitability was achieved through optimising the harvest date of sugarcane, accounting for geographical and crop differences in cane yield and CCS throughout the harvesting season. In terms of complexity, a mill region contains up to 10,000 paddocks on over 450 farms and harvested by up to 100 groups harvesting according to a fixed rotation, so that all farms are harvested proportionally over time. Harvest schedules using a fixed rotation do not allow the exploitation of geographical variation in CCS and cane yield at harvest date, meaning that productivity and revenue are sacrificed. Optimisation modelling has shown that by not using a fixed rotation and by exploiting geographical differences in productivity, significant gains in profitability are possible. Participatory action research has been a key component to realise the benefits of identified and agreed cane supply options for enhanced profitability.

Realising the profitability gains in the Australian sugar industry is a difficult challenge since:

- Individual farms are privately owned and farmers have for a long time been governed by regulated equity to ensure each farmer and harvester cuts a constant amount of cane throughout the harvest season. Socio-economic barriers for change from this system are strong in the Australian sugar industry and acceptance of change varies significantly across farmers and harvesting groups.
- Implementation of alternative cane supply options is not plug-and-play. In the past growers and harvesters have primarily adopted new technologies that involve a physical change (*e.g.* new harvesting equipment, new varieties, new pest control chemicals). Harvest schedules produced by the optimisation model are complex and need to be in a form that is easily interpreted by growers and harvester contractors.

- The concept of optimising more than one decision simultaneously is difficult to understand at first, even for scientists of non operations research disciplines, and optimising a harvest schedule at a farm or mill level is based on these principles. This would lead to growers, harvesters and millers being sceptical about the value of optimal cane supplies.
- Since the optimisation model uses statistical models to predict CCS and Tonnes of Cane per Hectare (TCH) response to harvest date, the optimal schedules produced by the model are not proof of concept themselves. Large-scale pilot implementation of the harvest schedules produced by the optimisation model is required to prove success.

Addressing these issues requires a three-fold approach with the ultimate aim of implementing cane supply options, leading to increased profitability across mill regions within the Australian sugar industry.

1. Strategic Research

- Develop cane supply optimisation model (Higgins, 1999; Higgins and Haynes, 2001) under CRC-Sugar Program 3, that captures the key inputs of CCS and TCH responses to harvest date and constraints associated with harvesting, transport and milling capacity.
- Develop strategies to overcome implementation barriers (Muchow *et al.*, 2000), pilot study schemes, software tools to assist implementation, and a process for monitoring and evaluation of pilot studies.

2. Case Studies

- Provide partnerships with the local industries in Mackay and Mossman to ensure the strategic research is carried out with continuous input from these local industries. Through options analysis groups in Mackay and Mossman involving growing, harvesting and mill representatives, necessary data was collated and models were ground-truthed and refined.
- Deliver options analysis that is relevant to the needs of the local industry with the ultimate aim of implementation.
- Provide proof-of-concept using pilot studies within the case study regions, which may lead to other mill regions within the Australian sugar industry adopting alternative cane supplies.

3. Wider Industry Delivery

- Assist other Australian mill regions in evaluating different cane supply and harvest scheduling options and implementation of these, given the outputs from strategic research and case studies.
- Develop the capacity for the widespread delivery of options for alternative cane supplies and the implementation of these.

OBJECTIVES

The primary objective of the project was to develop the capacity to conduct options analysis for alternative cane supply options across the Australian sugar industry. Other objectives were to deliver cane supply options to other mill regions through new partnerships, pilot alternative cane supply options in the Mackay region during 2001, improve the accuracy of the optimisation model at farm level, and develop a user-friendly cane yield estimation model for implementation within the industry.

The capacity to evaluate cane supply options in other mill regions was developed through constructing computer software, standardised databases and documentation to adapt the models to new mill regions. New partnerships were established with Maryborough, Tully and the Burdekin for the delivery of alternative cane supplies with Maryborough and the Burdekin proceeding to the implementation stage. Knowledge from previous studies was applied in Maryborough, which helped reduce the timeframe between start-up and implementation from four years down to one. Despite 2001 being a difficult year for growers in the Mackay region, a pilot study for alternative cane supplies was achieved and evaluated. The user-friendly cane yield estimation software was developed and is currently used by Mossman Agricultural Services, and is likely to be adapted for the Mackay region to improve satellite-derived estimates.

RESEARCH METHODOLOGY

Learning by experience and participation with the cane supply options analysis groups in each case study region underpinned the entire project. Embedded in the participatory research were action learning cycles to ensure that the project evolves towards beneficial change with full support of the industry with full review of progress. This approach was necessary since alternative cane supply options are new to the industry; optimisation is a difficult and sometimes a black-box concept, adoption will require the industry to change, and there are many potential barriers to implementation and difficulties with the use of software tools which could not be overcome without participatory research. This project (CTA044) linked knowledge and experience from the existing CRC-Sugar Sub-program 3.2 (Strategic Research) and SRDC MSA001 (Mackay Case Study Components). The well-established partnerships of CRC-Sugar Sub-program 3.2 were used for the Mackay and Mossman case studies and were more advanced in the delivery of cane supply options.

This project had three very different facets: (i) pilot implementation and wider industry delivery of alternative cane supply options; (ii) enhancing the modelling capability; and (iii) the development of software tools. Pilot implementation required the development of a strategy to overcome barriers to implementation, design a pilot scheme to provide proof of concept, and to ensure growers used the harvest schedules produced by the optimisation model. A statistical modeller, Dr Michele Haynes, was employed to design the pilot study scheme, including evaluation, to provide a statistical basis for establishing proof of concept. This was an integral component of developing a pathway towards pilot implementation, which was conducted using facilitated, creative solutions workshops with the milling, harvesting and growing sectors. In these workshops (held in late 1999 to mid-2000), two groups were formed to highlight issues (or potential barriers) associated with implementation, and to formulate plans to address them. The plans developed by the two groups were assessed and both groups collectively agreed on a final strategy for implementation during the 2000 harvest season.

The other key role of Michele was to develop models for predicting TCH and CCS, to improve the accuracy and credibility of the optimisation model at farm level, which was necessary for implementation. Prior to this project, the focus of the optimisation model was at industry level, exploiting CCS and TCH differences that occurred across productivity zones. As the implementation strategy was developed, individual farms and harvesting became the focus, and the optimisation model required major redevelopments such as predicting CCS and TCH across harvest dates using statistical methods, so that it could be accurately applied at farm level.

Developing software tools for industry required new specialist skills to be brought on board, with the appointment of Ms Samantha Peel. A major consideration for developing the software was future proofing against changes in data, operating systems and computer technology, considering the anticipated life of the software being between five and ten years. Prior to this project, the research tools were written in the Fortran 90 programming language, and the large range of data for the optimisation model was manually formulated for the model. Updating the input data, running the model or adapting the model to new mill regions, required time-consuming input from the scientist. The methodology used to develop the software tools was to firstly focus on the data management for the models, followed by the user-interfaces. Improved data management, which included developing data manuals, standardising data and the development of tools to clean and transform raw industry data into a form used by the model, removed the time-consuming input from scientists and transfers much of the implementation responsibility to industry. To ensure future proofing, MS Access was used to construct the databases for data management, since the sugar industry was familiar with MS Access and it is compatible with their existing/future database systems. The model interface and HarvSched were written in MS Visual Basic and the MS Access programming languages.

DETAILED RESULTS AND DISCUSSION

1. Computer Tool for Cane Yield Estimation

Accurate cane yield estimates (and re-estimates as the harvest season progresses) at paddock level are expensive for the mill and difficult to obtain, due to the large number of paddocks within a mill area. Given this issue, a mathematical model was developed in partnership with the Mossman Agricultural Services in 1998, which used historical block productivity data (Higgins *et al*., 1999). The model produced re-estimates at paddock and farm level that were up to 15 percent more accurate (*i.e.* up to 15 percent closer to the actual TCH) than those produced by the cane inspector technique. During 1998 and 1999, re-estimates were derived at the 6-week, 12-week and 15-week marks of the harvest season for the Mossman cane inspector. The main outcomes in the 1998 and 1999 trial years were the highlighting of well-consigned land and gaining the cane inspector's confidence.

Within this project, the cane yield estimation software was developed with guidance from the Mossman cane inspector. The interface was written in MS Visual Basic and the software can directly link with the field book and productivity databases on the Mossman Agricultural Services computer systems. This allows the cane inspector to produce re-estimates using the user interface, without the need of manual manipulation of data. The complete help manual of the software is contained in Appendix A. The software can provide the new estimates in several different formats, depending on the needs of the cane inspectors. The software was presented as a poster paper at ASSCT 2001 Conference (Peel and Higgins, 2001).

The Mossman cane inspector used the software at Mossman to derive re-estimates each week of the 2001 harvest season. After 5 weeks of the 2001 crush, the model produced a mill level re-estimate of 840,000 tonnes. The mill eventually cut 829,000 tonnes, much to the cane inspector's and Mill's satisfaction. They claim that the methodology is superior to the others used by the industry in 2001. The model also worked well at farm level and a comparison between the model re-estimate after 5 weeks versus the pre-season estimate is presented in Figure 1. Farms are sorted in order of industry estimate (pre-season). Model re-estimates after 5 weeks at farm level were an improvement over the pre-season estimate for about 70% of the farms. Since the cane inspector was satisfied with the software in 2001, it is very likely that it will continue to be used in 2002 and beyond. While it is difficult to put a dollar value on the outcome of using the software, we believe a satisfactory return from investment will result from its application.

The software will also be adapted to the Mackay region to enhance the accuracy of their satellite image methods, particularly in the case of a partial image resulting from cloud cover.

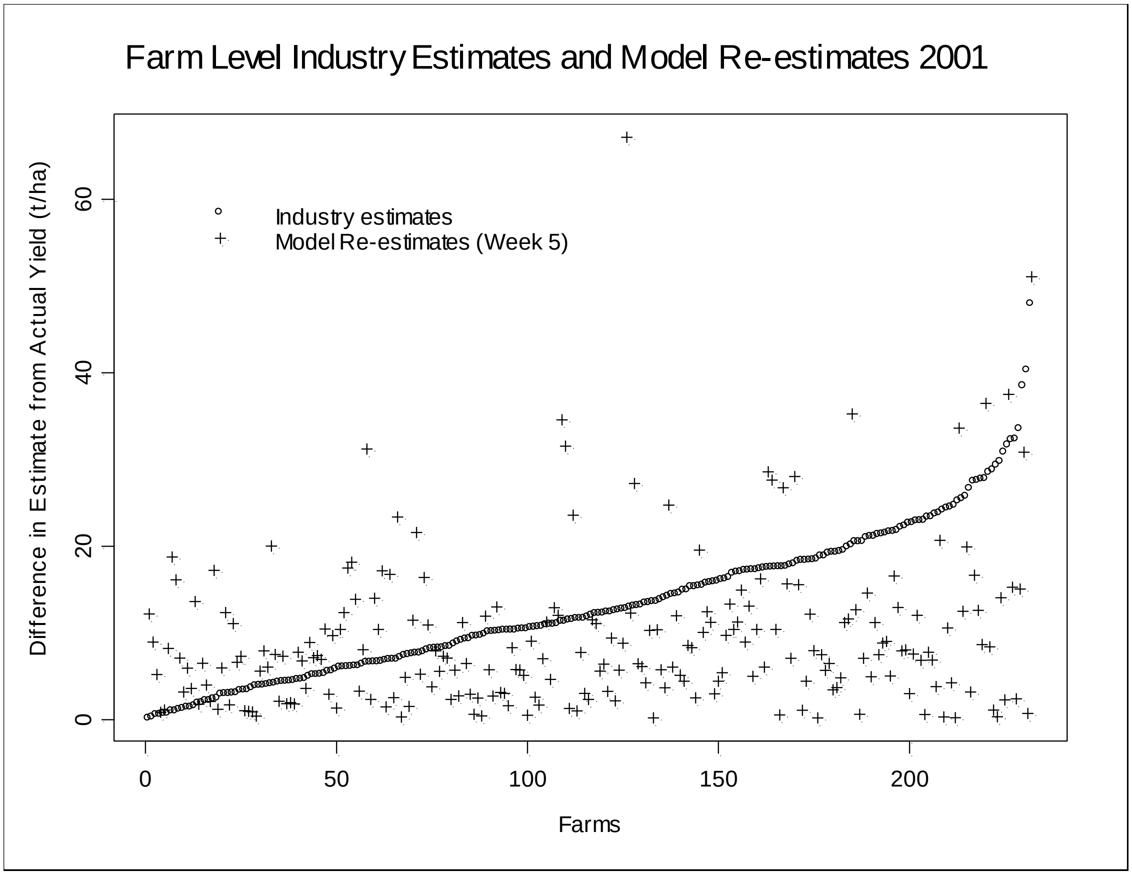


Figure 1. Accuracy of Farm level re-estimates versus pre-season estimates during the 2001 harvest season.

2. <u>Software Tools for Alternative Cane Supply Options</u>

Development of software for the cane supply options was the largest component of this project and required a full-time software development officer. The development of such software was essential because: (i) the research team did not have the capacity or desire to manually set up and run an optimisation model for every mill region in each new year; (ii) industry will eventually need to use the tools themselves; (iii) the tools need to be future proof; and (iv) implementation requires schedules produced by the model to be in a user-friendly and flexible format. The software need and the overall cane supply options project has evolved substantially since this project was formulated in late 1998 and the overall outputs/outcomes bears little resemblance to the original research plan. This evolution was mainly due to the development of pathways to implementation and learnings from the pilots. A flow diagram of the software as it stands at June 2002, is illustrated in Figure2.

In Figure 2, the cane supply options interface allows the user (who might be a person at Canegrowers or at a mill) to set up the optimisation model data from the raw industry data, and run the model for a specified cane supply option. DWIM is the central database that stores all data including the optimisation model input and output data, the CCS and TCH models for the optimisation model as well as parameters such as sugar price and milling/growing costs.

PortData reads in the raw industry data and transforms it into the standard format for DWIM. PortData may require merging block productivity data from two files such as in the case of rakes and paddocks data. Doing this manually is time consuming and requires knowledge of MS Access or programming. After the block and farm level data are imported, ValiData checks it for non-standard codes. The CSO model is the cane supply optimisation model written in Fortran 95. HarvSched is a large module in itself, written in MS Access, which is the main interface from the grower/harvester end (Section 2.3).

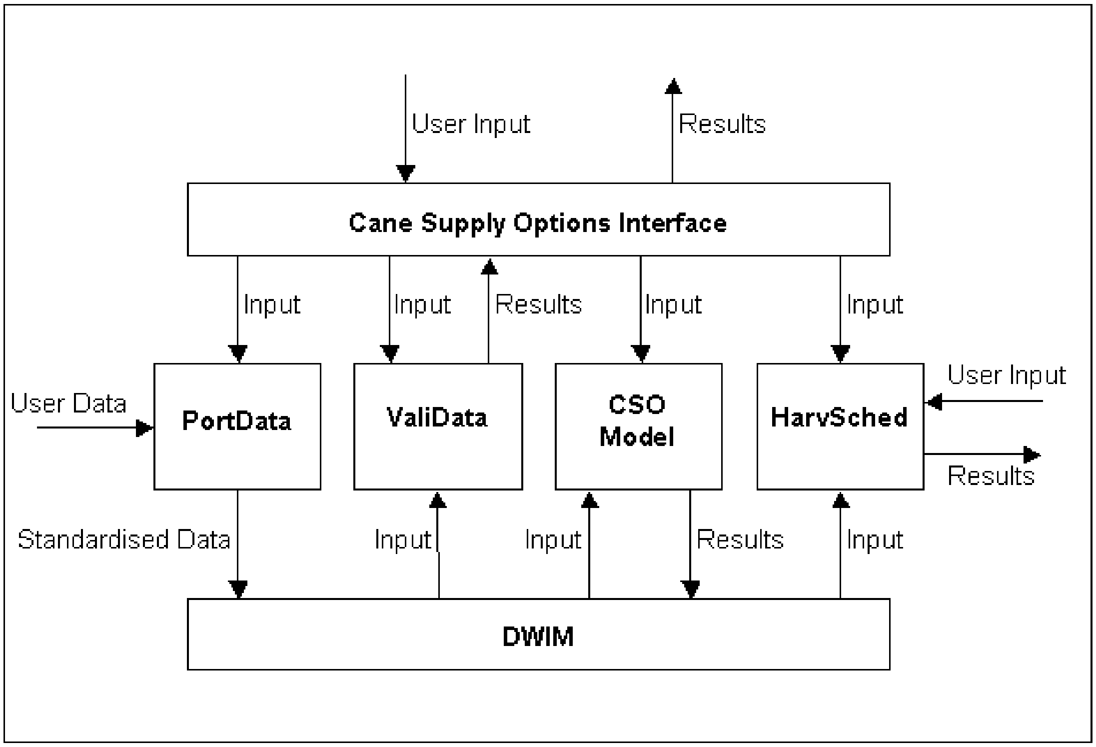


Figure 2. Cane supply options system.

In Figure 2, the cane supply options interface allows the user (who might be a person at Canegrowers or at a mill) to set up the optimisation model data from the raw industry data, and run the model for a specified cane supply option. DWIM is the central database that stores all data including the optimisation model input and output data, the CCS and TCH models for the optimisation model as well as parameters such as sugar price and milling/growing costs. PortData reads in the raw industry data and transforms it into the standard format for DWIM. PortData may require merging block productivity data from two files such as in the case of rakes and paddocks data. Doing this manually is time consuming and requires knowledge of MS Access or programming. After the block and farm level data are imported, ValiData checks it for non-standard codes. The CSO model is the cane supply optimisation model, which is written in Fortran 95. HarvSched is a large module in itself, written in MS Access, which is the main interface from the grower/harvester end (Section 2.3).

2.1. Standardised Data Formats for New Mill Regions

The first step in consolidating data for the cane supply optimisation model was to develop a document specifying all of the data required and the necessary format. Constructing the document helped the research team to carefully assess what data are really needed, and what data are optional, if the optimisation model was to be adapted to other mill regions, what was appropriate for the new regions. Some data are harder to obtain than others, *e.g.* data on wet paddocks is much more difficult to obtain than farm level and harvesting group data. However, data on wet paddocks is non-essential and this is noted in the manual. The manual for optimisation model data requirements and formats (see Appendix B) was used effectively to adapt the optimisation model to the Maryborough and Burdekin regions. In the case of Maryborough, all of the model data were ready within four weeks of one visit to Maryborough. The manual was also used for obtaining model data from the Tully region, though the project did not proceed in that region.

2.2. DWIM

Increasing the capacity to deliver cane supply options to the wider industry ultimately required an effective means of handling the huge quantities of the data (described in Appendix B), most of which needed to be updated every year. One could imagine in a modern society, the consequences of not having well-managed or well-maintained data, particularly in finance. The same applies to the alternative cane supply options research.

The overall system design of DWIM was critical, particularly as it required considering the industry needs and the other software components which link to it . DWIM needed to be designed for easy linkages to the optimisation model, the user interface and HarvSched. The well documented DWIM, online and MS Word (Appendix C), allow a non-expert to easily pick up the database and use the data. A separate DWIM is available for Mackay, Mossman and Maryborough, with the blank DWIM available for new mill regions. DWIM contains tables at block, farm group and mill level. It also stores the models for predicting CCS and TCH at harvest date. It also has functions to automatically calculate ratoon age of crop when the latest block productivity data (current year) are added to the database.

While DWIM has increased the efficiency and ease to develop options for alternative sugarcane supplies, other benefits include the availability of a comprehensive data set for further strategic research within the Australian sugar industry. This was evident in the development of more effective CCS and TCH responses to harvest date, for the optimisation model. Without DWIM, there would be almost 1,000 separate data files floating around on the system (at CSIRO) that have had some relevance to the cane supply options project. This could lead to problems in version control and a higher likelihood of flawed research results.

2.3. HarvSched

The need for HarvSched was not formulated until late 1999, six months after the start of this project. HarvSched was an underpinning tool for the pathway to implementation since it was the interface between the harvest schedules produced by the optimisation model and the growers/harvester contractors using them on the ground. It also produced summary graphs, such as the historical CCS trends of farms and varieties to provide back-up evidence that the model schedules were realistic. HarvSched is effectively an Intelligent Decision Support System since it allows growers/harvester contractors to fine-tune the schedule produced by the optimisation model to account for their on-farm constraints. That is, the grower/harvester contractor is making decisions based on an intelligent model and their own experience. A modelling and simulation conference paper (Haynes and Prestwidge, 2001a) giving more details about HarvSched is contained in Appendix D. HarvSched was also presented as a poster paper at the 2001 ASSCT (Haynes and Prestwidge, 2001b).

A more recent addition to HarvSched was linking with Geographical Information Systems (ArcView and MapInfo) to display harvest schedules using a farm map (see Appendix E). The development of the HarvSched package has increased motivation by the industry partners within the project and increased the likelihood of a grower/harvester contractor implementing the harvest schedules. Increased motivation was achieved through a better understanding of what the harvest schedule looks like, how it could work on individual farms, and some evidence (using historical data) that it will achieve benefits.

2.4. User Interface and Data Importing Routines

The user interface allows the non-expert to set up the optimisation model (*e.g.* set sugar price, cane supply option, run time) and to import the necessary data to run the model without having to manually manipulate data. The data importing routines can match rake and block data and transform harvest dates, varieties, crop classes and other block/farm data into a standard format for use in the model. Currently, a non-expert within CSIRO can now use the cane supply options system, though it has not yet been installed at any mill region. The Australian sugar industry's rapid advance in the use of WWW, GIS and other information technologies (*e.g.* the changes in Mackay Sugar from 1999 to 2002) will require careful consideration of how best to integrate the cane supply options system into the Australian sugar industry for longer term use. This will be addressed in the new SRDC project CSE003 "Adoption pathways for alternative cane supply options across the sugar industry". The current cane supply options computer system, with the user-interface, will be trialed for the 2002 harvest season implementation in Mackay, Mossman, Maryborough and Burdekin regions.

The user-interface and data importing routines were written in MS Visual Basic and views of these on-screen are contained in Appendix F.

3. Research to Improve Robustness of Optimisation Model at Farm Level and Across Season Types

Before the development of pathways to pilot implementation in 1999, the cane supply options research was focused at industry level options. For example, to determine the proportion of each productivity zone to cut at each harvest date, or what proportion of each crop class should be cut throughout the harvest season. Developing the pathways to implementation with industry partners in Mackay and Mossman highlighted the need for the cane supply options model and software to be relevant to grower level. The biggest inadequacy of the model at that stage was the way CCS and TCH responses to harvest date were calculated. This inadequacy was made clearer during the evaluation of the 2000 harvest season pilots (see MSA001 Final Report), since many growers commented that the schedule did not reflect the trends on their farm.

To overcome the problems of the 2000 harvest season, a statistical model for the CCS response to harvest date was developed. Modelling the CCS relative to the mill rather than actual CCS overcame the problem of accounting for year-to-year variability, since it allowed the effects of farm, varieties and crop classes to be more consistent across years. A description of the statistical modelling was applied as follows:

> Given paddock i and harvest week j, the estimated relative CCS for each farm paddock, rc^i_{jy}, was computed separately for each season, y $(y=1,\ldots,n_m)$, where n_m is the number of seasons represented in the block productivity data for mill region m ($n_m > 5$). A quartic polynomial model was fitted (using least squares regression) to paddock CCS, weighted by the tonnes of cane harvested, separately for each season so that the residuals, e^i_{jy}, represent paddock relative CCS computed using a smooth mill CCS trend that was

relevant to the season. Average variety and crop class were also removed from the e^i_{jy}, so that the rc^i_{jy} were computed from the analysis of variance model

$$e^i_{jy} = \mu + v_y + cl_y + (v.cl)_y + rc^i_{jy}$$

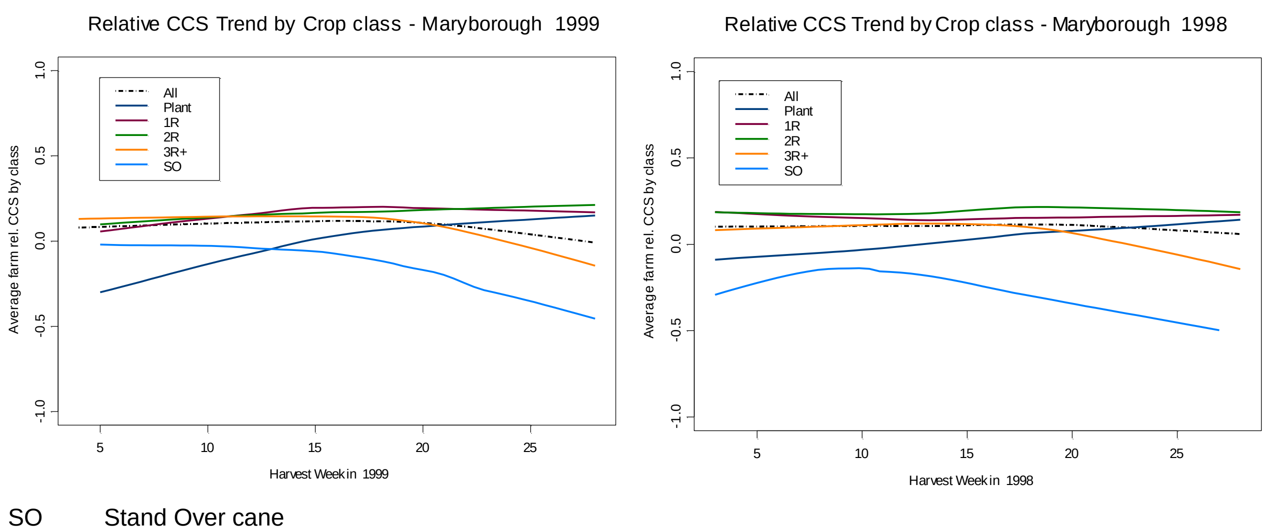
where the rcijy are the residuals of the model fitted to the eijy for paddock i harvested in week j, with effects for variety (vy), crop class (cly), their interactions during season y, and overall mean μ.

The linear model fitted to the adjusted relative CCS for all paddocks on a farm f is

$$rc^i_{jy} = \alpha_f + \beta_f * j + \eta^i_{jy}$$

where i=(1,…nf) and nf is the number of paddocks belonging to farm f , j=(1,…,w) is the harvest week where w is the length of the harvest season in weeks, af and bf are the estimated coefficients for the intercept, αf, and slope (or gradient), βf, of the linear model and the ηijy are the residuals of the model fit and are assumed to have a N(0,1) distribution.

Figures 3a and 3b illustrate the estimated relative CCS trends of different crop classes within the Maryborough region for 1999 and 1998 respectively. While the actual CCS curves changed from 1998 to 1999, the relative CCS trends stayed very consistent. This has significantly reduced the effects of year-to-year variability in CCS when generating options for alternative cane supplies, and resulted in increased confidence by industry stakeholders and grower participants for a beneficial implementation. The method was integrated into the optimisation model and applied to generate schedules for 2001 pilot implementation in Mackay, Mossman and Maryborough. The impacts were very positive in these regions with the accurate prediction of relative CCS throughout the 2001 harvest season (found through the end of year evaluation). Results for the 25 Mossman farms which were involved in the 2001 pilot implementation, showed the CCS model incorrectly predicted a “high early” or “high late” CCS for only three farms. More successful pilots were achieved in 2001 compared to the 2000 pilot scheme (see section 4). Full documentation of the model is presented in Appendix G and has been published in the paper by Higgins and Haynes (2001).



SO Stand Over cane

(a) (b)

Figure 3. Relative CCS of the crop classes at harvest date during the harvest season of 1999 (a) and 1998 (b).

4. Second Year of Pilot Implementation at Mackay

The inability to establish "proof of concept" in 2000, due to the very adverse season, led to the piloting scheme being repeated during 2001 and the remainder of the project milestones from March 2001 being modified to incorporate the repeated pilot scheme. The main advantages of piloting in 2001 versus the 2000 scheme were: (i) Maryborough region was now involved; (ii) the many learnings of the year 2000 pilots were incorporated to improve the process for 2001; and (iii) the CCS and TCH response to harvest date inputs to the optimisation model were improved to account for variety and better account for geographical location. Also, farm specific summaries of historical block productivity data were made available to individual growers to increase their confidence in the modelling process and hence increase their participation. As a result, about 15% of the Maryborough region and 29 farms from the Mackay region were committed to the 2001 pilot scheme.

As with the 2000 harvest season, three key options were to be piloted.

Option 1: Optimal harvesting across harvesting groups. This is where the largest potential gains in profitability are, since it involves the removal of both farm and harvesting group equity. In the 2000 harvest season, this option required a Harvesting Entitlement Exchange Scheme (HEES) to facilitate its implementation. While the scheme successfully allowed equity to be overcome for implementation, it increased the harvesting and transport complexity. As a consequence, the Mackay Cane Supply Options Analysis Group agreed to increase the emphasis on Option 2 in 2001.

Option 2: Optimal harvesting across farms within a harvesting group. Only farm equity is removed and farms are harvested when it is more beneficial in terms of their CCS relative to the mill. For example, if a harvesting group has farms that have a high early CCS relative to the mill, a larger proportion of these farms would be cut earlier, while maintaining harvesting group equity.

Option 3: Optimisation of harvest date of paddocks at farm level, accounting for differences in variety, crop class and age at harvest date. Both farm and group equity is maintained.

A further option was in demand in 2001 as follows:

Option 2/3: Optimal harvesting across farms for a grower manager. This option can apply when a grower has farms in contrasting high early versus high late areas, but others within the harvesting group do not wish to participate. In Mackay during 2001, the option 2/3 was implemented similarly to an option 3 since an individual paddock schedule was developed for all farms owned by the grower manager.

After the 2001 crush, evaluations were conducted for Mackay, Mossman and Maryborough. Statistically, proof of concept was not achieved for any of these three regions. The main reasons in the Mackay region were recovery from the 2000 season, substantial changes in field officer numbers and services to growers from Mackay Sugar in 2001, changes in varieties; insufficient numbers of growers committed to following the harvest schedules throughout the whole season; and changes in mill planning arrangements (*e.g.* establishment of a central communications centre). An evaluation workshop was held with the Mackay growers (who piloted in 2001) on 27 November 2001.

As part of the evaluation, an analysis was done to assess the gains in CCS (relative to the district) for the farms involved in the pilot scheme. Figure 4 contains a full description of the participating growers along with their compliance and gains in CCS. In Figure 4, the "Comp with Adjusted" field, is the compliance factor of the schedule used by the grower versus the schedule produced by the model. On many occasions, the grower changed the schedule produced by the model so much that it no longer had any resemblance with the optimal, thus a low compliance. The "Comp with model" field is the compliance between the actual harvest order on the farm versus the original optimised schedule produced by the model.

There were a total of six farms involved in Option 1 (harvesting across groups) for which farm 2552 had no historical data that matched with the 2001 data. There were 23 farms involved in either Option 2/3 (harvesting across farms within a group) or Option 3 (optimal harvesting of farm paddocks within a farm), for which an evaluation was conducted. Farms were not included in the evaluation if they had too many mismatches within the historical productivity data (prior to 2001) or if schedules were not finalised with the grower. Figure 5 shows a summary of the gain in "CCS relative to the mill" for 2001 versus the average for 1995 to 2000 for each option 2/3 and 3 farm in the pilot scheme. Since options 2/3 and 3 are not prescriptive, they are sorted in order of how the actual schedule complied with the schedule produced by the optimisation model. It was done this way to show the impact of compliance on the gain in CCS.

MACKAY CANE SUPPLY OPTIONS - GROWER PARTICIPATION LIST 2001

A schedule was prepared for the following growers -
(Note: only completed schedules have been entered in the GrowerData file)

								FINAL COMPLIANCE BASED ON RAKE DATA			
SURNAME	FIRST NAMES	FARM1	FARM2	FARM3	FARM4	TYPE	Option	Comp with Adjusted	Comp with Model	CCS gain A	CCS gain B
Adams	David	4187-1	4187-2			Management	2+3	61.1	28.7	0.08,-0.03	-0.09,0.07
Bonaventura	Lawrence	1016	1019	1022		Management	2+3	79.8	58.8	0.11, -0.38, -0.21	-0.33,-0.24,-0.34
Comelli	Gary & Wayne	1297-2	1293			Management	2+3	Too many mismatches			
Cowan	Ian & Glen	1282-1	1282-2			Management	2+3	64.3	52.2	-0.13,-0.15	-0.01,-0.79
Deguara	Andrew	1210-1	1210-2	1206	1738	Management	2+3	Too many mismatches			
Deguara	John	3074	3082			Management	2+3	58.9	44.2	-0.31, -0.25	-0.64,-0.74
Gordon	Sid & Sid	1234	1242-1	1246-1	1468-1	Management	2+3	Scheds never finalised			
Grech	Jeff	4055	4059	4244		Management	2+3	44.6	35.7	0.074,0.12,-0.23	-0.07,-0.07,-0.7
Hamilton	Lex	4085				Single farm	3	56.5	37.8	0.40	0.28
Lowrey (Graff	Ron	4180				Single farm	3	28.7	27.9	unknown	unknown
McMahon	Tony	4303	4304			Management	2+3	50.0	65.0	-0.81,-1.43	-0.76,-1.7
Millard	George, Bill, David	1155	1156			Management	2+3	65.0	50.8	-0.27	-0.47
Millard	George, Bill, David	3428				Single farm	3	Blocks numbers mismatch			
Powell	John	1178				Single farm	3	91.0	91.0	0.23	0.12
Pratt	Graham	1235	1238-1	1239	1254	Management	2+3	Scheds never finalised			
Sheehan	Mick	2066				Single farm	3	63.4	38.5	-0.46	-0.79
Vassallo	Emanual, Julian, A	3065				Management	HEES + 2+3	48.0	56.6	-0.31	-0.11
Vassallo	Emanual, Julian, Andrew		3071				HEES + 2+3	48.1	31.1	-0.03	-0.24
Volker	Albert & Gary	2078	2152			Management	2+3	Scheds never finalised			
Westcott	Warwick & Rowan	4155-1	4155-2			Management	2+3	49.9	41.6	0.32,0.62	0.27,0.19

Option 2 using HEES

SURNAME	FIRST NAMES	FARM1	FARM2	FARM3	FARM4	TYPE	Option	CCS gain A	CCS gain B
Argent	Dan and Paul	2445	2449			Management	1	0.42,0.47	0.25,0.04
Galea	Richard	2474	2448	3281		Management	1	-0.06, 0.42, 0.19	-0.13,1.44,0.47
Peoples	Jim	2552				Management	1	No historical data	No historical data

CCS gain A - Gain in CCS (relative to mill) versus average of 1995 to 2000
CCS gain B - Gain in CCS (relative to mill) versus 2000

Figure 4. Summary of participating growers for the 2001 harvest season.

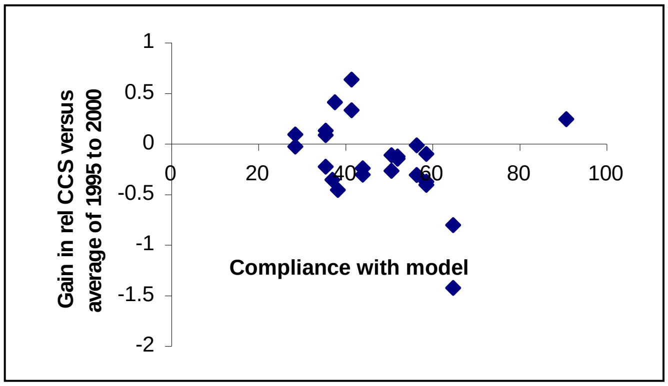


Figure 5. Gains in CCS for Mackay farms involved in Options 2 and 3 versus the compliance with the schedule.

In Figure 4, the results for Option 1 were promising compared to those for Options 2/3 and 3. The results for Option 1 were acknowledged by the Mackay Cane Supply Options Analysis Group, who agreed that the largest gains for industry were from Option 1. Out of the Option 2/3 and 3 pilots, there was only one farm (1178) with a very significant level of compliance (91 percent compliance and with a 0.23 units of CCS gain A). The low compliance also prevented any exploitation of geographical differences in CCS across harvest date (as in a pure option 1 or 2).

The growers who attended the evaluation workshop were asked to form groups of five or six people to discuss and provide responses to pre-determined questions. The purpose of this task was to obtain qualitative information to improve implementation in the future. Key messages that surfaced from the exercise was to need to account for new varieties and for growers to access schedules via the World Wide Web. The Mackay Cane Supply Options Analysis Group (MCSOAG) met immediately prior to the evaluation workshop with the purpose of providing a common understanding of the outcome from 2001 and to decide on future activities. The group agreed to continue the work in the future but to focus on linking the tools to the World Wide Web for growers and harvesters to use themselves in the absence of a designated project officer. The conclusion of the 2001 pilot scheme also marked the end of Mr Gerry Turner's term and role in the project. The MCSOAG acknowledged Gerry's valuable contribution to the project.

An evaluation workshop was held at Mossman on 20 February 2002. The evaluation was conducted for 13 growers (23 farms) who participated in 2001. While the evaluation did not statistically show that the alternative cane supply arrangements resulted in a significant CCS gain, most of the growers found the model schedules useful and expressed a wish to continue in 2002.

4.1 Issue of compliance to the option 3 schedules

At the time of the evaluation workshops and as a result of these workshops, the low compliance for options 2/3 and 3 pilots were attributed to: 1) extra ploughout; 2) harvester contractors not complying to the schedule, even when the grower asks them to; 3) complicated schedules to follow; and 4) possibly the non-prescriptive nature of Options 2/3 and 3 after the start of the harvest season. The Harvesting Entitlement Exchange Scheme for Option 1 locks the harvesting group into a schedule for the whole harvest season. Growers can withdraw from an Option 2/3 or 3 anytime during the harvest season, particularly when a difficulty is encountered.

Looking back at the life of the cane supply options projects has revealed further underlying reasons for the non-compliance of option 3 (and option 2/3, which were effectively an option 3 in practice). While results are not available to back up the claim, compliance is a much bigger issue with option 3, rather than option 1 and the pure option 2. Option 3 is a long list of "order of farm paddocks" requiring little change from a grower, rather than just following the order. Options 1 and 2 are a schedule simply stating how much cane should be cut from each farm throughout the harvest season, but requiring change to equity to achieve it, unless the grower or harvester contractor owns the land involved. Once there was a commitment to options 1 and 2, the schedule was straightforward to produce and implement (once the transport logistics were addressed). With option 3, there are many events that can cause the grower to divert from the schedule, including inaccuracy of consignment of block data, wet paddocks, new varieties and ploughout.

A potentially underlying reason why option 3 schedules were not followed closely in all case study regions is with the use of block productivity data to optimise the sequence of farm paddocks. This is a more recent learning and only applies with option 3. The statistical modelling applied to the block productivity data is used to capture the CCS trends for farms but also for variety by crop class (Higgins and Haynes, 2001). The effects for a farm are calculated using the data for that farm only, thus the estimated statistical model for the farm is reliable. This means there are no problems with schedules produced for options 1 and 2. Option 3 required the CCS effects, by harvest date, be estimated for individual paddocks. Since there is not enough historical data to estimate these effects for each paddock, we had to estimate the variety by crop class by harvester date effects using all paddocks within the mill region. This means that if paddock X is Q136, 1st Ratoon to be harvested in week 8, the estimate of CCS is captured using historical productivity data for all paddocks of Q136, 1st ratoon at week 8. Unfortunately variety, crop class and harvest date only explain about 50% of the CCS variation across farm paddocks. When optimising the harvest date of farm paddocks within a farm, the optimal sequence of farm paddocks within a farm is unlikely to be accurate because of this low explanation of CCS variation. For farms with paddock that are average in terms of CCS effects from variety and crop class, option 3 schedules are much more likely to be accurate. This appeared to be the case for the growers who had good compliance and achieved gains. Prior to the 2001 harvest season in Mackay, a large number of option 3 schedules were produced for growers and handed out to them. At the end of the season less than 10 percent of the farms had an 80% compliance. The main reason for growers not using the option 3 schedules at all or withdrawing early in the season, is the schedule being very different to the grower perception of the farm; thus those who do follow the schedule are the ones which the optimal schedule is accurate for their farm.

4.2 Other major lessons learned

Other shortcomings of option 3 that were learned over the life of the project are the huge resources required to produce an optimal sequence of farm paddocks in a format that growers are satisfied with. An interactive software tool (HarvSched, see Appendix E) was developed for Mr Gerry Turner (and project officers in other mill regions), to formulate schedules with the growers. The desirability of option 3 by growers and resistance to change from equity, move the emphasis from option 1 (2000 harvest season) to option 3 (2001 harvest season). At that time, however, the shortcomings of option 3 were not known.

Within the first three years of the cane supply option project within the CRC for Sustainable Sugar Production (1996 to 1999), the focus was to build a model to optimise the harvest date of farm paddocks. This model was to be used to assess various options for exploiting the geographical and crop differences in CCS and TCH across harvest dates, at a whole of mill level. A focus was to achieve options with a high level of accuracy. This meant that harvesting, transport and milling capacity needed to be modelled very accurately, and lead to a highly credible model for producing the options analysis. Once the project team and industry collectively realised that the alternative cane supply options were to be piloted by a few growers and harvesting groups, rather than a full uptake, the model had some major shortcomings for this purpose that were not known at the time and took two more years to fully understand. The shortcomings were as follows:

- The cane supply options produced by the optimisation model (particularly option 1) assumed the whole mill participated, which is the case when assessing potential gains over the current system. When only a subset of the mill participates, you need to only optimise across those harvesting groups.
- The CCS and TCH response to harvest date was based on productivity zone averages, up until the end of the 2000 harvest season pilots. This meant that a whole productivity zone (containing up to 50 farms) was categorised early maturing, follow the mill, or late maturing. Once the project team used GIS technologies in early 2001, it was realised that district averages were not valid and there was a huge geographical variation across farms within a zone. The CCS and cane yield responses, however, were improved for the 2001 harvest season (Section 3).
- For the purposes of implementation where there is not a full uptake within a mill region, there is no need for explicit information on transport capacity, wet paddocks, grub damage paddocks, crop cycle length, growing and milling costs. All of this information is subject to error and artificially constrains any schedule for the harvesting group or grower, who usually fine-tunes any optimal schedule to account for their individual on-farm practicalities. We learned that it is better to provide a grand optimal (with few constraints) and allow flexibility to fine-tune for harvesting group and farm practicalities.
- Following on from the previous dot point, it is very difficult and time-consuming for the project team and the local industry to update the model for each harvest season with information on transport capacity, milling and growing costs, inside and outside stops, wet paddocks, grub damage, and even pre-season block conditions/estimates. A focus for the 2002 and 2003 harvest seasons is to reduce this complexity drastically.
- For the implementation during 2000 and 2001 at Mackay, Mr Gerry Turner role was to organise and facilitate the pilot studies. Gerry effectively took on the role of a local champion. However, limited resources meant that this could not be an on-going arrangement, and the departure of Gerry meant the departure of the local champion. After the departure of Gerry in December 2001, it was much more difficult to keep the momentum in Mackay compared to Maryborough.

Nobody is at fault for any of the above learnings since every aspect of the cane supply options project was new for the project team and industry partners. However, the learnings are valuable for the follow-on project CSE003 and some future whole-of-system projects.

5. Wider Industry Delivery of Alternative Cane Supply Options

As a result of the broad-based industry workshop held at Townsville in March 1999 and attended by 140 industry representatives, several mill regions showed strong interest in evaluating alternative cane supply options for their regions. The first of these was Maryborough and a district level workshop was held May 1999 for which a commitment to move forward was established. During the district level workshop an emphasis was placed on the process and commitment required to deliver credible alternative cane supply options. This required collation of data, modelling of industry constraints using grower expert knowledge, adapting the optimisation model, ground-truthing preliminary model results, and communicating to the wider region. A small SRDC project (MSF001) helped support this process in Maryborough. The time between start-up and implementation was about one year, depending on the level of commitment from the mill region.

District level workshops were also held with Tully and the Burdekin in late 1999 and commitments similar to that of Maryborough were made. A district level workshop was held with the Tableland mill in late 2000. Key advantages of the Tableland mill region include: (i) the mill owns the harvesters and there are only four harvesters; (ii) the growers are very open to Option 1; and (iii) there is a much greater likelihood of a revolutionary uptake (*i.e.* the whole mill region implementing Option1). While the district level workshop highlighted that the Tableland is an ideal mill region for the project, no commitment was made to move forward.

A major highlight in wider industry delivery, was the rapid progress with the Maryborough region and their decision to pilot cane supply options during the 2001 harvest season. Using the tools developed in this project, the time required between startup and delivery of options analysis for alternative cane supplies was reduced from three years (in the case of Mackay) to less than one year. This also led to an earlier than expected implementation in Maryborough. The Database for Whole of Industry Models (DWIM) and HarvSched developed in this project provided the capacity needed for a large-scale pilot implementation in the 2001 harvest season. A workshop was held in Maryborough on 10 May 2001 to (a) develop a pilot implementation strategy, (b) select whole harvesting groups to invite for participation, and (c) organise a grower workshop to provide the growers and harvester contractors with the draft schedules and information. At this meeting, Mr Frank Sestak (Maryborough CPPB) was identified to take a lead role in facilitating the implementation and supporting the pilots within the Maryborough region, similar to the role of Mr Gerry Turner in Mackay. Twelve of the invited growers attended the grower workshop held on 13 June 2001, most of whom were keen to participate. Three large harvesting groups are participating in option 2, including the Maryborough Sugar Factory, which when combined, is a total of about 15% of the Maryborough 2001 estimate.

While a partnership was established with the Burdekin in late 1999, progress was slower than with Maryborough due to mixed feelings towards the project within the Burdekin Cane Supply Options Analysis Group and the increased difficulty of obtaining data. A cane supply options analysis workshop was held with the Burdekin Industry on 24 November 2000 to present and obtain feedback for cane supply options analysis results, address some implementation barrier issues and make a decision on pilot implementation in harvest 2001 or 2002. The project in the Burdekin did not proceed any further until late 2001. A Burdekin Cane Supply Options Analysis meeting was held on 16 October 2001 for which the key agenda item was for the group to decide whether to proceed to implementation in 2002 or terminate the project in the Burdekin. The group agreed to implementation in 2002 and a working group was formed to facilitate the process to implementation. This group met on 5 December 2001 and a path forward was established for implementation in the Pioneer and Invicta mills during the 2002 harvest season. A follow-up meeting for the steering group was held on 18 April 2002 to update progress for implementation during 2002. The project in the Burdekin is currently on track for about 15 growers to trial on-farm scheduling (option 3) and scheduling across farms within a group (option 2).

Through an invitation, a presentation was given to 80 growers in the Mulgrave region on 27 March 2001. Nearly all growers present agreed the work is highly relevant to the Mulgrave region, and wish to explore alternative cane supply options. The milling sector is also strongly supportive and a partnership may be formed in 2001. A meeting was held with Plane Creek growers on 29 May 2002 and a proposal was submitted to the region in early July. Major interest (July 2002) has also come from the Tully and Tableland mill regions with the latter keen for an option 1. The Tableland region will be followed up as part of CSE003. The Herbert region has also expressed interest, particularly from CSR and a district level workshop may be held in 2002.

ANALYSIS OF OUTPUTS AND OUTCOMES COMPARED TO OBJECTIVES

Project Objectives

This section lists project objectives, along with key outputs and outcomes associated with each.

Objective 1: Expand the capability of the Australian sugar industry to evaluate different cane supply and harvest scheduling options to enhance profitability

Output

A cane supply options system that can be used by a non-expert and is adaptable to mill regions across the Australian sugar industry.

Outcomes

The length of time from project start-up to implementation reduced from four years down to about one year.

Alternative cane supply options delivered to other sugar mill regions, Maryborough and Burdekin.

Objective 2**:** Develop "user-friendly" versions of tested optimisation models

Outputs

A "user-friendly" cane yield estimation system written in Visual Basic and Fortran and links with the Mill's databases.

An Intelligent Decision Support System "HarvSched", which interactively combines expert knowledge from growers with an optimal harvest schedules from the model, to produce an easily interpreted schedule that meets unquantifiable constraints.

A user-interface for the optimisation model that allows a non-expert to import data into DWIM, export the data from DWIM to the optimisation model, and set model parameters such as sugar price and the cane supply option.

Outcomes

Adoption of the cane yield estimation system by the Mossman cane inspector and used during the 2001 harvest season.

High level of satisfaction by the Mossman cane inspector and a commitment to use the software in 2002 and beyond.

Strong confidence in the cane yield estimation system with an accurate mill level re-estimate after 5 weeks of crush being 840,000 tonnes compared to the actual mill crush of 829,000 tonnes.

Capacity to deliver and implement alternative cane supply options to other regions on top of the Mackay and Mossman case study regions.

Increased implementation of alternative cane supplies by growers through a better understanding of how the schedules produced by HarvSched can add value to their farm.

Objective 3. Work with industry at district level to collate input data required for the models, and facilitate the use and adoption of models to evaluate alternative cane supply arrangements

Outputs

A Database for Whole-of-Industry Models (DWIM) that stores all of the model data in a central place and in a standardised format.

Established cane supply options analysis groups in the Burdekin and Maryborough regions, which include high-level representatives from the milling companies, Canegrowers and the harvesting sector.

Outcome

Cane supply options produced in partnership with the options analysis groups leading to a decision towards implementation.

Objective 4**.** Build the skills base in whole system analysis using mathematics and information technology linked to participatory approaches for R&D

Outputs

Improved modelling of CCS and TCH response to harvest date, which are the core drivers of options analysis produced by the cane supply optimisation model.

Additional Outcome

Pilot implementation of cane supply options within the Mackay, Maryborough and Mossman regions was undertaken during the 2000-2002 harvest seasons. The scope of the original milestones in the project was the delivery of cane supply options showing gains in profitability and a description of what the alternative cane supplies look like, compared to current farm and harvesting equity.

POTENTIAL BENEFITS AND LIKELY IMPACT TO THE AUSTRALIAN SUGAR INDUSTRY

The potential gains in profitability have been assessed for Mackay, Mossman, Maryborough and the Burdekin (Higgins and Muchow, 2002). Considering option 1 only, the average potential gains in profitability per year for Mackay, Mossman, Maryborough and the Burdekin are $119/ha (81,000 ha of harvested cane), $79/ha (13,000 ha), $77/ha (11,000 ha) and $235/ha (67,000 ha) respectively. This gives a total of $27M for these mills, over a total harvested land of 172,000 hectares. Given the Australian sugar industry has 400,000 hectares, the potential from option 1 could be approximated as $27M*400000/172000=$63M per year. Further gains are possible from optimising the harvest date of farm paddocks at farm level, exploiting the differences in varieties and crop classes at harvest date.

In the 2001 harvest season, pilot implementation of alternative cane supplies took place in Mackay, Mossman and Maryborough with the Burdekin commencing in 2002. In Mossman and Maryborough, the uptake was about 5% and 15% respectively and all mills will continue implementation in 2002. The uptake by Maryborough and Mossman is likely to increase in 2002 compared to 2001. Interest has also been shown by the Tableland, Mulgrave, Tully and the Herbert to explore options for alternative cane supplies. This further interest was partially a result of this project, since the capacity has been increased to deliver to other mill regions.

The overall likely impact to Australian sugar industry depends on the level of uptake in each mill region and the time it will take to achieve maximum uptake. Through the evaluation of the pilots of the 2000 and 2001 harvest seasons, issues were raised that need to be addressed to allow a full mill region uptake. These are listed in the recommendations.

PROJECT TECHNOLOGY

The following end-user software was produced within this project:

1. "HarvSched"
2. Cane Supply Options system (which can also include HarvSched)
3. Cane Yield Estimation System.

The software does use commercially sensitive block productivity data, which cannot be made available to those who do not have authorised access to the data.

RECOMMENDATIONS

This project, and its linkages with CRC-Sugar and MSA001, has provided major groundbreaking outputs in the development of optimisation models and delivery of whole-of-industry options for alternative cane supply arrangements. Through the participatory research, the project had major achievements in developing a process for implementation, and piloting the cane supply options with several harvesting groups and growers during the 2000 and 2001 harvest seasons. These accomplishments, which are beyond the objectives of the project, have not been achieved previously in a whole-of-industry project within the Australian sugar industry.

Despite the tremendous progress made with this project, a number of issues arose that need to be addressed for the industry to realise the full benefits from alternative cane supply arrangements. Issues raised by industry and researchers during the life of the project, which the project team recommends be addressed:

- Implementation of alternative cane supplies worked when there was only one decision maker involved. For example: on-farm scheduling; harvesting across farms when the grower owns all of the farms; or when the grower harvested their own cane (*e.g.* Maryborough Sugar Factory cane). The biggest benefits are when there is scheduling across farms (or groups), which requires co-operation of multiple decision makers (*e.g.* two harvesting groups or several growers). Despite a huge amount of effort being applied to the implementation of cane supply options involving multiple decision makers, it was unsuccessful. The questions of facilitating change in this area need to be addressed to realise whole-of-mill implementation of some whole-of-industry research.
- Alternative cane supply arrangements have co-dependencies with decisions made on irrigation, pest management, ripeners, variety selection, best-practice harvesting and trash management.
- Implementing and realising the full potential benefits of alternative cane supply arrangements may increase the costs and complexity of harvesting and transport. As a consequence, many growers, including those who had significant potential benefits, decided not to pilot alternative cane supply arrangements in the 2000 and 2001 harvest seasons. There are two potential avenues to address this. Firstly, develop institutions and incentives for harvesters and the mill to uptake alternative cane supply arrangements to ease the potential increased costs and complexity. Secondly, deliver options for integrated cane supply, harvest and

transport systems, providing maximum net benefit for the combined system, and potentially leading to a win-win situation across all sectors.

- The research depends heavily on the availability of quality historical block productivity data. Necessary cost cutting by some mills in 2001 means that the future data may not be less accurate compared to 1995-2000 data. Secondly, a new industry Privacy Act will make block productivity data less accessible to researchers due to the need to obtain permission from growers. The importance and potential impact of having quality data (versus cost of maintaining the data) needs to be understood throughout the industry.

In light of the learnings from Section 4, the following recommendations are made:

- Make the learnings available for future projects.
- Within existing regions that use option 3, no longer use schedules that show the harvest date (produced by the model) for every farm paddock. Instead produce simple/generic guidelines that show the preference order of variety by crop class across harvest dates. This will be addressed as part of CSE003.
- Not to consider option 3 in new mill regions (*e.g.* Plane creek, Tableland).

ACKNOWLEDGEMENTS

A large number of people participated in this project, particularly in the sugar regions of Mackay, Mossman, Maryborough and the Burdekin. Firstly, we would like to acknowledge members of the Mackay, Maryborough and Burdekin Cane Supply Options Analysis groups for driving the projects at the local level throughout the life of the project. In achieving the goals of the project, we particularly acknowledge: Dr Andrew Wood (CSR), Mr Frank Sestak (Maryborough CPPB), Mr Peter Downs and Mr Glyn Peatey (Maryborough Sugar Factory), Mr Dave Langham and Mr Drew McGilchrist (Mackay Sugar), Mr Gerry Turner, Ms Jennifer Hollis, Dr Lisa McDonald (CRC-Sugar) and Mr Alec Ford (Mossman Agricultural Services). We would like to thank Mr Alan Stafford of Mossman Agricultural Services who worked closely with the research team to develop the Cane Yield Estimation System and Ms Angela Murray of CSIRO Sustainable Ecosystems for her help in linking HarvSched to the GIS software. Other staff from the Coastal Cropping Systems group within CSIRO Sustainable Ecosystems also contributed to the success of the project: Mr Steve Attard for helping to facilitate the work in the Burdekin; Ms Clare Gambley for communication activities (magazine articles, fact sheets, posters etc) and Ms Barbara Philp for managing communication with industry collaborators and planning for meetings/workshops.

PUBLICATIONS RESULTING FROM THE PROJECT

1. HAYNES, M.A. and PRESTWIDGE, D.B. (2001a). Decision support framework for generating improved sugarcane harvest schedules. *Proceedings 2001 International Congress on Modelling and Simulation* pp.1659-1664.
2. HAYNES, M.A. and PRESTWIDGE, D.B. (2001b). A database system to guide implementation of alternative cane supply arrangements. *Proceedings Australian Society of Sugar Cane Technologists* 23: 467.
3. HIGGINS, A.J. (1999). Optimising cane supply decisions within a sugar mill region. *Journal of Scheduling* **2**: 229-244.
4. HIGGINS, A.J. and HAYNES, M.A. (2001). An integrated modeling approach to enhancing sugarcane profitability. *Proceedings 2001 International Congress on Modelling and Simulation* pp.1787-1792.
5. HIGGINS, A.J., HAYNES, M.A. and MUCHOW, R.C. (2000). Preparation of block productivity data for use in an optimisation model to improve can supply schedules. *Proceedings 1999 CRC-Sugar Productivity Data Workshop* (in press).
6. HIGGINS, A.J. and MUCHOW, R.C. (2002). Assessing the potential benefits of alternative cane supply arrangements in the Australian sugar industry. *Agricultural Systems* (In press).
7. HIGGINS, A.J. and PEEL, S.L. (2001). A Windows program that improves cane yield estimates of farm paddocks. *Proceedings Australian Society of Sugar Cane Technologists* **23**: 464-465.
8. HIGGINS, A.J., STAFFORD, A., MUCHOW, R.C. and RUDD, A.V. (1999). Improving cane yield estimates for farm paddocks. *Proceedings Australian Society of Sugarcane Technologists* **21**: 44-50.
9. MUCHOW, R.C., HIGGINS, A.J., ANDREW, W.J. and HAYNES, M.A. (2000). Towards improved harvest management using a whole industry systems approach. *Proceedings Australian Society of Sugar Cane Technologists* **22**: 30-37.
10. MUCHOW, R.C., HIGGINS, A.J., INMAN-BAMBER, N.G. and THORBURN, P.J. (2001). Towards sustainable sugarcane production using a whole systems analysis approach. *Proceedings 2001 International Symposium of Sugarcane Technologists* 24: 91 – 94.

LIST OF APPENDICES

APPENDIX A.

Cane Yield Estimation Help

Samantha Peel
CSIRO Sustainable Ecosystems
and
Cooperative Research Centre for Sustainable Sugar Production

Long Pocket Laboratories
120 Meiers Road
Indooroopilly
Brisbane 4068

Introduction

These instructions have been written for version 1.6 of the Cane Yield Estimation program.

System Requirements

This program requires the following hardware and software to work.

CPU:	Pentium (minimum)
Operating System:	Microsoft Windows 95, 98 or NT 4 – service pack 4
Database:	Microsoft Access 97 with service release 2
Other:	Cane Supply Options version 0.1.3

Data Requirements

The program requires the following data to be able to produce reasonable cane yield estimations. For more details see Appendix A.

- Several recent years of historical productivity data at paddock (sub-block) level,
- the current season's conditions for each paddock,
- pre-season estimates for the current season and at least one previous season, at paddock level, and
- field book data at paddock level (only for mid-season estimates).

The user will be required to import their data into a specifically designed Microsoft Access database, called DWIM. The Cane Supply Options program contains wizards that will help the user in this task. Once the required data has been imported, the Cane Yield Estimation program can be used to calculate pre-season or mid-season estimates.

Starting the Cane Supply Options Program

Briefly

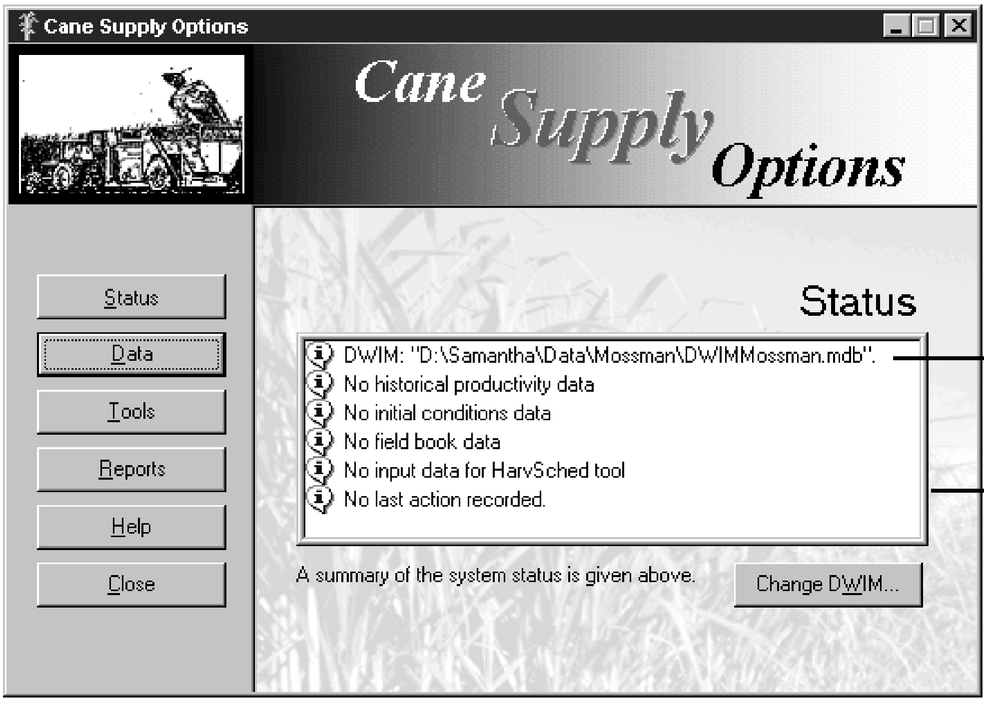
1. Starting the Cane Supply Options program.
2. Changing to a different pane.

Cane Supply Options program shortcut

In Detail

Starting the Cane Supply Options Program

To start the Cane Supply Options program, locate the "Cane Supply System" shortcut either on the desktop or via the **Start** button on the Windows taskbar (try under Programs). Select the icon (see figure above) and click to start the program. A splash form will appear followed by the Cane Supply Options program's main form.



The main form of the Cane Supply Options program

Changing to a Different Pane

Part of the Cane Supply Options program main form is a rectangular box that display a list of options. The combination of these options and the box is called a pane. The pane that displays when the Status button is clicked is called the Status pane. Each of the buttons on the left side of the main form (except the close button) has a corresponding pane to the right. To change to a different pane, click on the relevant button on the left side. For example, to display the "Help" pane, click the **Help** button. Alternatively, press the "Alt" and "H" keys simultaneously to display the Help pane. Notice that the word directly above the pane is the name of the currently displayed pane.

Connecting to a DWIM

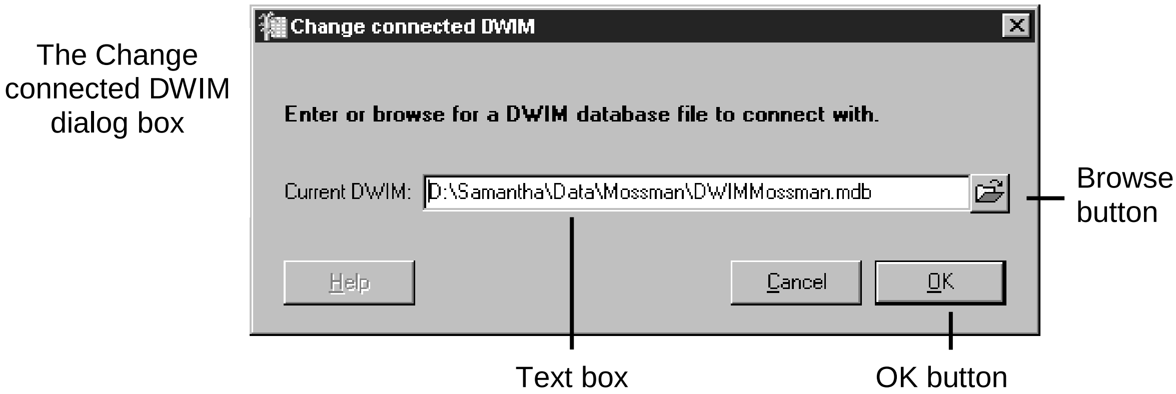
Briefly

1. Start the Cane Supply Options program or change to the “Status” pane (see appropriate section).
2. Show the “Change connected DWIM” dialog.
3. Enter the DWIM’s drive, path and file name or browse for the DWIM.
4. Verify on the “Status” pane that the correct DWIM is connected.

In Detail

Showing the “Change Connected DWIM” Dialog Box

The button located below the Status pane will have one of two phrases on its face. It will either say “Connect DWIM…” if a DWIM is not connected or “Change DWIM…” if a DWIM is connected. Click this button to show the “Change connected DWIM” dialog box. Alternatively, press the “Alt” and “W” keys simultaneously.



Changing the Connected DWIM

If a DWIM is already connected, this DWIM will be displayed in the text box on the dialog box. If a DWIM is not connected, this text box will be empty. Whichever DWIM is displayed in this text box when the **OK** button is clicked will become the connected DWIM. Thus, to connect to a different DWIM, alter the file name in the text box to that of another DWIM. Or, click the browse button to show the “Browse” dialog box. Locate the appropriate DWIM file in the directory tree and click the **Open** button. The Browse dialog box will close and the located DWIM’s file name will now appear in the text box. If the full name is not fully visible, click the text box and move the cursor with the arrow keys to scroll the name back and forth. Click the OK button to close the Change connected DWIM dialog box and connect to the specified DWIM. The Status pane of the Cane Supply Options program will be displayed.

Verify Connected DWIM

The connected DWIM’s drive, path and file name is displayed on the first line of the Status pane. Check that this is the correct DWIM. A cross will appear if there is no connected DWIM.

Importing Field Book Data

Briefly

1. Start the Cane Supply Options program or change to the “Status” pane (see appropriate section).
2. Check the connected DWIM and change if necessary.
3. Select the “Import Data” option from the “Data” pane.
4. Select “Field Book Data” item from the “Data” folder.
5. Continue through the wizard as instructed.
6. Verify that field book data has been imported.

In Detail

Checking the Connected DWIM

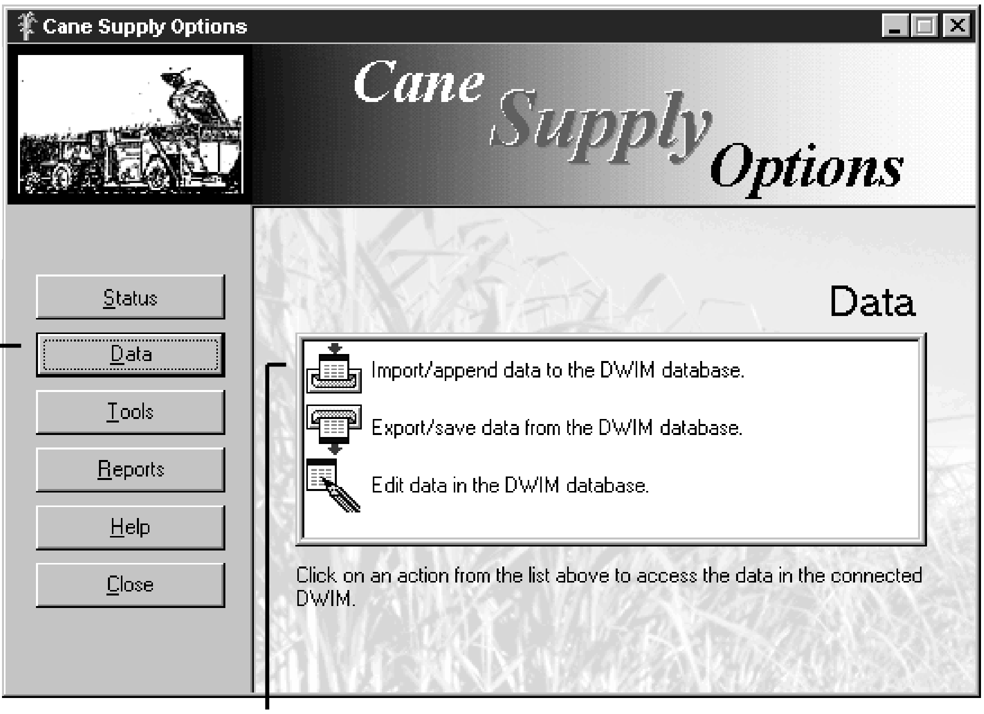
Check that the connected DWIM is correct. The connected DWIM’s drive, path and file name is displayed on the first line of the Status pane. A cross will appear if there is no connected DWIM. You must connect to a DWIM before proceeding (see section on connecting to a DWIM). See section on connecting to a DWIM if the connected DWIM needs to be changed.

Showing the Data Pane

To show the Data pane of the Cane Supply Options program, click the **Data** button. Alternatively, press the “Alt” and “D” keys simultaneously. The Data pane will replace the current pane on the Cane Supply Options main form.

The Cane Supply Options program with the Data pane displayed

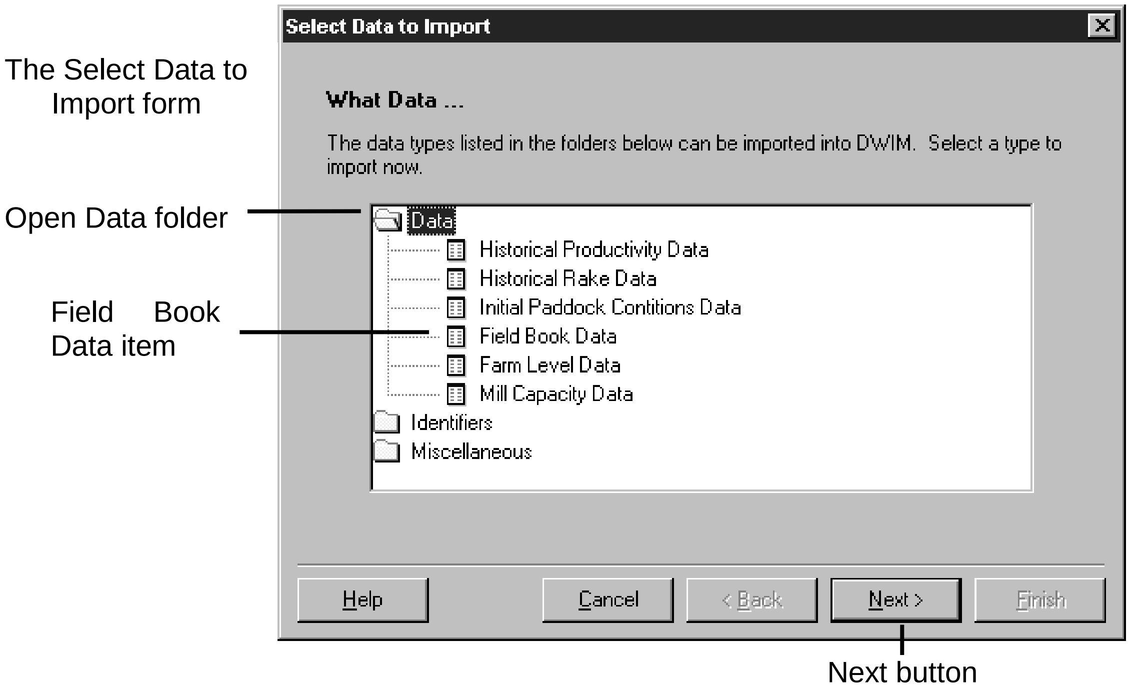
Data button



Import Data option

Selecting the Import Data Option

The Import Data option is the first line in the Data pane. When the mouse is over an option in the pane, the mouse icon will change from a to a and the option will become underlined. Click the **Import/Append data to the DWIM database** option. The “Select Data to Import” form will be displayed.



Selecting the Field Book Data Item

When the Select Data to Import page first displays, the Data folder may be closed . Double-click on the **Data** folder to open it and display the items it contains. Each of the items is a type of data that can be imported into the connected DWIM. There are other types of data that can be imported into DWIM in the Identifiers and Miscellaneous folders. Click the **Field Book Data** item in the Data folder to choose to import this type of data. Click the **Next** button to proceed to the first page of the Import Field Book Data wizard. The first page of the wizard is headed “Data File …”

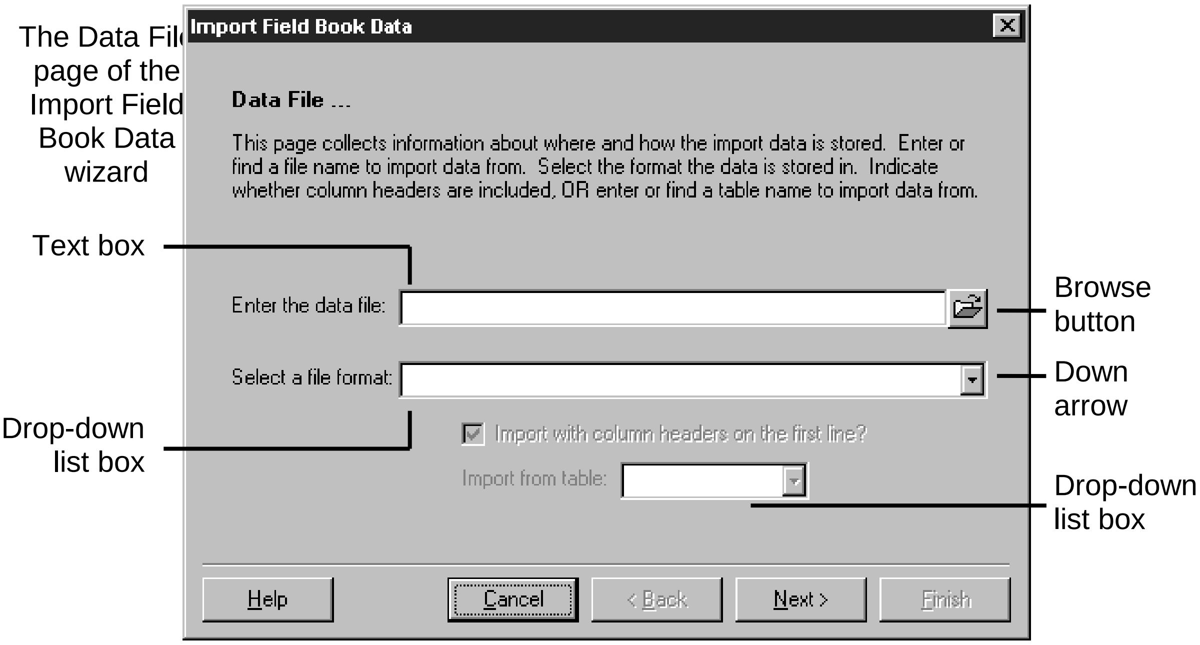
Entering Your Data File

There are two ways of entering your data file name. You can either

- type the file name (it is recommended that the file’s drive and path be included) in the “Enter the data file:” text box, or
- click the browse button and select your file in the dialog box that appears, as you would in other Windows programs. When you click the **Open** button on this dialog box, the file name is placed in the text box and the file type is selected from the file format drop-down list box automatically.

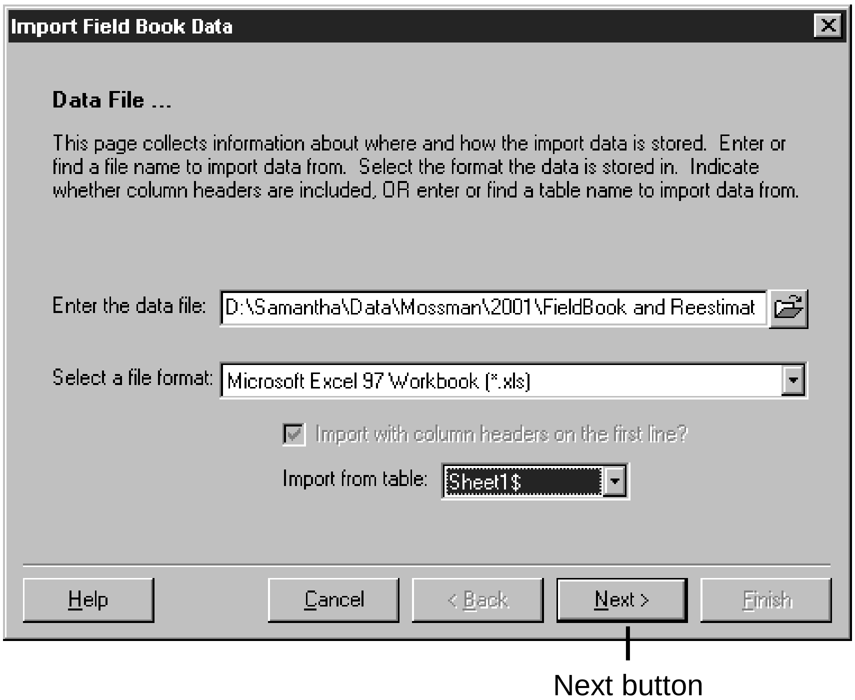
When you type the file name in the text box, you must also select the file format from the list of recognised file types. Click the down arrow on the right of the “Select a file format:” drop-down list box to expose the list of recognised file types. Click the appropriate file type.

The Data File page of the Import Field Book Data wizard



Once a file format has been selected, the "Import from table:" drop-down list box will become enabled (ie. not greyed out). Again, click the down arrow on the right of this box to expose the tables in your file. Click the name of the table containing the data you wish to import.

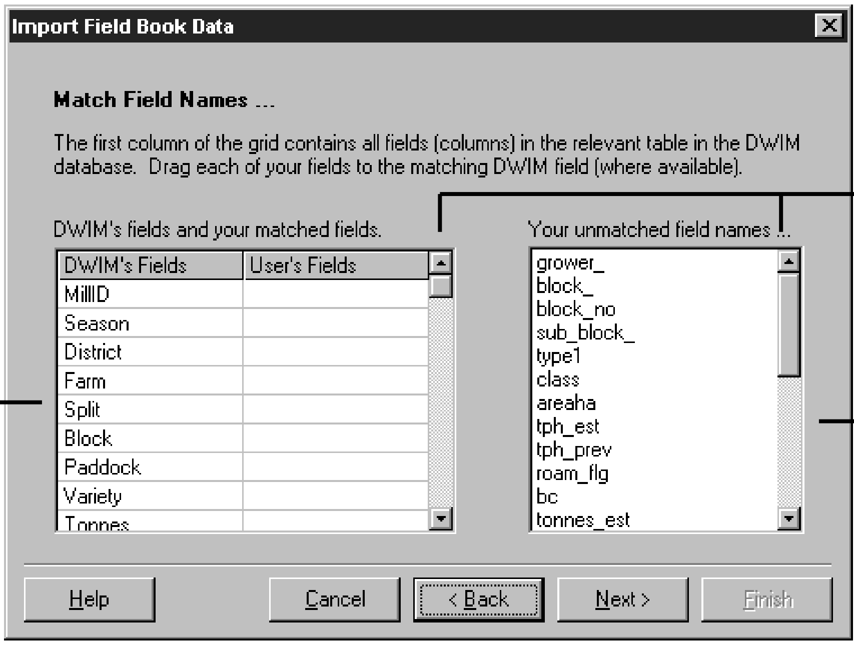
Completed Data File page of the Import Field Book Data wizard



ⓘ **Note**: Each worksheet in an Excel workbook is treated as a table. So, if your data is in "Sheet1" of the Excel workbook, your completed wizard page would look something like the above figure.

When you have entered the name of your data file and selected its file format and the table where the data is located, click the **Next** button to continue to the next wizard page. The next page of the wizard is headed "Match Field Names …".

The Match Field Names page of the Import Field Book Data wizard



Scroll bars

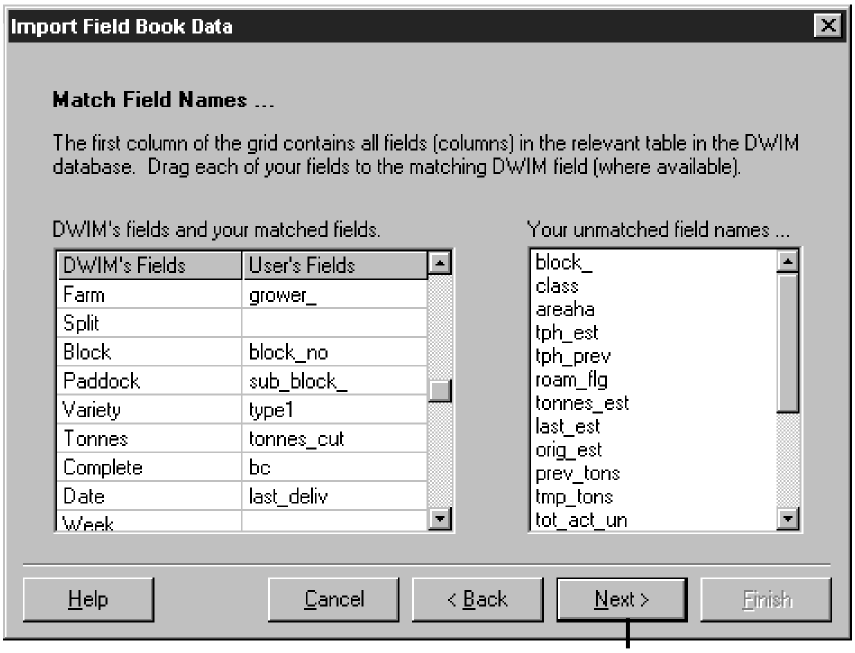
Matched field names grid

User's field names list box

Matching Your Fields to DWIM's

This page of the wizard displays the possible fields (columns of data) for field book data in DWIM as well as the user's fields from the data file table that was given in the previous wizard page. The DWIM fields are contained in the left column of the matched field names grid. These names can not be moved. The user's fields are contained in the list box on the right of the page. These names can be dragged across to the right column of the grid. The wizard saves matches for future imports so some of the user fields may already be matched to DWIM fields (ie. the right column of the grid may already contain user fields). Not all user fields need to be matched to a DWIM field before continuing to the next wizard page. Only the user fields that are matched to DWIM fields will be imported into DWIM.

Completed Match Field Names page of the Import Field Book Data wizard

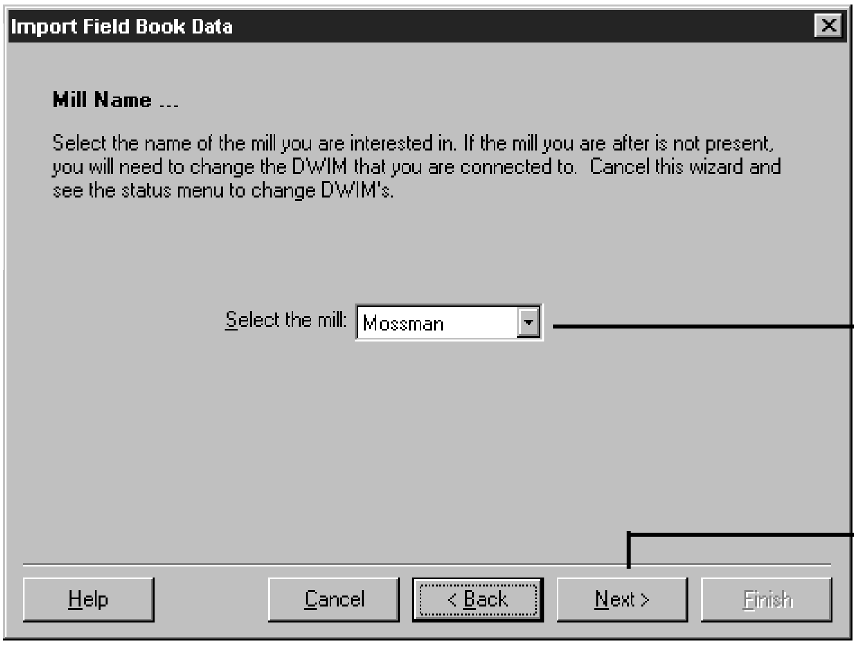


Next button

To match a user field to a DWIM field[3], select the user field from the list box. Drag the user field to the right column of the grid. The mouse icon will change from a to a ⃠ as you drag it across the wizard page. If the user's field is "dropped" whilst the mouse icon is a ⃠, the field will stay where it was. When the mouse icon changes to a as a user field is being dragged, the field can be "dropped" at this point on the wizard page. This "drop" icon will display as the mouse moves over an empty row in the right column of the grid. Drop the user field in the same row of the grid as the DWIM field you wish to match it to. If the field has been dragged to the wrong row of the grid, drag it back to the list box. It will be added to the end of the list of user fields. Use the scroll bars, if necessary, to position the list box and grid before dragging the user field to the grid. The finished page will look something like the one above.

Once the fields that need to be matched have been, click the **Next** button to continue to the next wizard page. The next page of the wizard is headed "Mill Name …".

The Mill Name page of the Import Field Book Data wizard



Drop-down list box

Next button

Selecting the Mill Name

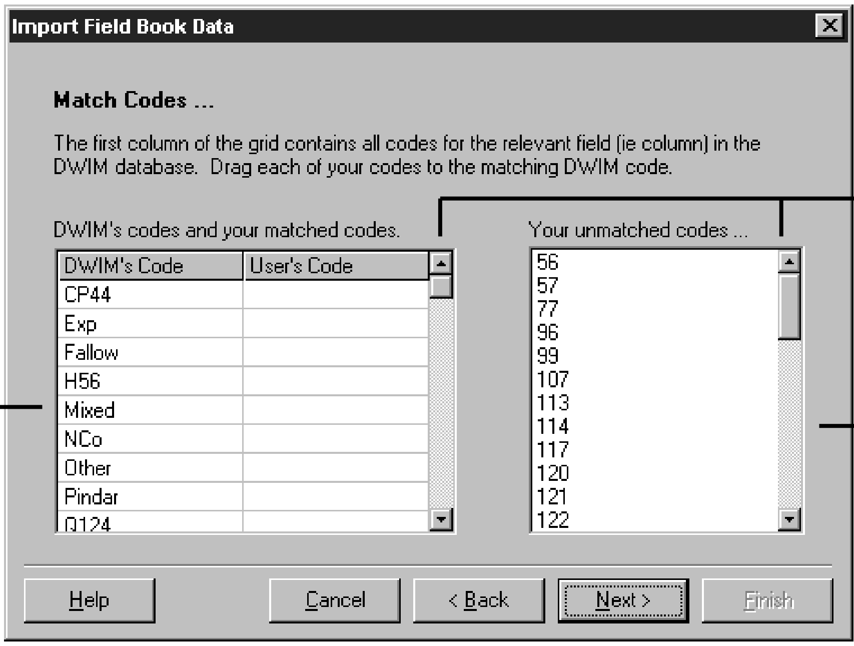
It is necessary to specify the mill that the field book data is from, especially for multi-mill regions. The name of a mill will automatically appear in the "Select the mill:" drop-down list box. If this is not the mill the data is from, click the down arrow on the right of the drop-down list box to expose the list of mill names. Click the correct mill name to select it. If the name of the mill is not present, report to the program distributor. Click the **Next** button to continue to the next wizard page. The next page of the wizard is headed "Match Codes …".

Matching Your Variety Codes to DWIM's

This page of the wizard displays the more common variety codes in DWIM as well as the user's variety codes (those codes found in the variety column of the user's data file table). The DWIM codes are contained in the left column of the matched codes grid. These names cannot be moved. The user's variety codes are contained in the list box on the right of the page. These codes can be dragged across to the right column of the grid. The wizard saves matches for future imports so some of the user variety codes may already be matched to DWIM variety codes (ie. the right column of the grid may already contain user variety codes).

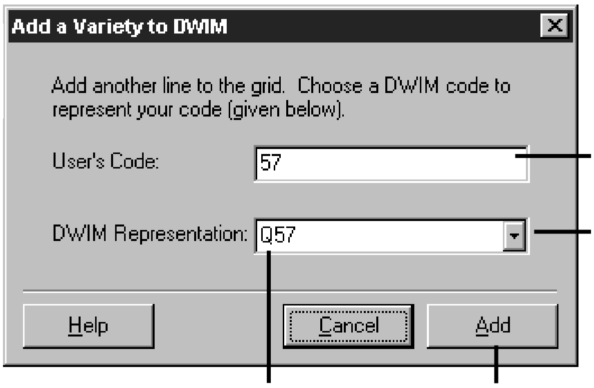
[3] See DWIM document for a description of DWIM fields.

All user variety codes need to be matched to a DWIM variety code before continuing to the next wizard page. Only one user variety code can be dropped on each line of the grid. New lines can be added to the matched codes grid (as an exhaustive list of varieties is not shown) when required.



To match a user variety code to an existing DWIM variety code, select the user variety code from the list box. Drag the user variety code to the right column of the grid and drop it in the same row of the grid as the DWIM variety code you wish to match it to (same as for field names).

The Add a Variety to DWIM dialog box



To add a new line to the grid, select the user variety code from the list box as before. Drag the user variety code to the left column of the grid and drop it anywhere (when the mouse icon changes to a). The "Add a Variety to DWIM" dialog box will appear with the "dragged" user variety code and a suggested DWIM variety code. The user variety code cannot be changed in this dialog box (click the **Cancel** button if you have selected the wrong code).

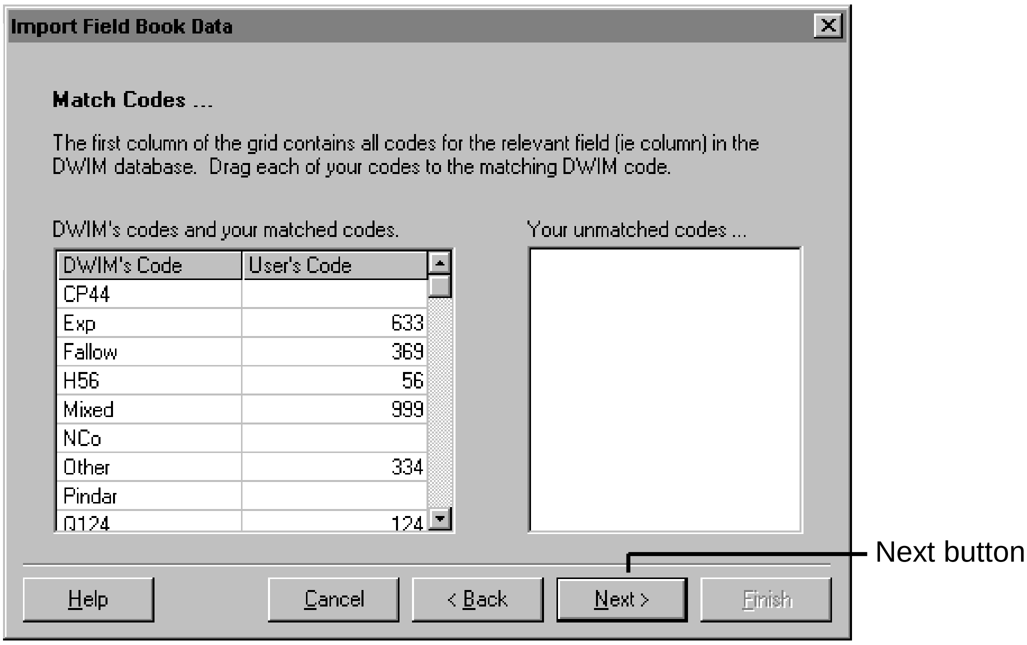
If the suggested DWIM variety code is acceptable, click the **Add** button to add this line to the grid on the Match Codes page of the wizard. The Add a Variety to DWIM dialog box will close and the codes will be added to the end of the grid. It may be necessary to scroll down to view the addition.

If the suggested DWIM variety code is not acceptable, click the down arrow on the right of the drop-down comb box to view alternative DWIM codes. Click an appropriate variety code to select it and then click the **Add** button as before.

If an appropriate DWIM variety code cannot be found in the drop-down combo box, then a code can be entered in the text part of this box. Click in the rectangular box to the left of the down arrow (clear the suggested code if necessary). Enter the appropriate code by typing the letters and numbers and click the **Add** button as before.

Warning: Whilst entering your own DWIM variety code is possible, the models that use this data will not necessarily be able to recognised this as a separate code. Use this facility with caution.

Completed Match Codes page of the Import Field Book Data Wizard



Once all the user variety codes have been matched to a DWIM variety code, click the **Next** button to continue to the next wizard page. The next page of the wizard has three sections of information, a “Units of data in fields” frame, a “Harvest start” frame and a “Complete indicator” frame.

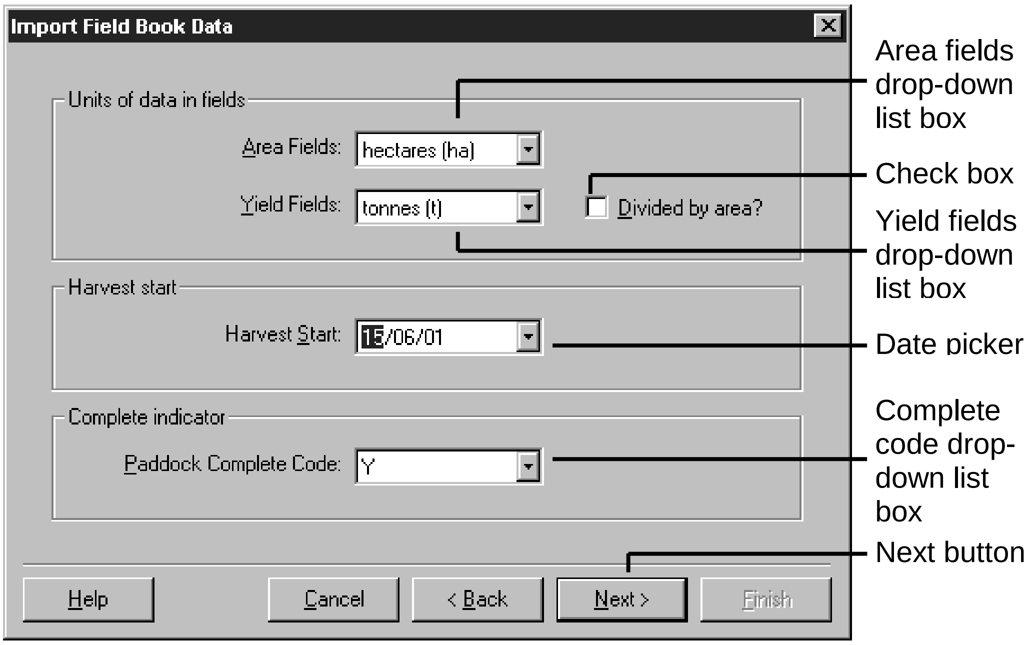
Completing the Three Framed Wizard Page

This page of the wizard allows the user to change the units of measurement for certain fields, the harvest start date and also specify their paddock complete code. Check the values displayed in each frame and change where necessary. Details on how to change the values in each frame are given below. When the page is correct, click the **Next** button to continue to the next wizard page. The next page of the wizard is headed “Finish …” and is the last page.

Changing Data Field Units

This frame of the wizard page allows the user to specify the units that their area and yield fields are measured in. For field book data, only a yield field (ie Tonnes) will be imported, hence it is not necessary to change the area field units. To change the specified units if the user’s Tonnes field is not in tonnes, click the down arrow on the right of the yield fields drop-down list box to view alternative units of measurement. Click the appropriate unit to select it.

The Import Field Book Data wizard page with three sections

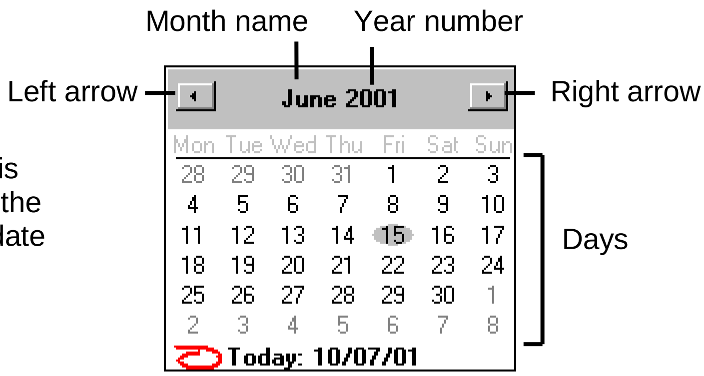


Warning: Yield fields that are expressed as percentages of area (eg. units of data values are t/ha) are not supported in this version. That is, checking the "Divide by area?" check box will have no effect on the importing process. Thus results from the model could be skewed as it is expecting yield values that aren't divided by their area.

Changing Harvest Start Date

The date picker in the harvest start frame allows the starting date for the harvest season to be set. There are two methods for changing the harvest start date. Method one: click the down arrow on the right of the date picker to expose the calendar and change the calendar (described below). Method two: click on the date and use the arrow keys on the keyboard or enter numbers.

The calendar that is exposed by clicking the down arrow on the date picker



The calendar displays one month at a time. Click the left arrow to change the month that is displayed to the previous month (eg. from June to May). Click the right arrow to change the month to the next month (eg. from June to July). Click down and hold with the left mouse button on either of the arrows to scroll through the months until the mouse button is released. Click the month name to reveal a list of months and click one of these to change to this month. Click the year number to reveal up-down arrows and click these to move a year at a time. The up-down arrows can also be clicked and held to scroll through the years. Click a day to change to that day in the current month and year.

Changing Paddock Complete Code

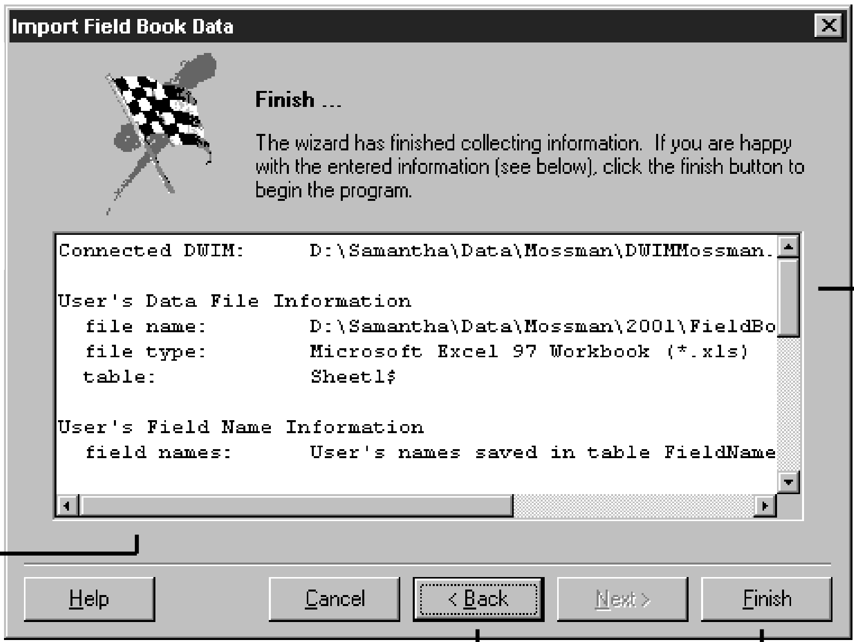
This frame of the wizard page allows the user to specify which code indicates that their paddock has been completely harvested (ie. all cane has been harvested from that paddock). The wizard automatically chooses a code but this may not be the correct one. To change the complete code, click the down arrow ▾ on the right of the complete code drop-down list box to view the other codes present in the user's data. Click the correct complete code to select it.

> **Warning**: It is only possible to assign one code as the complete paddock indicator. Importing data that has more than one complete paddock indicator will cause unreliable results to be produced. If your data has more than one complete code, contact the Program distributor or alter your data so that it has only one code for complete paddocks.

The Finish Page

The finish page displays a summary of the data entered into the wizard. Use the horizontal and vertical scroll bars to move the visible part of the summary. Check the summary and if it is correct, click the **Finish** button to begin importing field book data. Click the **Back** buttons if any of the summary is incorrect and needs to be changed.

The Finish page of the Import Field Book Data wizard



Vertical scroll bar

Horizontal scroll bar

Back button

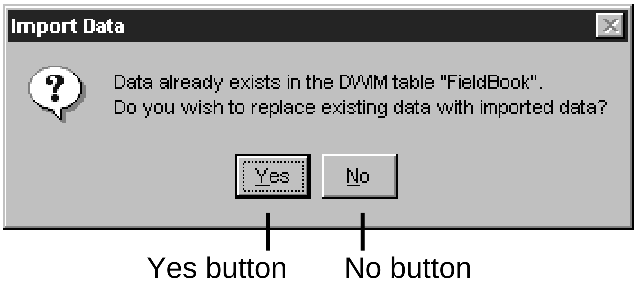
Finish button

> **Note**: The wizard will not go back past the "Data File" page.

Replacing Existing Data

The "Replace Existing Data" dialog box will be displayed if field book data has been imported previously. Click the **Yes** button to replace the previously imported data with the current data. Click the **No** button to stop the import of the data (and therefore not replacing the existing data). DWIM is not designed to store previous sets of field book data. The field book data table in DWIM will need to be copied and renamed if the previously imported data is to be kept.

The Replace Existing Data dialog box



Verify the Field Book Data has been Imported

Once the import has completed, the Cane Supply Options program will be left displaying the Data pane. Change to the Status pane by clicking the **Status** button. The last line of this pane specifies the last successful action performed by the program. If this does not say “Field book data imported” then the data has not been imported. The forth line of this pane specifies the date and time of the last successfully imported field book data. This time is recorded when the data has been successfully imported. Thus, if this date or time does not coincide with the import just performed, then the data has not been imported.

Trouble Shooting

Once the Finish button has been clicked the program begins the process of importing the user-specified data into the “FieldBook” table in the connected DWIM. The program creates a temporary table called “Temp_*tablename*” in the connected DWIM (where *tablename* is the name of the user’s table as specified in the “Data File” wizard page). The user’s data is imported into this table, one field at a time. Once the data has been imported successfully, it is copied from the temporary table into the field book table. However, if a problem occurs, the temporary table remains and the field book table is not altered. In this case, the temporary table gives an indication of where (ie. what field and what record) the importing failed.

If the import has failed, close the Cane Supply Options program and open the connected DWIM. You cannot open the connected DWIM whilst the Cane Supply Options program is opened. Click the **Tables** tab to view the list of tables in the database. Select the temporary table and click the **Open** button to view the data in the table.

Starting the Cane Yield Estimator Program

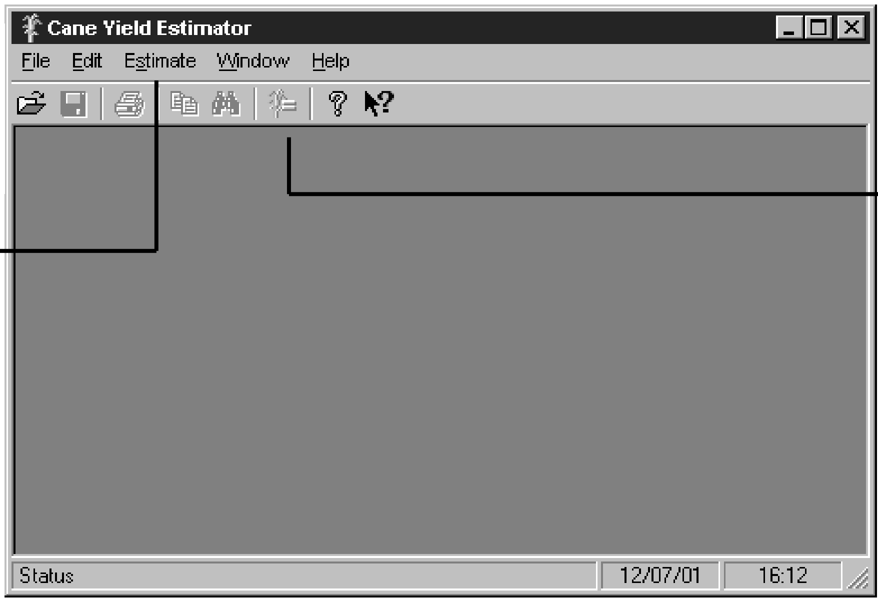
Briefly

1. Starting the Cane Yield Estimator program.

In Detail

Starting the Cane Yield Estimator Program

To start the Cane Yield Estimator program, locate the "Cane Yield Estimation" shortcut either on the desktop or via the Start button on the Windows Taskbar (try under Programs). Select the icon and click to start the program. A splash form will appear followed by the Cane Yield Estimator program's main form.



Cane Yield Estimator program's main form

Estimate menu button

Calculate estimate tool bar button

Producing Mid Season Estimates

Briefly

1. Import the latest field book data into DWIM (see appropriate section).
2. Start the Cane Yield Estimator program (see appropriate section).
3. Select “Calculate…” from the “Estimate” menu.
4. Continue through the wizard as instructed.
5. Save the Cane Yield Estimation results.
6. Close the Cane Yield Estimator program.

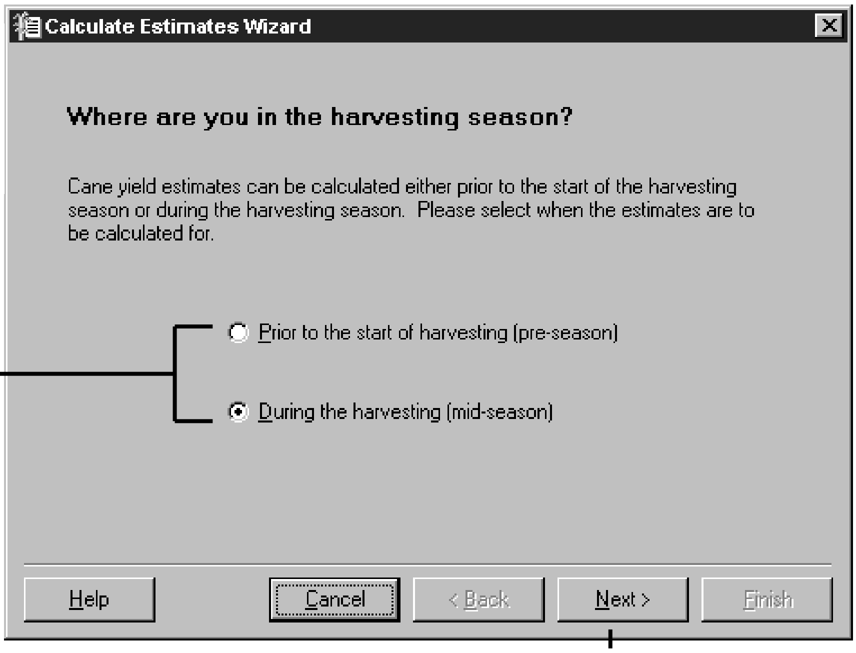
In Detail

Starting the Calculate Estimates Wizard

There are several methods for starting the Calculate Estimates wizard. They are

- click the **Estimate** menu button to expose the Estimate menu then click the **Calculate…** menu item,
- click the **calculate estimate** tool bar button, or
- press the “Alt” and “S” keys simultaneously, followed by the “C” key.

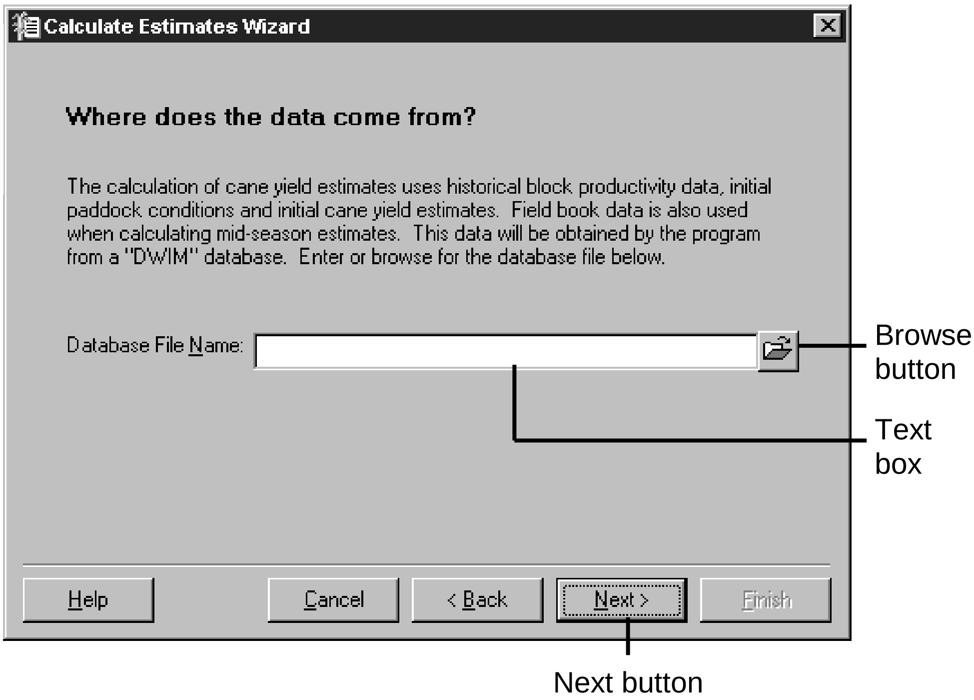
The first page of the wizard asks “Where are you in the harvesting season?”



Choosing Mid Season Estimates

The first page of the wizard allows the user to choose between calculating pre-season estimates or mid-season estimates. The mid-season estimates option is the default. The selected option has a black dot in the white circle to its left. Clicking the other option button will make it the selected option. Make sure that the mid-season option is selected. Click the **Next** button to continue to the next wizard page. The second page of the wizard asks “Where does the data come from?”

The data file name wizard page



Specifying Where the Data Is

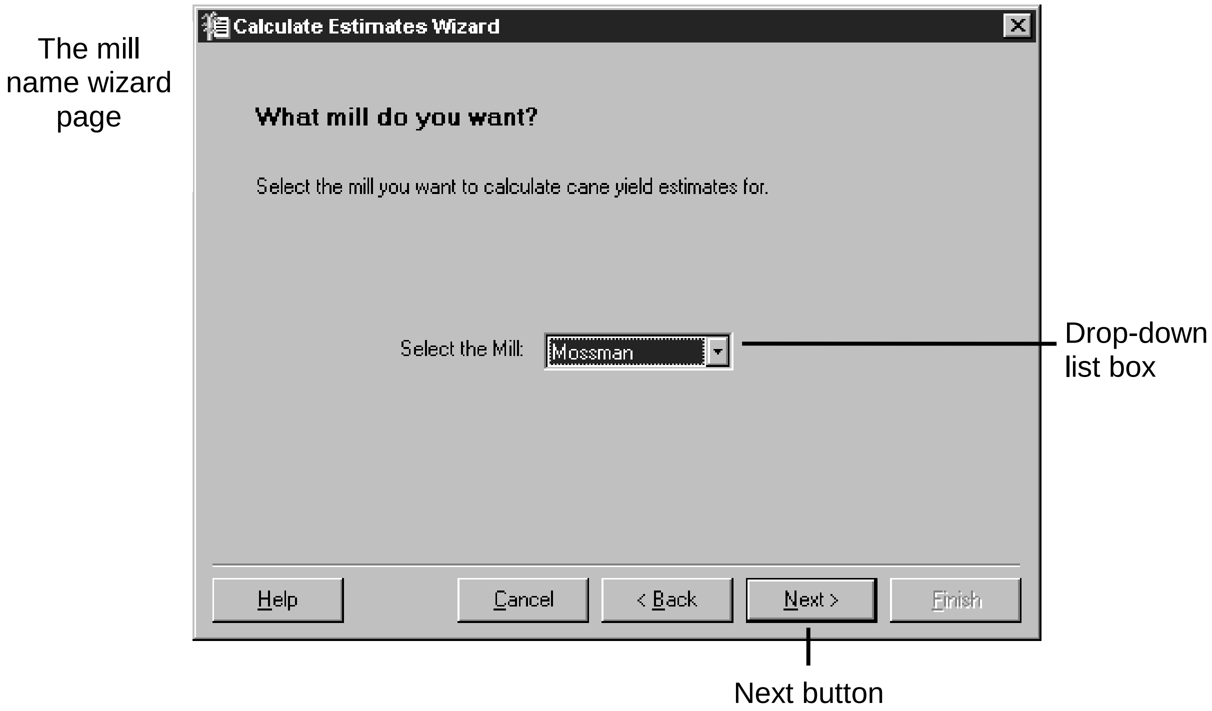
The second page of the Calculate Estimates wizard contains a text box for entering the name of the DWIM file containing the input data. It also contains a browse button that can be used to help locate the DWIM file. Please specify the same DWIM that the data for this estimate (eg field book data) was imported into.

To enter the DWIM file in the text box, click in the text box to activate the cursor (blinking line), then type the file name. It is recommended that the drive and full path of the DWIM file name be entered. Use the browse button to help specify the drive and full path. Click the browse button and select the DWIM file in the “Set DWIM database” dialog box that appears, as you would in other Windows programs. When you click the **Open** button on this dialog box, the selected DWIM file name (including drive and full path) is placed in the text box. Click the **Next** button to continue to the next wizard page. The third page of the wizard asks “Where does the data come from?”

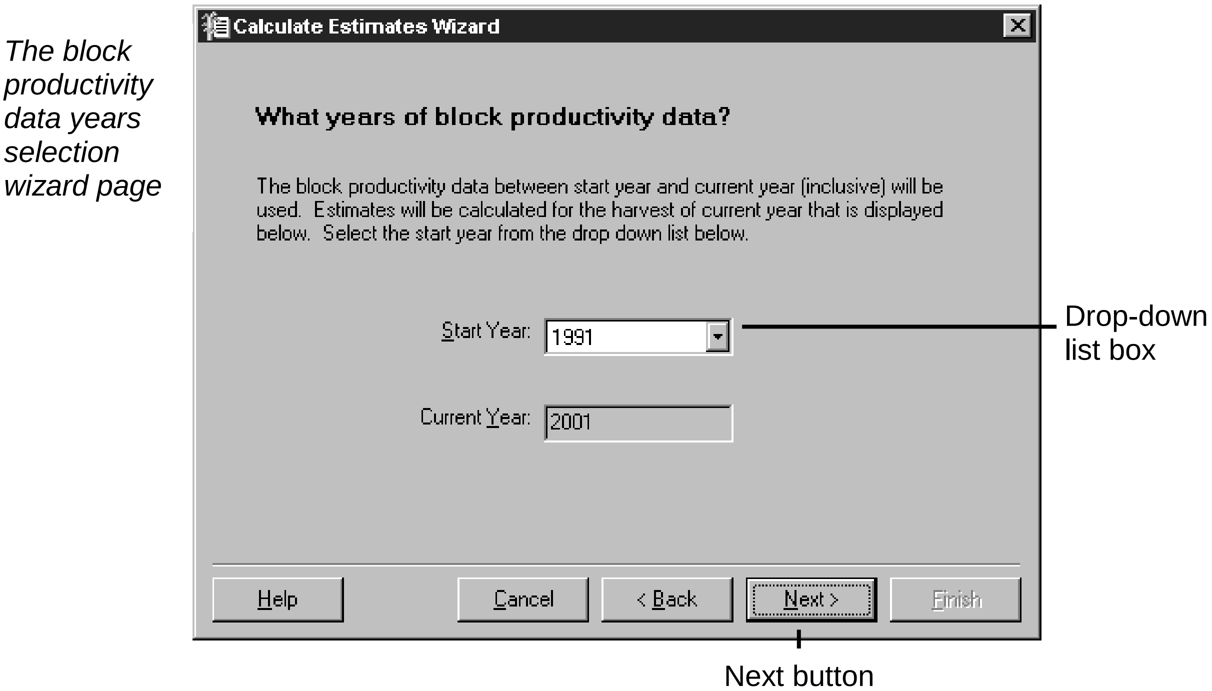
Selecting the Mill Name

The third page of the Calculate Estimates wizard is used to specify which mill estimates are to be produced for. The name of a mill will automatically appear in the “Select the mill:” drop-down list box. If this is not the mill that estimates are required for, click the down arrow on the right of the drop-down list box to expose the list of available mill names. Click the correct mill name to select it. If the name of the mill is not present, report to the program distributor.

ⓘ **Note**: For multi-mill regions, estimates can only be produced for one mill at a time. Run the Calculate Estimates wizard for each mill where estimates are required. A document containing the resulting estimates will be created, in the main form, each time the wizard is run.



Click the **Next** button to continue to the next wizard page. Before displaying the next page, the wizard will access the specified DWIM file to gain information that it needs for the next page. Whilst it is doing this, the mouse icon will change from an to an . The forth page of the wizard asks “What years of block productivity data?”

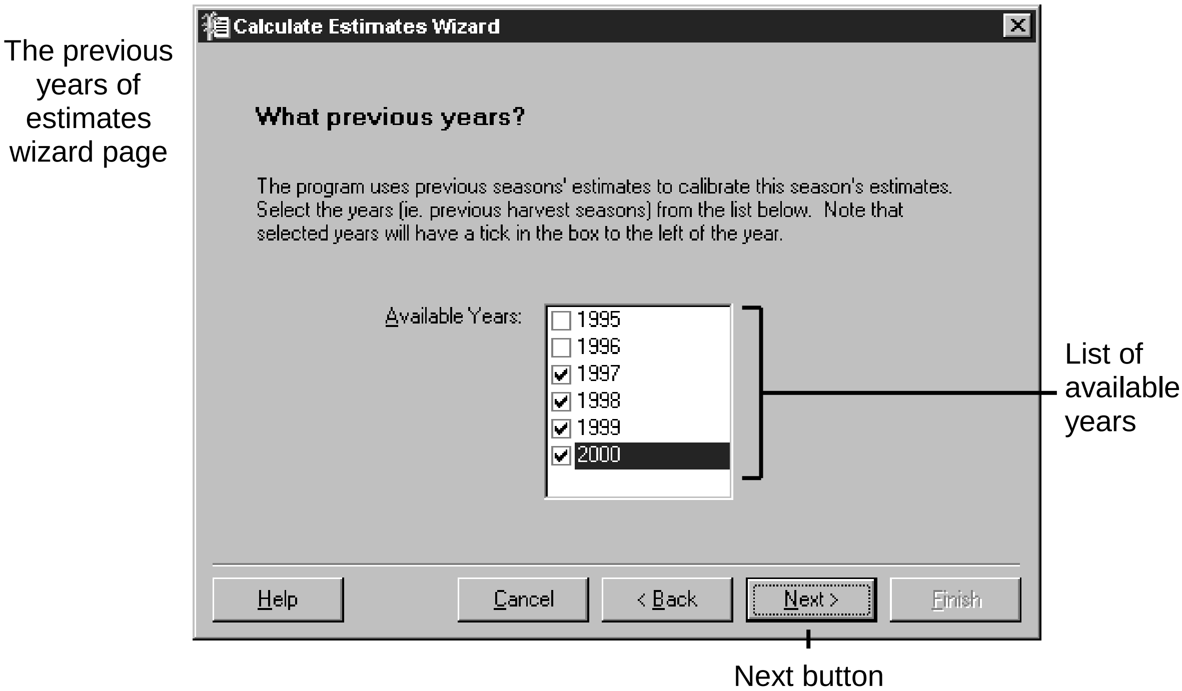


Years of Block Productivity Data

The forth page of the wizard allows the user to truncate the years of block productivity data that will be used to calculate estimates. The oldest available year of data will automatically be displayed in the “Start Year:” drop-down list box (eg in the figure above, this is 1991). It also displays the current year or newest year of data (eg in the figure above, this is 2001). Thus, for the figure above, the estimator would use 1991 – 2000 block productivity data and 2001 paddock initial conditions data (block productivity data without the results). To change the start year, click the down arrow on the right of the drop-down list box to expose the list of data years. Click the required year to select it. For example, if 1994 was chosen, the estimator would use 1994 – 2000 block productivity data and 2001 paddock initial conditions data.

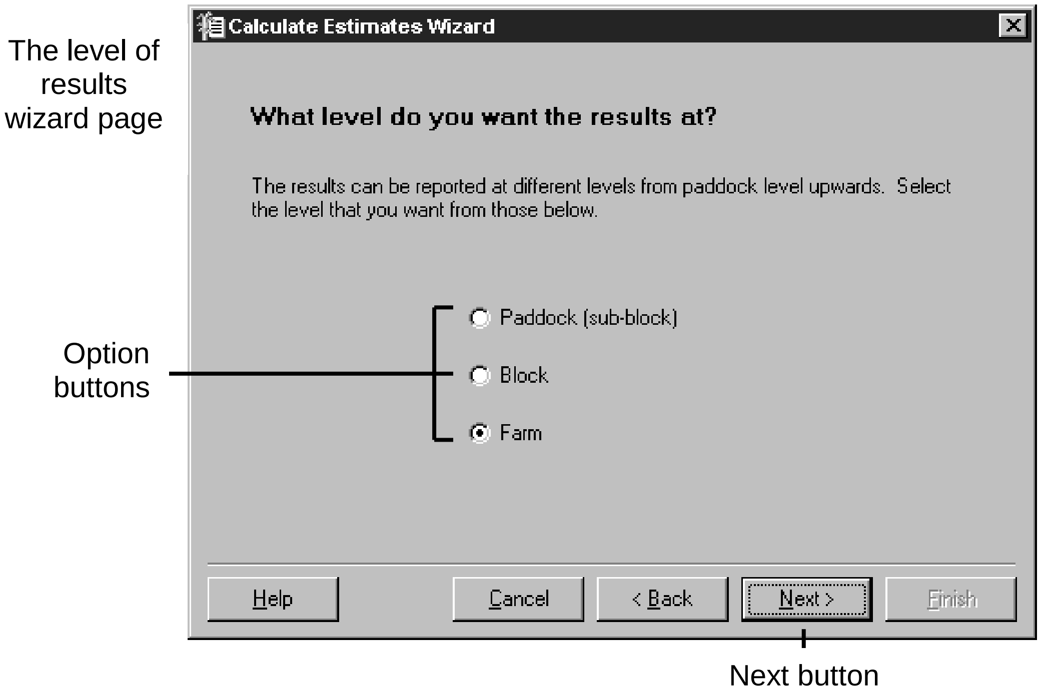
ⓘ **Note**: The current year is based on the initial conditions data imported into the specified DWIM. To alter this, a new set of data from a different year would need to be imported into the DWIM.

Click the **Next** button to continue to the next wizard page. Before displaying the next page, the wizard will access the specified DWIM file to gain information that it needs for the next page. Whilst it is doing this, the mouse icon will change from an to an . The fifth page of the wizard asks “What previous years?”



Previous Years of Estimates

The fifth page of the wizard allows the user to select which of the previous years of estimates the estimator will use. The wizard automatically selects the last four available years by ticking the box beside the year. To change the selection status of any of the years, click the box to the year’s left to give it the opposite state. For example, the box beside the year 2000 in the figure above is ticked. Clicking the box will make the tick disappear and hence make the year unselected. Clicking the box again will make the tick reappear and hence make the year selected again. (If your clicks are too close together the computer will treat them as a double-click which, in this case, is the same as a single click.) When the required years have been selected (at least one must be), click the **Next** button to continue to the next wizard page. The sixth page of the wizard asks “What level do you want the results at?”



Level of Results

The sixth page of the wizard allows the user to specify what level the results will be presented at. The default option is results at “Farm” level. The selected option has a black dot in the white circle (option button) to its left. Click the option button to the left of the required option to select this option. Click the **Next** button to continue to the next wizard page. The last page of the wizard asks, “Have you finished?”

ⓘ **Note**: All calculations are made at the paddock level and then aggregated if the user requests block or farm level. Thus, paddock level results will take slightly less time, as they do not need the extra summation that the other levels do.

The Finish Page

The finish page allows the user one last chance to change what they have entered in the wizard. Click the **Back** button (on each wizard page) if changes need to be made. Otherwise, click the **Finish** button to begin calculating the estimates.

After the Finish Page

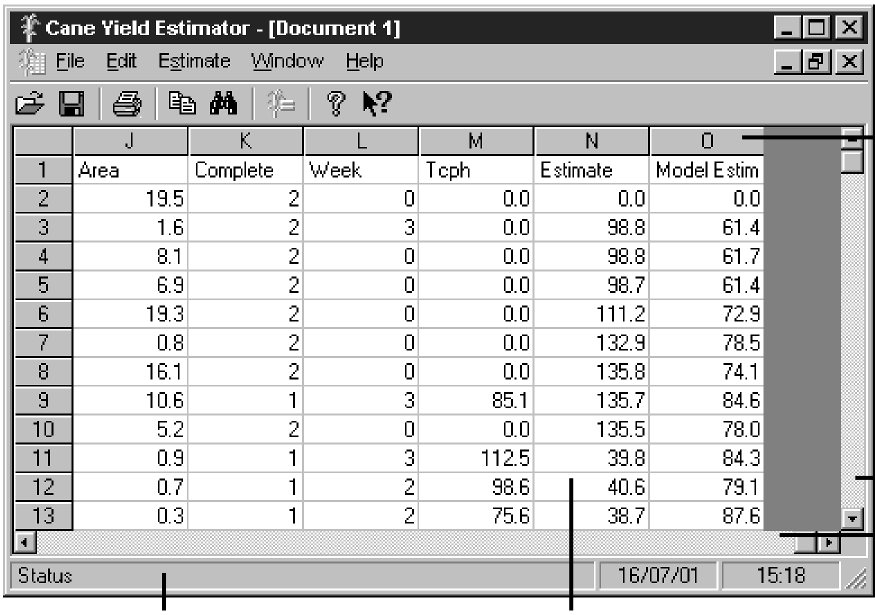
Once the Finish button on the Calculate Estimates wizard is clicked, the program will gather the specified data, calculate cane yield estimates and present the results at the specified level in a spreadsheet like document. The mouse icon will change from an to an when the finish button of the wizard is clicked and remain as an until the results have been displayed. An empty document form will open in the main program form. This form will contain the estimates once they have been calculated. The status bar of the main program form will indicate the task the program is currently completing. Once calculated, the estimates are displayed in the document with other identifying information (eg farm, block and paddock numbers). The calculated estimates can be found in the column headed “Model Estim”. Use the scroll bars to move the visible part of the document around to view other columns and/or rows.

ⓘ **Note**: The paddocks presented in the results are those from the initial conditions data. The field book data paddocks have been matched to these, where possible. The "Tcph" column contains tonnes of cane per hectare for completed paddocks only.

The Finish page of the Calculate Estimates wizard

Back button Finish button

The results at paddock level



Column of calculated estimates

Scroll bars

Status bar Document containing results

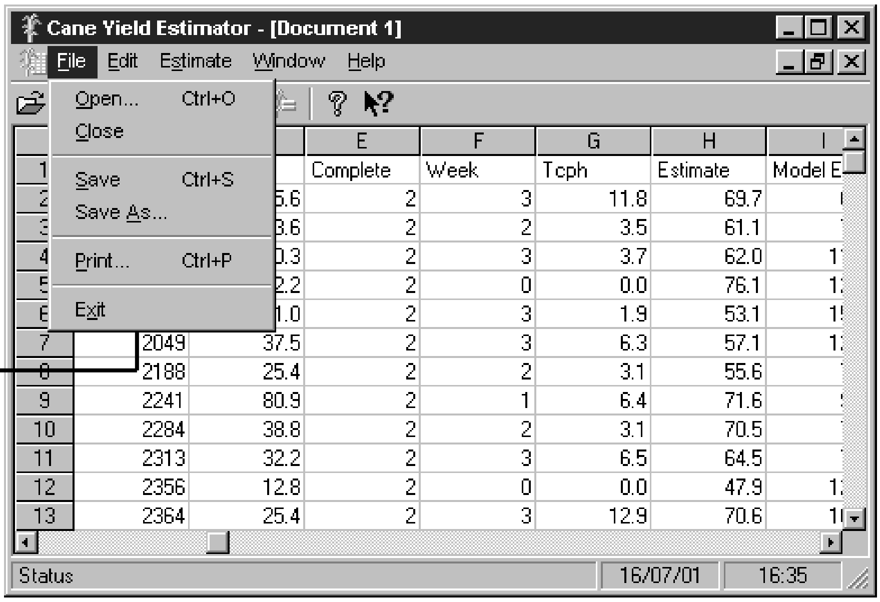
Saving the Document

The "Save As" dialog box will allow the user to specify where to save the document. There are several methods for showing this dialog box. They are

- click the **File** menu button to expose the File menu then click the **Save** or **Save As…** menu item,
- click the **save** tool bar button 💾, or
- press the "Ctrl" and "S" keys simultaneously.

Once this dialog box is open, specify where to save your document as you would in other Windows programs. Click the **Save** button to save the document to the specified file. Once you have saved the file, the dialog box will not appear unless you select the "Save As…" item from the File menu.

The main program form with the file menu exposed



File menu

> ⚠ **Warning**: Saving the document as a Microsoft Excel file has proved incredibly slow on some machines. In these cases, selecting the entire document, copying and then pasting into a Microsoft Excel workbook may work a bit faster (see below).

Copying To Excel 97

To copy the entire document (grid of results) and paste in a Microsoft Excel Workbook, follow the steps below.

1. Select the entire document by
   - clicking the blank rectangle to the left of the "A" column and above the row labelled "1", or
   - clicking the **Edit** menu button to expose the Edit menu and then clicking the **Select All** menu item, or
   - press the "Ctrl" and "A" keys simultaneously.
2. Copy the selected area by
   - clicking the **Edit** menu button to expose the Edit menu and then clicking the **Copy** menu item, or
   - press the "Ctrl" and "C" keys simultaneously.

   This operation may take some time to complete depending on the number of records in your data. The mouse icon will become an ⌛ whilst the program is copying the data. Also, the word "Copying…" will display in the status bar until the copying operation has completed.
3. Open Microsoft Excel or open a new Workbook in Microsoft Excel.
4. Paste the copied data by
   - clicking the **Edit** menu button to expose the Edit menu and then clicking the **Paste** menu item, or
   - press the "Ctrl" and "V" keys simultaneously.

Glossary

This glossary explains some of the terms used in the main body of the text.

Button	an object displayed on a form that when clicked performs a function, reminiscent of a button on a calculator or control panel.
CPU	central processing unit, the part of the computer that does the processing (or computing).
Click	push down and release the primary mouse button (on a right-handed mouse this will be the button on the left).
Cursor	a (usually) blinking object that indicates the position where the next character will be placed.
Database	a device that stores data using tables with fields and records
Dialog box	a form that gathers information from the user for a specified operation. It does not contain other forms (eg. the form that is displayed when you select open from the file menu in Word). It usually needs some action by the user before continuing with other actions in the program (called a modal form).
Double-click	two clicks of the primary mouse button in quick succession
Drag	push the primary mouse button down over a selected item but don't release, move the mouse (whilst holding the primary mouse button down).
Drop	usually follows a drag operation, release the primary mouse button when the mouse icon is at the appropriate place.
DWIM	an Access database with a structure designed specifically for use with the Cane Supply Options and Cane Yield Estimator programs.
Field	a column in a database table.
Form	the graphical front (usually rectangular) to a program or operation that interfaces with the user. A form can contain other forms (eg. when you open Word a form is displayed and then when you open a document a form is displayed within the initial form).
Monitor	the visual display unit (or screen) attached to the computer.
Mouse	a piece of hardware attached to the computer that is used to move the mouse icon around the monitor.
Mouse button	a button (usually) on the mouse.
Mouse icon	a picture that the user moves around the monitor (via the mouse). Its most common form is a white arrow. This picture has a hot-spot which is (usually) centrally located. The hot-spot is where any action by the mouse will be interpreted.
Pane	an area on the Cane Supply Options program that displays (and encloses) options related to the last clicked button.
Record	a row in a database table.
Select	click on an object to highlight it.
Shortcut	an icon that is linked to a program located somewhere else in the computer's directory structure. Selecting and clicking this icon will start the linked program.
Splash Form	a form that is displayed briefly while the main program is loading.

Table a database table is a structure that has two dimensions (ie. matrix) and stores data.

Taskbar a Microsoft Windows (95, 98, NT, 2000) operating system feature that contains the Microsoft Windows Start button and system time plus buttons for any other open programs.

Wizard a set of dialog boxes (displayed one at a time) that steps the user through an operation or process. The dialog boxes (usually) have next and back buttons for navigation through the set.

Wizard page one of the dialog boxes in the wizard.

APPENDIX B.

Optimal Cane Supply Input Data

Instructions and Questionnaire

S. L. Peel and A. J. Higgins

COOPERATIVE RESEARCH CENTRE *for*
SUSTAINABLE SUGAR PRODUCTION

Sugar
Research
And
Development
Corporation

CSIRO

PART 1 - INSTRUCTIONS

1 About this Document

This document provides a record of the necessary input data for the generic Cane Supply Optimisation model[4]. Every effort has been made to provide a widely useable version (within the industry) of the model's input data from experiences with several regions. However, differences may exist and should be discussed with the CSIRO contact at an early stage. Note that, in some cases, it was necessary to choose between terminologies used by industry partners to create this generic format. This does not imply that those not chosen are invalid or inferior. Nor is it expected that industry partners (present and future) adopt the terminologies or formats described here. A proposed development of the model will be the inclusion of an input database. This database will assist industry partners to gather the input data and transform it to the required format for the Cane Supply Optimisation model. From discussion of the generic input requirements with industry partners (usually a mill representative involved with data collection and storage), specific input requirements for a mill region can then be identified and incorporated into the model. Therefore, allowing the model to more closely reflect the current state of the mill region.

Once input requirements have been agreed upon, the document assists industry partners to provide the input data in the right format for the model. It also allows a record to be kept of the progress of data collection. When data collection is complete, the document facilitates quick transfer of information from the industry partner to the CSIRO staff member.

The document is separated into two parts. Part one contains the instructions for providing the generic input data, for a single mill region, in the required format. Part two contains a questionnaire that is to be completed and returned, along with the electronic data, to CSIRO. The electronic data should be either space delimited text files or Microsoft Excel spreadsheets. The data is used to efficiently generate the key model input data for producing optimal cane supplies and conducting options analyses. Any questions or comments, please contact

Dr Andrew Higgins
CSIRO Tropical Agriculture
120 Meiers Rd
BRISBANE 4068
Ph: (07) 3214 2340
E-mail: andrew.higgins@tag.csiro.au

[4] Higgins, A. J. (1999). Optimizing cane supply decisions within a sugar mill region. *Journal of Scheduling* **2**, 229-244.

2 FARM DATA

2.1 Historical Block Productivity Data

The Historical Block Productivity data contains information about each paddock for all districts over a number of harvesting seasons. The information should be stored in a single text file (using space delimited columns) or a single work sheet in Microsoft Excel. Please have the data organised into the following columns with the given headings:

Column Number	Heading	Description
1	season	Season of harvest information (eg. 1994)
2	district	Product district or zone identification number
3	farm	Farm identification number
4	split	Split farm identification within the farm. For example, if the farm is split over two locations, give one location a split number of 1 and the other location a split number of 2. Enter 1 for those farms that are not split.
5	block	Block identification number within the farm
6	paddock	Paddock or sub-block identification number within the block (eg. for paddock A enter 1). If the paddocks are the same as the blocks (ie. the blocks are not divided into smaller areas), enter 1 for all values.
7	variety	Variety of cane grown on the paddock (eg Q124) up to 10 characters in length. Apart from the normal variety designation, “mixed”, “exp” (for experimental), “fallow” or “other” can be used. Note that the case of the letters is not important.
8	class	Crop class of the cane on the paddock 1 for plant crop (after fallow or plough out) 2 for first ratoon 3 for second ratoon ⋮ 7 for sixth ratoon and older ratoons 9 for stand over crop 10 for fallow paddock
9	plant	Plant type of the paddock’s crop 0 for not known or undetermined 1 for planted after fallow 2 for planted after plough out
10	area	Area of the paddock (as specified in column 6 above) in hectares. Include an area for paddocks that are fallow.
11	week	Major week of harvesting of the paddock expressed as the number of weeks that have elapsed from the beginning of the season. It is assumed that the first week of the season is the 15th to 22nd June in any year. Give cane harvested before this week a value of one.
12	weekpyr	Major week of harvesting of the paddock in the **previous** season, expressed as the number of weeks that have elapsed from the beginning of the season (see column 11 above). **This column of data is desirable but not necessary.**
13	age	Age of plant crops (not necessary for ratoon crops). The age is the number of weeks between planting and harvesting. For example, the age of a crop planted in the first week of July and harvested in the second week of July (the following season) is 53 weeks. **This column of data is desirable but not necessary.**

14	tonnes	Tonnes of cane cut in a completed harvest season from this paddock.
15	ccs	Actual ccs of the cane cut in a completed harvest season from this paddock (eg. 13.20).
16	irrigated	Whether the paddock was irrigated or not 0 for not irrigated 1 for irrigated **This column of data is desirable but not necessary.**
17	green	Percentage of green cane (of total tonnes of cane) for each paddock. Represent the percentage as a number between 0 and 100 (ie. green cane divided by total cane multiplied by 100). **This column of data is desirable but not necessary.**

2.2 Wet Blocks

The wet block information is used to establish which paddocks are susceptible to moisture and generally can not be harvested, by the available machinery, whilst wet. This information may be too difficult to implement in some regions due to the subjective nature of the data required and/or the regions weather patterns.

If a paddock is generally wet at certain times of the year, please provide this information. It is only necessary to provide wet block data for the last completed harvest season (last year of data provided in the historical block productivity data). This information should be stored in a single text file (using space delimited columns) or a single work sheet in Microsoft Excel. It is only necessary to provide the farm, split, block and paddock identification number for those paddocks that are susceptible to moisture. Please have the wet data organised into the following columns with the given headings:

Column Number	Heading	Description
1	farm	Farm identification number
2	split	Split farm identification within the farm. For example, if the farm is split over two locations, give one location a split number of 1 and the other location a split number of 2. Enter 1 for those farms that are not split
3	block	Block identification number within the farm
4	paddock	Paddock or sub-block identification number within the block (eg. for paddock A enter 1). If the paddocks are the same as the blocks (ie. the blocks are not divided into smaller areas), enter 1 for all values.
5	dates	Enter the dates that a paddock is generally too wet to harvest (eg. “31/05 – 01/07, 01/10 – end” means that the paddock is too wet from the end of May through to the beginning of July and from October onwards). Note that it is expected that the entries in this column will not all be the same length.

2.3 Farm Level Information

The farm level information includes data about farm ownership and harvester group usage for each farm (and split). The farm ownership information is used in the calculation of farm equity. The farm level information should be stored in a single text file (using space delimited columns) or a single work sheet in Microsoft Excel. Please have the data organised into the following columns with the given headings:

Column Number	Heading	Description
1	farm	Farm identification number
2	split	Split farm identification within the farm. For example, if the farm is split over two locations, give one location a split number of 1 and the other location a split number of 2. Enter 1 for those farms that are not split
3	farmer	Farmer identification name (i.e. use the same name for each farm that is owned by the same farmer) up to 50 characters in length
4	group	Harvester group identification number (i.e. the harvester group that the farm, taking into account the split, belongs too) *OR* Harvester group name (up to 50 characters in length)

2.4 Initial Paddock Status

This is the initial status of each paddock at the beginning of the new season. The status of a paddock includes the variety and class of crop growing there. This information may not be available until a few months before the beginning of the season. When the data is unavailable, the program will use the last season of block productivity data to initialise the paddocks. However, when the data is available, it should be stored in a single text file (using space delimited columns) or a single work sheet in Microsoft Excel. Please have the data organised into the following columns with the given headings:

Column Number	Heading	Description
1	district	Product district or zone identification number
2	farm	Farm identification number
3	split	Split farm identification within the farm. For example, if the farm is split over two locations, give one location a split number of 1 and the other location a split number of 2. Enter 1 for those farms that are not split.
4	block	Block identification number within the farm
5	paddock	Paddock or sub-block identification number within the block (eg. for paddock A enter 1). If the paddocks are the same as the blocks (ie. the blocks are not divided into smaller areas), enter 1 for all values.
6	variety	Variety of cane grown on the paddock (eg Q124) up to 10 characters in length. Apart from the normal variety designation, “mixed”, “exp” (for experimental), “fallow” or “other” can be used. Note that the case of the letters is not important.
7	class	Crop class of the cane on the paddock 1 for plant crop (after fallow or plough out) 2 for first ratoon 3 for second ratoon ⋮ 7 for sixth ratoon and older ratoons 9 for stand over crop 10 for fallow paddock
8	plant	Plant type of the paddock’s crop 0 for not known or undetermined 1 for planted after fallow 2 for planted after plough out

9	area	Area of the paddock (as specified in column 5 above) in hectares. Include an area for paddocks that are fallow.
10	weekpyr	Major week of harvesting of the paddock in the **previous** season, expressed as the number of weeks that have elapsed from the beginning of the season. It is assumed that the first week of the season is the 15th to 22nd June in any year. Give cane harvested before this week a value of one. **This column of data is desirable but not necessary.**

2.5 Growing Costs

The growing costs are split into two groups. They are

- the cost of planting, maintaining and harvesting a new crop and
- the cost of maintaining and harvesting a ratoon crop.

These costs are required, at district level, to establish the differences between a plant crop and a ratoon crop. In some cases, the differences between costs for individual farmers within a district will be too great to give an accurate representation at district level.

Tables are provided in the questionnaire (see section 3 Farmer Costs) to record these numbers. There is a separate column provided for each district, if needed. Otherwise (costs are the same for all districts), enter the costs in the first district's column and tick the "same for all districts" box at the top. Note that it is understood that not all costs are applicable in a district and that non-applicable costs will be left blank.

2.6 Farmer Equity

The farmer equity information is the number of rounds of harvesting and what percentage of cane was cut in each round. It is only necessary to provide farmer equity information for the last season of harvesting.

A table is provided in the questionnaire (see section 4 Farmer Equity) to record these numbers. Please enter the percentage of cane cut for each round. For example, if 10% of the cane was cut in round one, 40% in round two, 40% in round three and 10% in round four, then the completed table in the questionnaire would be

Round	1	2	3	4	5	6	7	8
Percentage	10	40	40	10				

3 Harvester Data

3.1 Harvesting Groups

The harvesting group information supplies the model with names for each group. If not provided in the Farm Level Information section, the information should be stored in a single text file (using space delimited columns) or a single work sheet in Microsoft Excel. Please have the data organised into the following columns with the given headings:

Column Number	Heading	Description
1	group	Harvester group identification number (if numbers are used in the Farm Level Information section)
2	name	Harvester group name (up to 50 characters in length)

3.2 Season Start and Finish

For the model, it has been assumed (due to historical block productivity restrictions for Mossman and Mackay case studies) that the harvesting season starts in the third week of June and finishes at the end of November. However, for all mills, this is not generally the case for either starting or finishing dates. The actual starting and finishing dates for the seasons covered by the historical block productivity data are required. This information should be stored in a single text file (using space delimited columns) or a single work sheet in Microsoft Excel. Please have the data organised into the following columns with the given headings:

Column Number	Heading	Description
1	season	Season of harvest information (eg. 1994)
2	start	The first date of harvesting (dd/mm/yyyy) in the season
3	finish	The last date of harvesting (dd/mm/yyyy) in the season

4 Milling Data

4.1 Relative CCS Cane Payment Formula

Please provide the formula used to calculate the payment to the farmers for the cane they supplied. A space is provided in the questionnaire (see section 6 Cane Payment Formula).

4.2 Capacity and Stoppages

The capacity of the mill and the percentage of stoppages should be provided for each fortnight of the season. A table is provided in the questionnaire (see section 7 Milling Data) to record these numbers. Please enter the capacity of the mill in tonnes per hour (use actual capacity from the last completed season if necessary). Also enter the percentage of time lost in each fortnight due to all types of stoppages and outside stoppages.

$$\% \text{ Total Stops} = \frac{\text{Hours not crushing due to problem (eg. maintenance, breakdown)} \times 100}{\text{Total available crush time in hours}}$$

$$\% \text{ Outside Stops} = \frac{\text{Hours not crushing due to wet weather (ie. no supply)} \times 100}{\text{Total available crush time in hours}}$$

Note that % Total Stops should include stoppages due to wet weather. For example, here is a section of the table from the questionnaire (filled with fictitious numbers).

	Fortnight Ending					
	29 Jun	12 Jul	27 Jul	10 Aug	24 Aug	…
Rated Capacity (t/hr)	500	504	521	521	521	…
% Total Stops	13.0	13.0	11.5	8.5	9.5	…
% Outside Stops	6.3	4.8	5.4	7.0	5.1	…

4.3 Costs

Some costs associated with the operation of the mill are required. The provision of this data should be discussed with Dr Russell Muchow (CSIRO Tropical Agriculture) and will be treated with the utmost care with respect to confidentiality.

5 District Information

This section in the questionnaire (see section 8 District Information) is provided so that interesting features of particular districts or what made a district a district can be noted. Any maps that can be provided would be appreciated. Please feel free to add to this (or make no comments at all).

6 Transport Data

6.1 Rail

Please skip this sub-section if the region does not use railways to transport the harvested cane.

Several types of information are required to properly access the rail transport capacity and cost. First, a couple of definitions for terms used in the description of the required transport data.

Loco area: The line taken (or area serviced) by a loco in a round trip. It is possible that a loco will travel through another loco area before reaching its designated area.
Round trip: The trip from the mill out to the end of the run and back to the mill.
Shift: The time a loco and its crew are together each day (for example a loco running for 8 hours with the same crew).

6.1.1 Loco Areas

It is possible that the loco areas are different to the districts described in section 5 above. Please use the space provided in the questionnaire (see section 9 Rail Transport Data) to give details about the different loco areas. Any restrictions should also be noted here. Restrictions may take the form of upper limits on supply of cane per day from a particular area, or, certain areas may only be able to take smaller capacity bins.

For those regions where loco areas differ from districts, please provide a list of farms in each loco area. The data should be stored in a single text file (using space delimited columns) or a single work sheet in Microsoft Excel. Please have the data organised into the following columns with the given headings:

Column Number	Heading	Description
1	farm	Farm identification number
2	split	Split farm identification within the farm. For example, if the farm is split over two locations, give one location a split number of 1 and the other location a split number of 2. Enter 1 for those farms that are not split
3	loco area	Loco area identification number (please use the same order or number as used in the details from the previous paragraph)

Also required is the maximum capacity (maximum tonnes of cane) per shift that each loco area is capable of transporting back to the mill. These numbers will be based on the number of round trips possible per shift, maximum tonnage that a loco can pull and load restrictions on all or part of the line. Note that load restrictions may be automatically included as they would effect round trip times. Assume an unlimited supply of bins and cane is possible in each area when calculating this number. This information should be entered in the table provided in the questionnaire (see section 9 Rail Transport Data). For example, the completed table for Pleystowe Mill is:

Loco Area	Maximum Capacity (t/shift)
1 Over Gap	414
2 Brightley-Mainline	538
3 Barrie	521
4 Savannah	1120
5 Victoria Plains	961
6 Glens-Wiford-Cooks	919
7 Winston-Crow Hill	386
8 Palms Palmyra Cassanda	1050
9 Nebia Shannons	896
10 Te Kowai	896
11 Oakenden	521

6.1.2 Loco Shifts

The number of hours in a shift, the cost of a shift (per day) in dollars, the maximum possible number of shifts per day the mill can run, and the current number of shifts per day are required. The maximum number of shifts per day should be based on crew availability and loco numbers. This information should be entered in the table provided in the questionnaire (see section 9 Rail Transport Data). For example, the completed table may look like this.

Length (hrs)	Cost ($ per shift)	Maximum Number (per day)	Current Number (per day)
8	$320.00	30	27

The maximum number of shifts that can be allocated to each loco area in a day (24-hour period) is required. Consider each loco area separately and assume, initially, that unlimited shifts are available in each. That is, ignore that other loco areas need to be serviced at the same time. Then apply restrictions on having unlimited shifts such as the capacity of sidings in a loco area. The number of shifts for each loco area should be entered in the table provided in the questionnaire (see section 9 Rail Transport Data). For example, the completed table for Pleystowe Mill is:

Loco Area	Maximum Number of Shifts per Day	Current Number of Shifts per Day
1 Over Gap	8	
2 Brightley-Mainline	6	
3 Barrie	6	
4 Savannah	5	
5 Victoria Plains	3	
6 Glens-Wiford-Cooks	4	
7 Winston-Crow Hill	2	
8 Palms Palmyra Cassanda	7	
9 Nebia Shannons	9	
10 Te Kowai	2	
11 Oakenden	5	

6.1.3 Bins

The capacity and number of each bin type, that the mill currently has, is required. This information should be entered in the table provided in the questionnaire (see section 9 Rail Transport Data). For example, there are two types of bins and a total of 2000 bins. 1500 of these bins have a capacity of 5.6 tonnes each and the other 500 have a capacity of 3.6 tonnes each. The completed table, for this example is:

Capacity (t)	5.6	3.6	
Number	1500	500	

Note that any restrictions on bins in certain loco areas should be noted in the loco area information section in the questionnaire.

6.2 Road

Please skip this sub-section if the region does not use roadways to transport the harvested cane. Note that it is assumed that any vehicle can visit any farm (ie. road areas are not used).

6.2.1 Capacity and Cost

Several types of information are required and are to be entered into two tables in the questionnaire. The following information is entered in the first table. The vehicle type, vehicle capacity (in tonnes), number of shifts per day, and cost per shift (in dollars) are required for all types of vehicles used in the region. A table is provided in the questionnaire (see section 10 Road Transport Data) to record these values. For example, if there were two types of vehicles (semi and b-double) with a capacity of 23 and 34 tonnes respectively, the completed table in the questionnaire would be

Vehicle Type	Semi	b-double						
Capacity (t)	23	34						
Number of Vehicles	130	12						
Shifts per Day	2	1						
Cost per Shift ($)	800	880						

Note that in the above example, the 130 semis have two shifts per day and each shift costs $800.00.

Secondly, the number of round trips (mill to district to mill) each vehicle type makes to each district in a shift. A table is provided in the questionnaire (see section 10 Road Transport Data) to record these values. For example, given the above vehicle types, each vehicle can make 2 round trips per shift to district 1, and 3 round trips per shift to district 2. The completed table in the questionnaire would be

Vehicle Type	**District**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
semi	2	3					
b-double	2	3					

PART 2 - QUESTIONNAIRE

1 Contact Information

Name of the region or mill..
Contact person ....
Contact phone ....
Contact email..

2 Farm Data

Name of the file storing the historical block productivity data ....
Name of the file storing the wet block data (if available)..
Name of the file storing the farm level information data..
Name of the file storing the initial paddock status data (if available)..

3 Farmer Costs

Please enter the costs that were incurred in each district in the tables on the following pages.

☐ Same for all districts

Plant Costs	Cost in each District ($/ha)													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
FORM[1]:														
Fuel & Oil														
R&M – machinery														
R&M – irrigation														
Planting:														
Buy plants														
Planting														
Lime														
Pesticides:														
Fungicide														
SuSCon														
Lorsban														
Chemicals:														
Roundup														
Gesapax Combi														
Actril DS														
Gromoxone														
Diuron														
Fertiliser:														
Mixture														
Irrigation:														
Water costs														
Electricity														
Expendables														
TOTAL														

[1] Land preparation and crop management.

Ratoon Costs	Cost in each District ($/ha)													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
FORM[1]: Fuel & Oil														
R&M – machinery														
R&M – irrigation														
Chemicals: Roundup														
Actril DS														
Gromoxone														
Diuron														
Fertiliser: Mixture														
Irrigation: Water costs														
Electricity														
Expendables														
TOTAL														

4 Farmer Equity

Enter the percentage of cane cut in each round.

Round	1	2	3	4	5	6	7	8
Percentage								

5 Harvester Data

Name of the file storing the harvesting group name data (if necessary) ....
Name of the file storing the start-finish dates data ....

6 Cane Payment Formula

Enter the CCS cane payment formula here.

....

7 Milling Data

Please enter the mill capacity and the percentage of stoppages for each fortnight of the season.

	Fortnight Ending												
	29 Jun	**12 Jul**	**27 Jul**	**10 Aug**	**24 Aug**	**07 Sep**	**21 Sep**	**05 Oct**	**19 Oct**	**02 Nov**	**16 Nov**	**30 Nov**	**After**
Rated Capacity (t/hr)													
% Total Stops													
% Outside Stops													

8 District Information

....

....

....

....

....

....

....

....

9 Rail Transport Data

Record loco area information here.

....

....

....

....

....

....

....

....

....

....

Name of the file storing the loco area of all farms (if necessary) ....

Enter the maximum capacity per shift of each loco area.

Loco Area	Maximum Capacity (t/shift)

Enter the shift details.

Length (hrs)	Cost ($ per shift)	Maximum Number (per day)	Current Number (per day)

Enter the maximum number of shifts per day that can be allocated to each loco area.

Loco Area	Maximum Number of Shifts per Day	Current Number of Shifts per Day

Enter the capacity and number of each bin type.

Capacity (t)								
Number								

10 Road Transport Data

Enter the length, in hours, of a shift..

For each type of vehicle, enter its type (eg. semi), its capacity, the number of shifts per day and the cost per shift.

Vehicle Type										
Capacity (t)										
Number of vehicles										
Shifts per Day										
Cost per Shift ($)										

Enter the number of round trips per shift.

Vehicle Type	District													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14

APPENDIX C.

DWIM
Database for Whole of Industry Models

1. Purpose

For CSIRO, the purpose of the DWIM database (and its associated user interface) is to

- facilitate the collection of collaborating regions' relevant data whilst the results are being produced by us,
- automatically convert the data into the standard format (necessary for interaction with the models),
- provide a common place where error checked data can be found and used by all team members, and
- gather a large range of diverse data (ie fields) whose inclusion in the final models is under research.

For Industry, the purpose of the DWIM database is to

- facilitate the collection of the region's relevant data for a model run,
- automatically convert the data into the standard format (necessary for interaction with the models),
- present the data using the mill's codes so that it is instantly understood by them, and
- store only the data (ie fields) required by the final versions of the models.
- Input and Output?

2. Structure Overview

There will be a DWIM for each region (not mill). It will have the region's name appended to the end (eg DWIMMackay). The DWIM databases will contain data only (input and output). That is, the user interface will be contained in a separate database (called portData) that can link to any of the DWIM regional databases.

Within each DWIM there will be at least

- a mill table that contains a mill identification number for each mill name in the region [Mill],
- a current block productivity table that contains the initial conditions for each paddock before the commencement of the harvesting season [BlockProd_Current],
- a historical block productivity table that is updated at the end of each season with the completed block productivity data [BlockProd_Historical],
- a field book table that contains the block productivity data at a certain point in the harvesting season [FieldBook],
- a farm table that contains information that occurs at farm level (eg farm name, district number),
- several tables that provide unique identifiers within the region (eg tables Group and District), and
- several other tables that record standard codes for the codes used by the mills (eg tables FieldNames and UserCodes)

Note that historical information will only be retained in the historical block productivity table. Any data imported into a field book table, for example, will overwrite the existing data.

The data will be stored using standard codes. Thus, conversion from mill codes to standard will be required in the importing of data routines (ie in the user interface). However, “lookups” will be utilised to allow the mill operators to view the data in their codes. It is practical to store the data using standard codes so that the models are not performing conversions (from mill to standard codes) each time they run. Will this work for CSIRO staff viewing?

3. Standard Codes

Note that, for the Week and WeekPrevious fields, the week number provided by the mill region is used. That is, this number is not standard across all mill regions. The WeekBeginDate field is provided to allow some comparison between mill regions or other data (eg climate data) if required. Where the mill provides a date (rather than a week number), it shall be converted to a week number using the formulas specified in Appendix B.

The standard codes[5] for each field, as decided by CSIRO staff, are

Field	Codes
Area	A real number that gives the harvested area in hectares of the paddock referred to in the record.
ArrayElement	An integer that indicates that the value is this element of an array. If this value is 0 or Null then the value is a scalar!
Block	An integer that indicates which block within the farm-split the data in the record is from.
CaneShift	A real number that gives the maximum amount of cane that can be transported per fortnight in the transport area.
Ccs	A real number that gives the CCS of the cane harvested from the paddock referred to in the record.
CCSGroup	An integer number that indicates which CCS group within a mill the data in the record is from. The integer “22” will be assigned when it was not possible to place the record in a CCS group. This is provided by a statistical grouping method.
Cclass	See “Class” below.
Class	An integer that indicates the crop class of the paddock referred to in the record. The following crop classes are represented by 1 to indicate a plant crop (after fallow or replant) 2 to indicate a first ratoon crop 3 to indicate a second ratoon crop 4 to indicate a third ratoon crop 5 to indicate a forth ratoon crop 6 to indicate a fifth ratoon crop 7 to indicate a sixth or greater ratoon crop 9 to indicate stand over crop 10 to indicate a paddock that is fallow
ClassCode	A string (text value) that indicates the code used by the mill for the standard code referred to in the record.

[5] Note that real numbers are single reals and that integer numbers are long integers. This is due to the connection between Fortran and Visual Basic using a Dynamic Link Library (DLL).

Code1	An integer that indicates the row number of the parameter value contained in the record.
Code2	An integer that indicates the column number of the parameter value contained in the record.
Company	A string (text value) that is the name of the company that runs the mill (eg Mackay Sugar).
Complete	An integer number that indicates whether the paddock referred to in the record has been completely harvested 1 paddock has been completely harvested 2 paddock hasn't been completely harvested (includes no harvesting)
ConstName	A string that is the name of the CSO model constant whose value is stored in this record.
ConstType	An integer that indicates the data type of the CSO model constant value (i.e. Boolean=1, Text=10, Single=6, LongInteger=4). This uses DAO type constants (see DataTypeEnum in DAO 3.6 object library).
ConstValue	A string that is the value of the CSO model constant for this record's mill.
Contact	A string (text value) that is the name of the person who is the initial contact for information or data for this mill.
DatePlanted	A date that indicates the planting date, in the previous season, of a paddock that contains a plant crop (see Class). A zero date should be entered for any other crop class type. This field can be used to calculate the age of a plant crop at the time of harvest.
Description	A string (text value) that is a description of the value contained in the record.
District	An integer number that indicates which product district or zone within a mill the data in the record is from. This is as provided by the mill.
DistrictName	A string (text value) that is the name of the district or zone in this record (as industry provides).
Dwim	A string (text value) that is the code used by DWIM for the user code in the record.
Farm	An integer number that indicates which farm within a mill the data in the record is from. The combination of the MillID and Farm numbers should give a unique identifier when working within a region (if the region is not already using unique farm numbers). This is as provided by the mill.
FarmName	A string (text value) that is the name of the farm-split in this record (as industry provides).
FieldName	A string (text value) that is the name of a field in a DWIM table.
FieldNameMill	A string (text value) that is the name that the mill uses to represent a DWIM table field.
FarmNos	An integer number that indicates the number of farms in the CCS group.
Fortnight	An integer number that indicates which fortnight the data in the record is for (fortnight 1 begins on the 15th June in the current year).
Group	An integer number that indicates which harvesting group within a mill the data in the record is from. See MillGroupID for a unique identifier within a region. This is as provided by the mill.
GroupName	A string (text value) that is the name of the harvesting group in this record (as industry provides).

HDPY	An integer value that is used by HarvSched to indicate harvest time in the previous year. Same as WeekPrevious field except that values are fortnights not weeks and are scaled to start at one (excluding where no harvest is recorded as zero).
Intercept	A real number that is the intercept of the farm CCS relative to the mill at the start of the season. This is provided by a statistical grouping method.
InterceptGroup	A string that describes a group of intercepts of farm CCS relative to the mill at the start of the season. This is provided by a statistical grouping method.
MaxShift	An integer number that indicates the maximum number of shifts (per day) that is possible in a transport area.
MillCapacity	A real number that indicates the capacity of the mill (in tonnes per hour) in a fortnight (see Fortnight). This is provided by the mill.
MillID	An integer number that indicates which mill the data in the record is from (linked with the Mill table).
MillName	A string (text value) that is the name of the mill in this record.
OptimalFortnight	An integer that indicates the fortnight that the paddock should be harvested in (as calculated by the CSO model).
Option	An integer that indicates which option the data in the record is from (eg. 111, 511, 541).
OutsideStopsRail	A real number that gives the percentage of outside rail stoppages for a mill in a fortnight. That is, $\%\text{ Outside Stops} = \frac{\text{Hours not crushing due to rail stoppages}}{\text{Total available crush time in hours}}$.
OutsideStopsRoad	A real number that gives the percentage of outside road stoppages for a mill in a fortnight. That is, $\%\text{ Outside Stops} = \frac{\text{Hours not crushing due to road stoppages}}{\text{Total available crush time in hours}}$.
Paddock	An integer number that indicates which paddock or sub-block within a block the data in the record is from.
Parameter	A string that is the name of the parameter in this record.
Percent	An integer that gives the percentage of a farm that is to be cut in the round described by the record (eg 50 for half of the farm). The number must be between 0 and 100 inclusive. The default is 100 when Round = 1.
Period 1	A real number that is the optimal proportion of tonnes to be cut in the first period. A period is equal to two fortnights, so the first period is the first and second fortnight of the harvesting season.
Period 2	A real number that is the optimal proportion of tonnes to be cut in the second period. A period is equal to two fortnights, so the second period is the third and forth fortnight of the harvesting season.
Period 3	A real number that is the optimal proportion of tonnes to be cut in the third period. A period is equal to two fortnights, so the third period is the fifth and sixth fortnight of the harvesting season.
Period 4	A real number that is the optimal proportion of tonnes to be cut in the forth period. A period is equal to two fortnights, so the forth period is the seventh and eighth fortnight of the harvesting season.
Period 5	A real number that is the optimal proportion of tonnes to be cut in the fifth period. A period is equal to two fortnights, so the fifth period is the ninth and tenth fortnight of the harvesting season.

Period 6	A real number that is the optimal proportion of tonnes to be cut in the sixth period. A period is equal to two fortnights, so the sixth period is the eleventh and twelfth fortnight of the harvesting season.
Plant	An integer that indicates the plant type (only where crop class is of type plant) of the paddock referred to in the record. The plant types are represented by 0 to indicate an unknown or undetermined type 1 to indicate a crop planted after fallow 2 to indicate a crop planted after plough out
PV_Intercept	A real number that is the probability value associated with the significance of the intercept. < 0.1 implies intercept significantly different from 0. This is provided by a statistical grouping method.
PV_Slope	A real number that is the probability value associated with the significance of the slope. < 0.1 implies slope significantly different from 0. This is provided by a statistical grouping method.
Records	An integer that indicates the number of records over the years 1993-1999 from which the farm CCS group is computed. Less than 30 implies that the result is not significant.
Region	A string (text value) that is the name of the region the mill is in (eg Mackay, Burdekin).
Round	An integer that indicates the round for this record (eg 1 for the first round). A round with a zero percentage does not need to be stored in the round table. The default is 1.
Season	A four digit integer that indicates which season of harvest the data in the record is from (eg 1994).
selected	A boolean value that is used by HarvSched to indicate farms selected by the user for output.
Slope	A real number that is the slope of the farm CCS relative to the mill across harvest weeks. This is provided by a statistical grouping method.
SlopeGroup	A string that describes a group of slopes of farm CCS relative to the mill across harvest weeks. This is provided by a statistical grouping method.
Split	An integer that indicates which location on a farm the data in the record is from. In Mackay, this field can be used to distinguish between separate locations owned by the same farmer (ie. given the same farm number). Note that when a farm is not split, a zero should be used in this field.
TableName	A string (text value) that is the name of a DWIM table referred to by this record.
TotalStops	A real number that gives the percentage of total stoppages for the mill in this fortnight. That is, $\% \text{ Total Stops} = \frac{\text{Hours not crushing due to problem (eg. maintenance)}}{\text{Total available crush time in hours}}$. Note that this includes stops due to wet weather and breakdowns.
Tonnes	A real number that gives the tonnes of cane harvested from the paddock referred to in the record.
TonnesEstimate	A real number that gives the pre-season estimated tonnes of cane to be harvested from the paddock referred to in the record.
TonnesEstimate Model	A real number that gives the model pre-season estimated tonnes of cane to be harvested from the paddock referred to in the record.
Transport	An integer number that indicates which transport area within a mill the data in the record is from. This is as provided by the mill.

TransportCapacity	A real that is the daily transport capacity (in tonnes) of a transport area.
TransportName	A string (text value) that is the name of the transport area in this record (as industry provides).
Type[6]	A string (text value) that gives the type of code that the record refers to (eg Class).
Type[7]	A string (text value) that gives the type of parameter that the record refers to (eg CcsCoefficient).
User	A string (text value) that is the code used by user (or mill) for the DWIM code in the record.
Value	A real (double precision) that is the value of the parameter (specified by the parameter name, row and column) in the record.
Variety	A string (text value) that indicates which variety is planted in the paddock referred to in the record. The following varieties are represented by CP44 to indicate a CP44 variety CP51 to indicate a CP51 variety Exp to indicate experimental varieties Fallow to indicate a paddock that is fallow H56 to indicate a H56 variety Mixed to indicate a paddock that has more than one variety NCo to indicate a NCo variety Other to indicate another variety Pindar to indicate a Pindar variety Q1 to indicate a Q1 variety ⋮ Q999 to indicate a Q999 variety
VarietyCode	An integer value that is the code used by the CSO model to represent this variety (take Q off; take H off; take CP off; rest become 2?).
VarietyMajor	A boolean value that indicates whether the variety in the record is a major variety. This value is calculated.
Warning	A boolean value that indicates there is an error with the data in this record.
Week	An integer value that is used by the models to indicate harvest time. See notes on specific region for information on how this value is provided (either by the mills or computed from WeekStart, WeekMajor and WeekFinish values).
WeekBeginDate	A computed value that is the date of the beginning of the week in the Week field. Note that week 1 could start on a different date each year.
WeekPrevious	An integer number that indicates the week of harvesting of the paddock in the previous season. This value will normally need to be calculated using data provided by the region. Zero (?) will be used to indicate that the paddock could not be found in the previous season and hence the previous year's week of harvesting is unavailable. Crops that have been planted in the current year (i.e. Class = 1) will be given a WeekPrevious value of zero.

[6] For the UserCodes table.
[7] For the Parameters table.

4. Mill Data Format

This section discusses, in general terms, the format of the data we receive from the mills. Important features, with regards to the standard, are noted for future reference. Detail on mill codes can be found in Appendix B.

4.1 Mackay

Mackay block productivity data comes from two different sources, a paddocks file and rake files. Both sources contain paddock identifiers (farm, split, block and paddock numbers form a paddock identifier) for each record that are used to link data. A rake record is used when it can be matched to a record in the paddocks' file using the paddock identifiers. The following table indicates the source of each field of data for the productivity table in DWIM.

Paddocks' File	District, Farm, Split, Block, Paddock, Variety, Class, TonnesEstimate, Area, Tonnes, WeekStart, WeekFinish
Rake Files	Ccs, WeekMajor

4.2 Mossman

Mossman mill data is received in Excel spreadsheets. Their completed block productivity data has, in the past, been sourced from the one file. However, in the last couple of years, district information for farms has been missing from the records in this file. The district field for these years was filled from the previous years information, were possible. However, new farms have appeared in this time for which the district was unknown (given a zero in the data). In 2000 we were provided with farm level data (in an Excel spreadsheet). This contained, among other information, the district that each farm belonged to in 2000. We were thus able to retrospectively replace many of the zero districts in the historical block productivity table. The farm level data (known to Mossman as Harvesting Group information) should be requested in the future. Note that some farms appear to have been reassigned to a different district some time in 1996-1997.

Prior to 1999 Mossman did not record a paddock (sub-block) number for records. However, blocks did have divisions and these were treated in one of two ways in the data. Either, more than one record would refer to the same block but the variety and/or the crop classes were different for each record OR the block's variety was denoted as mixed. It has not been confirmed that all records with a mixed variety denote a block with divisions (ie paddocks). It does not appear that the idea of paddocks has been fully implemented in Mossman yet. That is, there are cases where more than one record still refers to a block.

While fallow paddocks appear in the initial conditions data (see table BlockProd_Current) and field book data, they have not been included in the completed block productivity data received from the mill at the end of the season. Should the fallow paddocks from the initial conditions be added to the historical block productivity table along with the completed block productivity data at the end of each season? Experience would indicate that this might not be wise due to the higher occurrence of errors in the initial condition's data. Do the models require fallow paddocks (apart from the calculation of the plant field)? Can the DWIM database be used with the GIS data without the fallow paddocks?

The variety and class codes have varied slightly over the years in the completed block productivity data. They also vary from those codes used in the field book data. These should be verified with mill staff upon analysis of received data. For plant crops, the plant field is traditionally supplied as part of the class field. That is, two codes represent plant crops in the mill's class field. One code indicates that the plant crop was planted after a fallow year and the other indicates that the plant crop was planted after plough out (possibly with a very short fallow period of a month or two).

The method used to indicate harvest time has varied over the years in the completed block productivity data. When data is received from the mill, check that the harvest time fields are self-explanatory. If they are not, verify their meaning with the mill. I suspect that Mossman stores harvest time as dates but have been supplying us with weeks in some cases. If weeks are supplied, a date for the beginning of the first week will need to be obtained.

5. Error Checking

This section discusses problems that could occur with the data. It does not deal with the process of converting from the mill code to the standard code.

5.1 Nulls or Missing Values

The DWIM databases shall contain complete data with no missing values for those fields being used by the models. Therefore, it will be necessary to check the imported data immediately for missing values (nulls, empty cells or zeros where not appropriate) and replace these with appropriate values. It may be necessary to contact the mill to obtain a replacement for the missing value. When a suitable replacement can not be found, the record may need to be excluded from the model input.

5.2 Season

Check that all records are four digit numbers. If not, a decision will need to be made, after viewing the values present in the data, on how to rectify the problem.

5.3 District and Farms

Check that a farm-split stays in the same district through time (ie. from season to season and within a season). Check that the districts present in the data are valid districts (ie the district is present in the District table). Check that the farm-splits in the data are valid and present in the farm table (if this data has also been imported).

5.4 Fallow Paddocks – Class versus Variety

As fallow paddocks can be recorded in either the Variety or Class fields, a check needs to be made to see that these are consistent for each record. That is, that all the records with a fallow in the Variety field also have a 10 in the class field and vice versa. The decision on what to do when there isn't consistency may depend on what region the data is from. For example, a region that does not have a fallow indicator in the Class field would need to be modified to do so.

Check that fallow paddocks have no results. That is, check that fallow paddocks have Ccs, Week and Tonnes equal to zero. Non-zero results attributed to fallow paddocks may indicate a problem with the records crop class classification. Refer back to data source (ie mills) when this occurs. Note that fallow paddocks can have a non-zero area.

5.5 CCS

It is possible that Ccs equals zero when this data has not been able to be found for a particular paddock (eg. Mackay data comes from two separate data sources, rake and paddock information, which may not be able to be matched). It is also possible that Ccs is greater than zero when Tonnes is zero, in Mackay data (see section 6.1). Do these go into the model as input?

5.6 Weeks

For each record, check that WeekStart ≤ WeekMajor ≤ WeekFinish when the mill region (eg Mackay) provides more than one 'week of harvest' field.

6. Uncorrectable Problems

6.1 Mackay

For 1992, 1993 and some of 1994 data, paddocks were not recorded on the rakes. This data is unreliable at paddock level.

As the Ccs value is obtained from the rakes data files, it is not always possible to match all paddocks with a Ccs value in these files. That is, some Ccs values are zero even though the paddock was harvested. The Ccs for each paddock i is calculated as

$$\text{Ccs}_i = \frac{\sum_{j=1}^{n} \text{Ccs}_{i,j} \times \text{Tonnes}_{i,j}}{\sum_{j=1}^{n} \text{Tonnes}_{i,j}},$$

where $\text{Ccs}_{i,j}$ is the jth record in the rake files that matches paddock i and n is the number of records that match paddock i in the rake file. If no matches occur for paddock i then $\text{Ccs}_i = 0.0$.

As the data in the rakes file is "rawer" than that in the paddock's file, it is possible that a value for Ccs, TonnesRakes, etc exists when Tonnes is zero. However, Ccs is still valid at mill level (this cane was cut and did receive the given ccs value) and are thus retained in the data. Care will need to be taken that these values are not used at inappropriate times (eg at farm level).

Several problems with the time of harvest indicators (WeekStart, WeekMajor and WeekFinish) were detected. These problems were either with numbers not making sense (eg WeekMajor outside the range of WeekStart and WeekFinish) or some of the numbers were missing altogether (eg no WeekFinish value). The Week field is included to provide a single field that can be used to provide a time of harvest indicator and hence is computed using the other three fields.

6.2 Maryborough

There are five records with a district of zero in the historical block productivity table. We think these are new farms but are unsure. They are all fallow paddocks. In the 2000 data, these farms are in district 3.

It appears that block 99 has been used by Maryborough to separate one farm from another in the irrigation data. This data has been imported into the paddock initial conditions table as is. Have not yet verified with Maryborough if this is true.

One of the paddocks in farm 4701 has an irrigation source code of 1 which does not exist in the irrigation source code list. Have not checked with Maryborough how or if this can be replaced.

In the 2000 final block productivity data, 140 records have an area of zero but have cut cane and received a ccs value. This is an ongoing consignment problem. We have been advised to ignore these records (from Glyn Peatey, Maryborough Sugar Factory).

In the 1991-1999 historical block productivity data, 54 records have Ccs greater than zero for fallow crop. This may indicate that the crop class indicator is incorrect.

6.3 Mossman

Current data in Mossman DWIM is based on a fixed season start of the 15th June (no matter what year). We have set this season start date, not Mossman. Occasionally, Mossman has started the season before this date which means that there are records with a zero in the Week field that have had cane harvested (ie Tonnes > 0). Fallow paddocks are also given a zero in the Week field. Be careful to always use the Class or Variety field to indentify fallow paddocks, not the Week or Tonnes field.

APPENDIX A – Glossary

Assignment	The recording of paddock numbers, crop class, variety, etc by the harvester on the bin at the time of harvesting.
CCS:	Commercial cane sugar.
Cell:	One element or rectangle of a database table. The variety of the 5th record is an example of a value in a cell.
DWIM:	A name used to refer to the collection of regional databases, such as DWIM_Mackay, used by the Whole of Industry models. Also may be used to refer to the common structure that underpins these regional databases.
Field:	A column in a database table or query.
Mill:	The area or data processed by a single mill (eg Farleigh, Mossman, or Tully).
Mill Codes:	The set of codes used by a mill to represent the data. For example, a mill may use “1R” to indicate that a paddock contains a first ratoon crop. It may be that all mills in a region use the same codes but the importing routines at this stage will be for a mill’s data not a region’s data.
Record:	A row in a database table or query.

Region: One or more mills that work in cooperation (eg the Mackay region has mills Farleigh, Marian, Pleystowe and Racecourse).

Standard Codes: The standard set of codes used by the models to interpret the data from all mills. For example, the plant crop indicator for a paddock varies but can be “Plant” or “0” depending on which mill the data is from. The standard code indicating that a paddock contains a plant crop is the integer number 1. Thus the need to transform the data from a mill as it is imported into the database.

APPENDIX B – Mill Codes

Only those fields where the mill codes differ from the standard codes shall be listed.

B.1 Farleigh, Marian, Pleystowe and Racecourse Mills

<table>
<tr><th>Field</th><th>Standard Code</th><th>Mill Code</th></tr>
<tr><td rowspan="4">Split</td><td>0</td><td><Null></td></tr>
<tr><td>1</td><td>A</td></tr>
<tr><td>2</td><td>B</td></tr>
<tr><td>etc</td><td>etc</td></tr>
<tr><td rowspan="9">Variety</td><td>CP44</td><td>44</td></tr>
<tr><td>Exp</td><td>15</td></tr>
<tr><td>Fallow</td><td>0 (???), 10 and 90</td></tr>
<tr><td>H56</td><td>56</td></tr>
<tr><td>Mixed</td><td>20, 21, 22, 23, 24, 25 and 26</td></tr>
<tr><td>NCo</td><td>31</td></tr>
<tr><td>Other</td><td></td></tr>
<tr><td>Pindar</td><td>40</td></tr>
<tr><td>Q???</td><td>???</td></tr>
<tr><td rowspan="9">Class</td><td>1</td><td>1</td></tr>
<tr><td>2</td><td>2</td></tr>
<tr><td>3</td><td>3</td></tr>
<tr><td>4</td><td>4</td></tr>
<tr><td>5</td><td>5</td></tr>
<tr><td>6</td><td>6</td></tr>
<tr><td>7</td><td>7, 8, 9 and 10</td></tr>
<tr><td>9</td><td>11, 12, 13, 14, 15, 16, 17, 18, 19 and 20</td></tr>
<tr><td>10</td><td>0</td></tr>
<tr><td rowspan="2">Plant</td><td>1</td><td><Null> for class equal to 1</td></tr>
<tr><td>2</td><td>R for class equal to 1</td></tr>
<tr><td>Week</td><td colspan="2">Note that week 1 is as defined by Mackay mill (refer to Geoff Flemming).

<table>
<tr><th colspan="3">Values Available</th><th></th></tr>
<tr><th>Start</th><th>Major</th><th>Finish</th><th>Value used in Week</th></tr>
<tr><td>Y</td><td>Y</td><td>Y</td><td>f(Start,Major,Finish)</td></tr>
<tr><td>Y</td><td>Y</td><td>.</td><td>Start</td></tr>
<tr><td>Y</td><td>.</td><td>Y</td><td>Start</td></tr>
<tr><td>Y</td><td>.</td><td>.</td><td>Start</td></tr>
<tr><td>.</td><td>Y</td><td>Y</td><td>Major</td></tr>
<tr><td>.</td><td>Y</td><td>.</td><td>Major</td></tr>
<tr><td>.</td><td>.</td><td>Y</td><td>Finish</td></tr>
<tr><td>.</td><td>.</td><td>.</td><td>0</td></tr>
</table>
where Start = WeekStart, Major = WeekMajor, Finish = WeekFinish and

$$f(\text{Start},\text{Major},\text{Finish}) = \begin{cases} \text{Start} & \text{Finish} - \text{Start} \le 2 \\ \text{Major} & \text{Finish} - \text{Start} > 2 \text{ and Start} \le \text{Major} \le \text{Finish} \\ \text{Start} & \text{Otherwise} \end{cases}$$</td></tr>
</table>

B.2 Maryborough Mill

Field	Standard Code	Mill Code
Split	0	Not available.
Paddock	1	No letter on block (eg. 34)
	1	Letter “A” on block (eg 34A)
	2	Letter “B” on block (eg 34B)
	etc	etc
Variety	CP44	44
	CP51	51
	Exp	21, 77
	Fallow	0
	H56	56
	Mixed	19
	NCo	
	Other	
	Pindar	
	Q???	???
Class	1	0
	2	1
	3	2
	4	3
	5	4
	6	5
	7	6, 7
	9	8
	10	9
Plant	0	Not available from Maryborough Sugar Factory
Week		

B.3 Mossman Mill

Field	Standard Code	Mill Code
Split	0	Not available.
Paddock	1	0
	1	A
	2	B
	etc	etc
Variety	CP44	
	Exp	633
	Fallow	369
	H56	56
	Mixed	999
	NCo	
	Other	334
	Pindar	123
	Q???	???
Class	1	1 and 2 or PLANT and REPLANT
	2	11 or 1R
	3	12 or 2R
	4	13 or 3R
	5	14 or 4R
	6	15 or 5R
	7	16, 17, 18 and 19 or 6R, 7R, etc and OR
	9	
	10	40 or FAL
Plant	0	Class not equal to 1 or 2
	1	Class = 1 or PLANT
	2	Class = 2 or REPLANT
Complete	1	Y
	2	N
Week	Week = FLOOR((Date – 15/06/Season)/7) + 1	

APPENDIX C – Database Tables and Their Fields

C.1 Paddock Initial Conditions Table

Name: BlockProd_Current

Fields: MillID, Season, District, Farm, Split, Block, Paddock, Variety, Class, Plant, Area, TonnesEstimate, TonnesEstimateModel, WeekPrevious, HDPY, DatePlanted, OptimalFortnight, Warning

Status: Rewritten every year before the harvest using data supplied by the mills. Fields HDPY and WeekPrevious will usually need to be calculated. Field DatePlanted is used by HarvSched to store planted dates as the user enters them. Field OptimalFortnight will contain results from CSO model when run using a 1 year planning horizon.

C.2 Historical Block Productivity Table

Name: BlockProd_Historical

Fields: MillID, Season, District, Farm, Split, Block, Paddock, Variety, Class, Plant, Area, Tonnes, TonnesEstimate, Ccs, Week, WeekBeginDate, Warning

Status: Updated every year before the harvest using data supplied by the mills. Field WeekBeginDate will usually need to be calculated.

C.3 CCS Group Table

Name: CCSGroup

Fields: CCSGroup, InterceptGroup, SlopeGroup, Description

Status: Static across all mills and mill regions (in this version).

C.4 Crop Class Codes Table

Name: Class

Fields: ClassCode, Class

Status: Static across all mills and mill regions.

C.5 CSO Model Constants Table

Name: CSOModelConstants

Fields: MillID, ConstName, ConstValue, ConstType, ArrayElement

Status: Checked before model run (a few values will be altered by interface).

C.6 District or Zone Table

Name: District

Fields: MillID, Season, District, DistrictName

Status: Checked when data from the mills is imported or updated.

C.7 Farm Table

Name: Farm

Fields: Farm, Split, FarmName, MillID, Season, District, Group, CCSGroup, Transport, selected, Warning

Status: Rewritten every year before the harvest using data supplied by the mills and a statistical grouping method.

C.8 FarmCCSGroup Table

Name: FarmCCSGroup

Fields: Farm, Split, Season, CCSGroup, Intercept, Slope, PV_Intercept, PV_Slope, Records

Status: Calculated once (in this version) using a statistical grouping method.

C.9 Field Book Table

Name: FieldBook

Fields: MillID, Season, Farm, Split, Block, Paddock, Variety, Area, Tonnes, Complete, Week, Warning

Status: Rewritten every year during the harvesting season (possibly more than once) using data supplied by the mills.

C.10 Field Name Table

Name: FieldNames

Fields: TableName, FieldName, FieldNameMill

Status: Checked when data is imported or updated.

C.11 Mill Capacity Table

Name: Fortnight

Fields: MillID, Season, Fortnight, MillCapacity, TotalStops, OutsideStopsRail, OutsideStopsRoad

Status: Rewritten every year before the harvest using data supplied by the mill.

C.12 Harvesting Group Table

Name: Group

Fields: MillID, Season, Group, GroupName

Status: Rewritten every year before the harvest using data supplied by the mill.

C.13 Level 1 Results Table

Name: Level1Options

Fields: Option, MillID, Group, CCSGroup, FarmNos, Period 1, Period 2, Period 3, Period 4, Period 5, Period 6

Status: Output from cane supply optimisation model and used as input for HarvSched. Rewritten each time the cane supply optimisation model is run.

C.14 Level 2 Results Table

Name: Level2Options

Fields: Option, MillID, Group, CCSGroup, FarmNos, Period 1, Period 2, Period 3, Period 4, Period 5, Period 6

Status: Output from cane supply optimisation model and used as input for HarvSched. Rewritten each time the cane supply optimisation model is run.

C.15 Level 3 Results Table

Name: Level3Options

Fields: Option, MillID, CCSGroup, Cclass, Variety, Period1, Period2, Period3, Period4, Period5, Period6

Status: Output from cane supply optimisation model and used as input for HarvSched. Rewritten each time the cane supply optimisation model is run.

C.16 Mill Table

Name: Mill

Fields: MillID, MillName, Region, Company, Contact

Status: Created when the region's DWIM is formed. Updated when information, such as the contact, changes.

C.17 Parameters Table

Name: Parameters

Fields: MillID, Type, Parameter, Code1, Code2, Value

Status: Manual update prior to season?

C.18 Rounds Table

Name: Round

Fields: MillID, Season, Farm, Split, Round, Percent

Status: Rewritten every year before the harvest using data supplied by the mill.

C.19 Transport Area Table

Name: Transport

Fields: MillID, Season, Transport, TransportName, TransportCapacity, MaxShift, CaneShift

Status: Checked when data is imported or updated.

C.20 User Codes Table

Name: UserCodes

Fields: Type, Dwim, User, Description

Status: Checked when data is imported or updated.

C.21 Variety Major Table

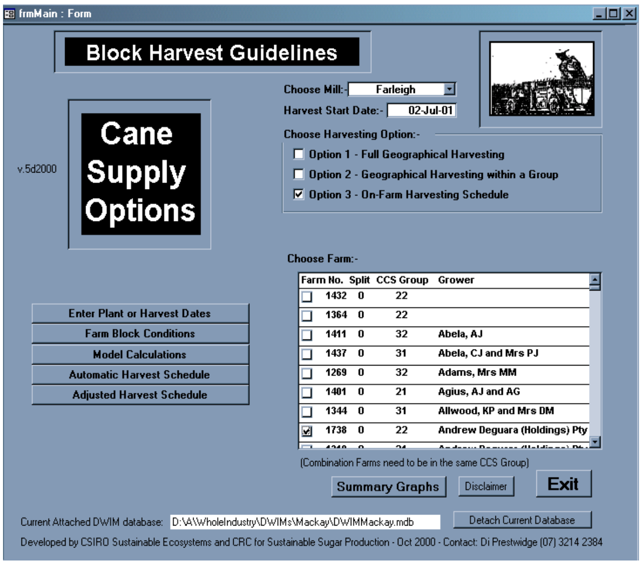
Name: VarietyCodes

Fields: Variety, VarietyMajor, VarietyCode

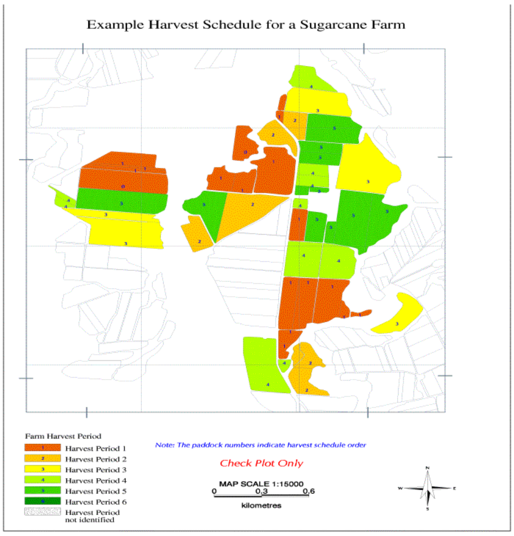
Status: Rewritten every year before the harvest using data supplied by the mill.

APPENDIX E.

Screen shots of HarvSched



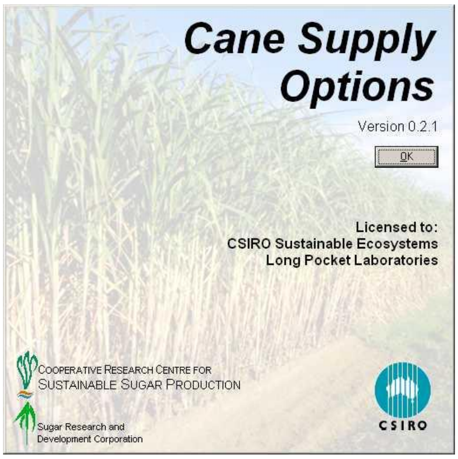
Farm Split	Block	Paddock	Class	Variety	Harvest Fortnight 2000	Estimate	Cum.	% Cum.	Round	Period	Harvested ?	CCS		Area	Grower Payment	Industry Payment	Start Date	Finish Date
1738-0	2	4	6R+	Q124	1	192	192	9%	1	1		12.0	12.1	2.4	2110	3390		
1738-0	7	2	Plant	Q121	0	369	561	25%	2	3		13.8	12.6	4.6	2200	3840		
1738-0	7	1	Plant	Q138	0	365	925	41%	3	3		13.3	12.1	4.6	2080	3690		
1738-0	2	3	1R	Q121	8	239	1164	52%	4	3		13.5	12.3	3.1	2060	3620		
1738-0	6	1	1R	Q124	4	170	1335	59%	4	3		13.5	12.4	2.2	2110	3700		
1738-0	2	2	2R	Q124	1	176	1511	67%	4	3		13.4	12.2	2.4	1920	3390		
1738-0	11	1	Plant	Mixed	0	80	1591	71%	4	3		13.6	12.4	0.9	2280	4000		
1738-0	9	1	6R+	Q138	4	188	1779	79%	5	3		12.5	11.3	2.4	1850	3420		
1738-0	2	1	6R+	Q124	1	249	2028	90%	5	4		13.7	11.9	3.1	2060	3880		
1738-0	8	2	4R	Q124	4	127	2154	96%	6	5		13.9	12.0	1.9	1700	3190		
1738-0	8	1	4R	Q124	4	98	2252	100%	6	6		13.0	12.2	1.5	1740	2990		
			Total Estimate			2252							Total Area	29.21	2030	Av.$/ha:		

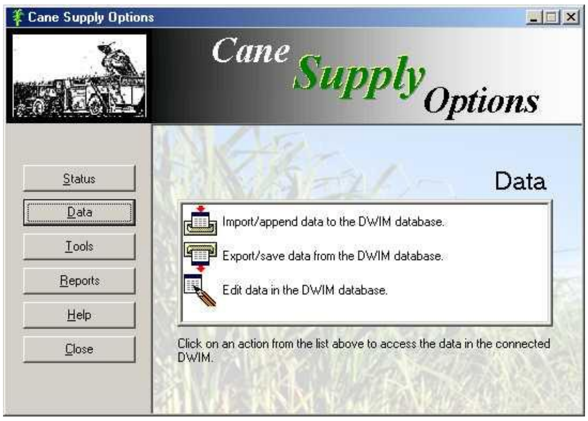


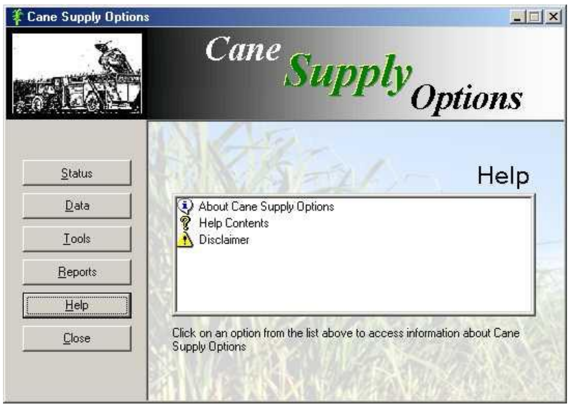
APPENDIX F.

Screen shots of the user-interface and data importing routines

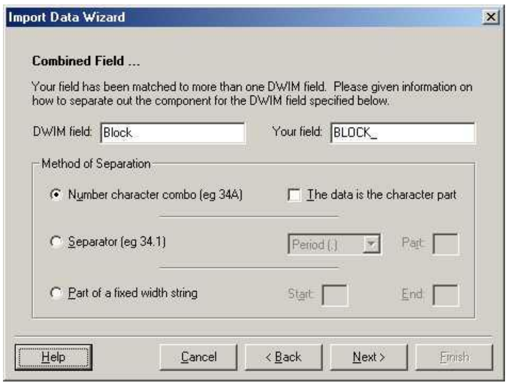
User Interface

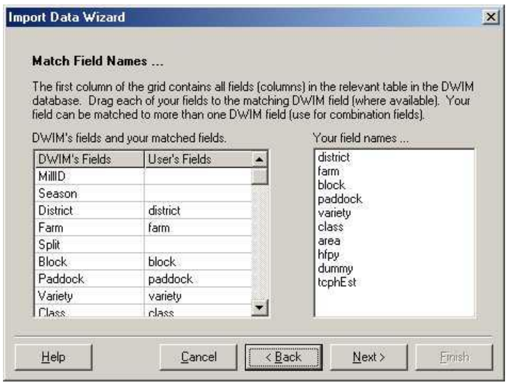


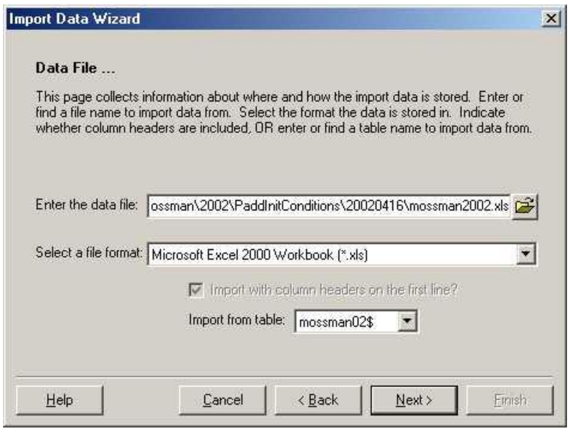


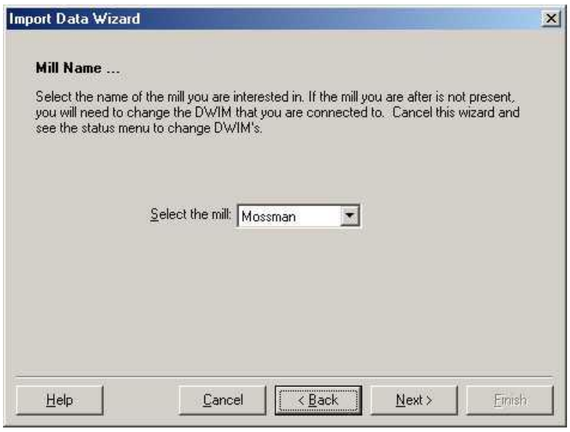


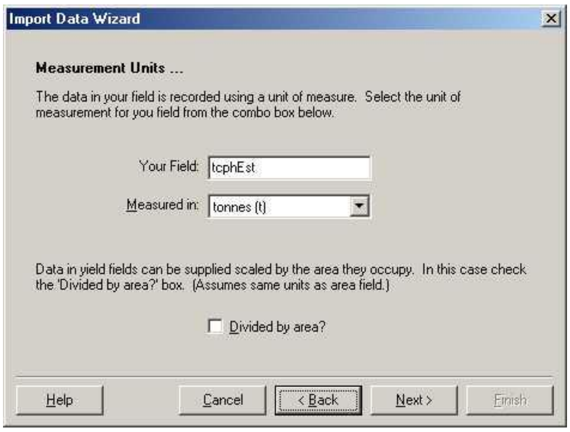
Importing data



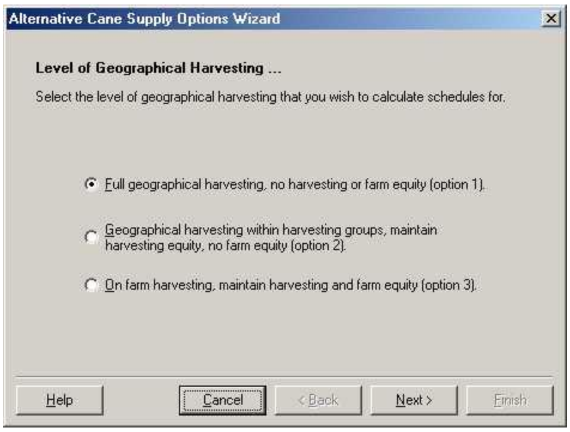


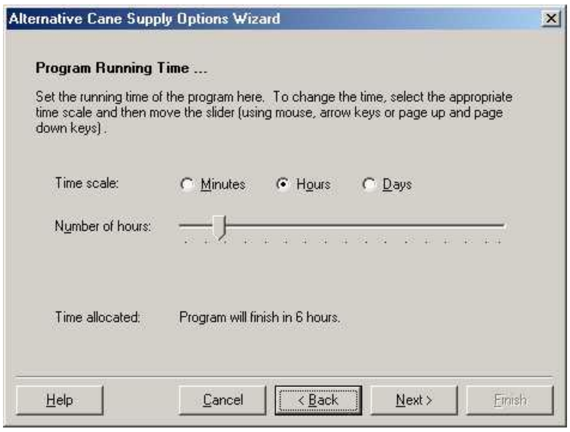


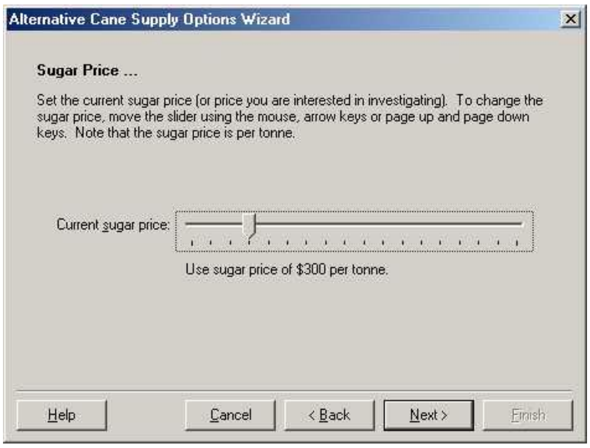


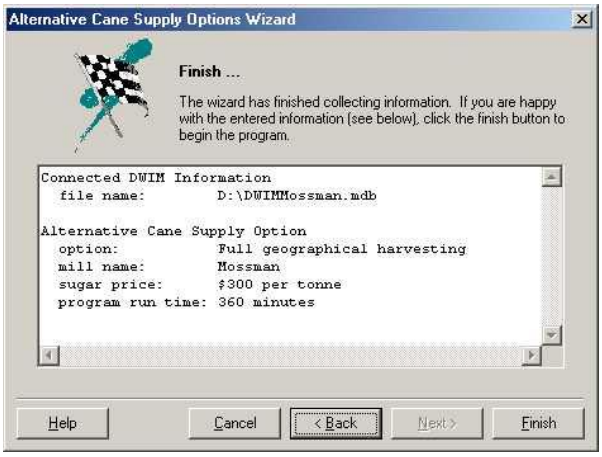


Running the model






Alternative Cane Supply Options Wizard
Sugar Price ...
Set the current sugar price (or price you are interested in investigating). To change the sugar price, move the slider using the mouse, arrow keys or page up and page down keys. Note that the sugar price is per tonne.
Current sugar price:
Use sugar price of $300 per tonne.
Help
Cancel
< Back
Next >
Finish


Alternative Cane Supply Options Wizard
Finish ...
The wizard has finished collecting information. If you are happy with the entered information (see below), click the finish button to begin the program.
Connected DWIM Information
file name: D:\DWIMMossman.mdb
Alternative Cane Supply Option
option: Full geographical harvesting
mill name: Mossman
sugar price: $300 per tonne
program run time: 360 minutes
Help
Cancel
< Back
Next >
Finish

APPENDIX G.

Deriving the Statistical Models for Relative CCS

Michele Haynes
Coastal Cropping Systems
CSIRO Sustainable Ecosystems
9 April 2002

Project: Cane Supply Options Analysis Project (CSOAP).
See Andrew Higgins (Principal Investigator) for details of the Project and the Cane Supply Optimisation (CSO) Model.

Use: To predict paddock level CCS relative to the mill average for the approaching season, based on historical block productivity data, for input to the cane supply optimisation model.

Application:
Sugar regions of Mossman, Mackay (Farleigh, Marian, Pleystowe, Racecourse), Maryborough and The Burdekin (Invicta, Pioneer).

References for Modelling Approach:

Higgins, A. and Haynes, M. (2001) An integrated modelling approach to enhancing sugarcane profitability, *Proc. of the International Modelling and Simulation Congress*, Australian National University, Canberra, 2001.

Haynes, M. and Prestwidge, D. (2001) A decision support system for generating improved sugarcane harvest schedules. *Proc. of the International Modelling and Simulation Congress*, Australian National University, Canberra, 2001

Haynes, M. (2002) Modelling Sugarcane Productivity – CCS and Cane Yield. In preparation

Haynes, M. and Murray, A. (2002) Spatial Patterns in Sugarcane Productivity. In preparation.

1. Data Storage and Software

Historical block productivity data (BPD), recorded by the sugar mills, has been forwarded to the CSOAP research team within the CSIRO Coastal Cropping Systems group, with permission for its use in this project.

The format of the BPD can differ substantially among the sugar mills. Therefore, it has been necessary to design and develop an ACCESS 2000 database called DWIM (Database for Whole of Industry Management) to standardise and store the historical BPD for each of the sugar mill regions participating in the project. A separate DWIM has been created for each mill area.

DWIM also stores other relevant data including farm paddock initial conditions for the approaching season and output from the modelling procedures. Details on the structure and content of DWIM can be found in the latest version of a document called DWIM.doc. The date on which this document was last updated usually proceeds this filename e.g. 20020308-DWIM.doc. This document resides in a directory called C3_Samantha/C3BCommon/June2002/DWIMs/Documentation on TAGNET. The actual DWIMs are accessed through the same directory within a sub-directory named as the sugar region for which the data is stored e.g. Mackay.

To develop a statistical model for relative CCS, the appropriate historical data is extracted from the relevant DWIM. Prior to April 2002, historical BPD has been stored in DWIMs for the Mossman, Mackay and Maryborough mill regions. Statistical models for relative CCS have been developed for the 6 mills in these regions and the farm CCS groups and model coefficients stored in the DWIMs.

For the 2002 harvest season, models for relative CCS are also required for the 2 Burdekin mill regions of Pioneer and Invicta. A DWIM has not yet been constructed for these mill regions and so the data has been extracted directly from EXCEL files provided by the mills.

To develop the statistical models, the relevant data fields are exported from the DWIM (or EXCEL file) to an EXCEL file. This file is then imported to the statistical software S-PLUS (Version 6 for Windows) for subsequent data manipulation, analysis and model development.

2. Importing Data and Manipulation in S-PLUS

2.1 Variables and records to import from DWIM

The following variables (columns) should be imported into S-PLUS from DWIM:

Season, District, Farm, Split, Block, Paddock, Variety, Class, Area, Tonnes, Ccs, Week.

The seasons for which historical BPD is recorded, varies from mill to mill. It is also important to have some knowledge of the accuracy of the historical data in determining which seasons should be included for modelling purposes. For example, for Mackay it was decided to include data for the seasons 1995 to 2000 as it was not considered appropriate to include data prior to 1995. For Mossman and Maryborough, data from 1992 to 1999 were included as early data was considered to be reasonably accurate, although 2000 data was not included because of the unusual climatic conditions throughout the harvesting and growing seasons. As new seasonal BPD becomes available, it may be added to the dataset and included in the modelling procedures to update the model.

2.2 Preparation of data for modelling

The response variable in these models is CCS. If the value of CCS is zero then this corresponds to a fallow paddock with no crop harvested or a missing value. Therefore, all records containing CCS=0 should be removed from the data.

For the purposes of the CSOAP we are interested in modelling the linear trend in a farm's average CCS relative to the mill average CCS, with harvest week. For a given harvest week, the farm's average CCS is computed by averaging over paddock CCS values. So before averaging over paddocks harvested, the effects of factors associated with the paddock, such as crop class and variety, must be removed from the CCS response data. For this reason, the categories representing the levels of the factors of crop class and variety, must be carefully coded to ensure that model is sensible and that sufficient data is available for each season, week, crop class and variety combination.

Crop class

It can been shown in exploratory analyses that the effect of crop class on CCS declines from plant crop down to 2^{nd} or 3^{rd} ratoon. Generally, the difference in the effects of 3^{rd} ratoon and beyond is not detectable or important and therefore it is appropriate to consider all crop classes of 3^{rd} ratoon or greater, as one category. If standover cane is recorded in the data then it is important to denote this class by a separate category.

Recode the variable **Class** into the variable **adjcclass** as follows:

Crop class	Class	adjcclass
Plant	1	1
1^{st} ratoon	2	2
2^{nd} ratoon	3	3
3^{rd} ratoon	4	4
4^{th} ratoon	5	4
.	.	.
.	.	.
nth ratoon	n+1	4
standover	9	9

Variety

The composition of varieties on a farm and over a whole mill region can vary quite dramatically from year to year. Also, varieties that were grown several years ago may have been superseded so that they are no longer harvested. Therefore, it is necessary to tabulate the number of records (indicating paddocks) or percentage of area harvested, by season and variety, to determine the major varieties that have been grown across the region. For varieties which contribute only a small percentage to the overall composition, it is appropriate to combine these varieties into one category. Generally, a variety has been classified separately as a major variety if it occurred more than 100 times in the previous year and as “other” if the crop is mixed or does not occur very often. However, this threshold value is subjective and may vary from mill to mill.

Recode the variable Variety into the variable vrcode, with the codes corresponding to major varieties remaining the same but with the code “other” representing all other varieties.

Note that the major varieties for each of the mill regions investigated to date, are indicated in the VarietyCodes table of the corresponding DWIM.

2.3 S-plus data frames and functions

Instructions

1. Copy the required DWIM into a directory on the hard-drive of your PC before attempting to extract data.

2. Import historical data from the BlockProd_Historical table in DWIM, into an S-plus data frame (saved in File Menu – Chapter). This can be achieved by:

a) Firstly saving the data directly from the Access table to an Excel spreadsheet and then importing this spreadsheet into an S-plus data frame through the **File-Import**

Data-From File menu. However, the number of records to be imported is often too great to be saved in an Excel file in which case the second method (b) can be used.

b) Importing the data directly into an S-plus data frame from the Access table, through an ODBC connection, using the menu commands **File-Import Data-From ODBC Connection** (Di Prestwidge or Yvette Everingham may be able to assist with this).

3. When importing the data, name the data frame **millbpd.df** (replace mill by the region name e.g. mary (Maryborough), moss (Mossman)). Unwanted columns may be deleted from the data frame at this stage if preferred.

4. Remove records from the data frame which contain CCS values of zero. To do this, highlight the column for Ccs in the data frame, and from the menu choose **Data – Subset.** In the window that appears, enter Ccs == 0 in the space for "Subset rows with: ". Save the new data frame to **posmill.df**.

5. In the data frame **posmill.df**, create the new variable columns for **adjcclass** and **vrcode**, by recoding the categories for Class and Variable as described in Section 2.2 above. You will need to refer to the DWIM documentation to understand what each of the levels of Class and Variety represent, before recoding them. For example in the raw data, Class=8 often refers to a plant crop after fallow and so must be recoded to adjcclass=1.

NOTE: If it is more convenient, the new variables adjcclass and vrcode could be created in DWIM before importing to S-plus.

6. For some mill areas, e.g. the Burdekin district, it may be necessary to recode district categories as well, if district numbers have been changed over the seasons. This could also be achieved in DWIM.

3. Modelling Paddock/Farm CCS relative to the Mill

3.1 Computing paddock CCS relative to the mill

The mill CCS trend across a harvest season is found by computing the average CCS from all paddocks harvested for each week within a harvest season. The average is weighted by the tonnes harvested from each paddock. For one season of BPD this trend can be computed through fitting a linear model to the CCS response with week specified as an independent categorical variables (or factor). The model is

$$Ccs = \mu_1 + week.fac_i + residual \qquad (1)$$

In model (1), every paddock CCS value is explained by the overall mean μ_1, a week effect and a residual. Therefore the residuals represent the difference between the paddock CCS and an average of CCS for all the paddocks harvested across the mill region within the same week. An average effect on CCS is computed for each week i, giving an average trend in mill region CCS across the harvest season.

But rather than consider the residuals in (1) as the paddock CCS relative to the mill average CCS, we should also remove any other effects that we suspect might be contributing to the paddock CCS value. Otherwise the residuals may be influenced by the type of crop as well as the week in which it is harvested.

From previous analyses we know that crop class and variety and their interactions with time of harvest (week), may have an effect on CCS response, and information about the levels of these factors corresponding to a paddock, are recorded in the BPD. Therefore we can also include these independent factors in the model to remove the corresponding average effects on CCS response. For i paddock records, the model becomes

$$Ccs_i = \mu_2 + week.fac_j + vrcode_k + adjcclass_l + week.fac_j{:}adjcclass_l +$$

$$week.fac_j{:}vrcode_k + adjcclass_l{:}vrcode_k + e_i \qquad (2)$$

Therefore the residuals in model (2), denoted by e_i, represent the difference between the paddock CCS and an average of CCS for all the paddocks harvested across the mill region within the same week but also with the average effects of variety, crop class and their interactions with week removed. We use these residuals to represent the paddock CCS relative to the mill average CCS.

3.2 Assessing linear trends in relative CCS with harvest week, for farms

For the purposes of exchanging harvesting entitlement, both between and within groups, we are interested in detecting a linear trend in relative CCS with harvest week for individual farms. This is to investigate whether some farms tend to have a higher average CCS early in the season, or a higher average CCS later in the season. The trend in relative CCS may not be strictly linear however the fitted model can be used to detect whether there is a significant change in average CCS from one end of the season to the other.

To model the linear trend we need to compute the residuals e_{iy} for each season y. A separate model of the form (2) will need to be fitted separately for each year represented in the BPD. Different factor effects will be computed for each year as there will be some variation in these estimated effects. The model becomes

$$Ccs_{iy} = \mu_y + week.fac_{jy} + vrcode_{ky} + adjcclass_{ly} + week.fac_{jy}{:}adjcclass_{ly} +$$

$$week.fac_{jy}{:}vrcode_{ky} + adjcclass_{ly}{:}vrcode_k + e_{iy} \qquad (3)$$

It is now assumed that for each year, the average trend in CCS with harvest week across the whole mill region, as well as the average crop class, variety and interaction effects, have been removed from the paddock CCS records. For each farm f, the residuals e_{iyf} across all years will now be treated as one data set. For a farm f, fit the linear model

$$e_{iyf} = a_f + b_f.Week + r_{iyf} \qquad (4)$$

where Week is the continuous variable representing the week in which the paddock was harvested, r_{iyf} are the residuals of the fit, a_f is the coefficient representing the average relative CCS value for the farm at the start of the season (intercept) and b_f is the coefficient representing the slope of the linear trend in average relative CCS for the farm. If a linear trend for a farm is consistent from year to year then this will show up as a strong significant slope effect.

3.3 Allocating farms to CCS Groups

The following 3 paragraphs are an extract from the paper by Higgins and Haynes (2001) and explain the formation of CCS groups. Note that the coefficient β_f corresponds to b_f in model (4) above.

"Farm relative CCS with harvest week was modelled as a simple linear trend for the purpose of detecting whether average relative CCS for a farm is consistently higher either early or late in the season. A more complex trend was not considered, due to the large amount of variation in the relative CCS data for paddocks. A farm *f* was considered to be a potential candidate for Options 1 and 2 if $\beta_f \leq -0.02$ or $\beta_f \geq 0.02$. These values correspond to a change of at least 0.4 units in CCS over a period of 20 weeks. This was regarded by our industry partners to be the minimum gain in relative CCS that a grower would consider to be worthwhile pursuing. Farms were grouped according to the criteria that ($b_f \leq -0.02$ or $b_f \geq 0.02$) and ($a_f < 0$ or $a_f > 0$) with a 10% level of significance. This resulted in the formation of nine groups of farms. For the three groups in which b_f was not significantly different from zero, there was no evidence that the farm average CCS followed a different trend in CCS to the mill. Hence, the grower would not be encouraged to participate in the cane supply arrangements of Options 1 and 2.

The significance of trends in farm relative CCS with harvest week, and the model's ability to estimate farm CCS accurately, varied considerably for a number of reasons. For example, large time intervals between visits by the harvester to a smaller farm may result in a sparsity of CCS entries for every harvest week in the block productivity data, even over six or more seasons. If the timing of the visits is not consistent over several seasons or if a strong linear trend was not present in the majority of seasons for which data is available, then the evidence of a linear trend with harvest week may be weak. Also, there may be variations in the data due to unknown sources that have not yet been accounted for in the model which, consequently, distort any trend with harvest week. For these reasons it was acknowledged that some farms will have been incorrectly allocated to groups, and hence estimation for these farms will not be as accurate as for others. However, for the purposes of encouraging participation in a pilot study, a conservative approach was taken in which farms were not allocated to a "high early" or "high late" group unless the evidence was reasonably strong.

Due to the variation in farm size and the unbalanced structure of data by variety, crop class and harvest week for a single farm, a linear model as represented by (7) was fitted to the relative CCS data for all farms within a group to provide an estimate of the average linear trend in farm relative CCS with harvest week for the group. The relative CCS for a paddock is then estimated by combining the average effects of variety and crop class with the linear trend by harvest week effect contributed by the farm (confounding location and management effects)."

3.3 Estimating paddock level CCS effects for optimisation model

Having allocated farms to 9 CCS groups, it is now necessary to estimate CCS values relative to the mill for farm paddocks within a CCS group. This is achieved by

- Removing the mill CCS trend with harvest week from the CCS response data, separately for each year. Here we assume that Week is constant and fit a quartic polynomial to the CCS in each year to remove a smooth trend.

  $$Ccs = a + Week + Week^2 + Week^3 + Week^4 + residual \qquad (5)$$

- Fitting a model similar to (2) to the residuals from (5) over all years to estimate overall average effects for crop class, variety and their interactions with (Week + $Week^2$, to allow some curvature in the interaction).
- To these effects add the estimated linear trend as defined by the CCS group to which the farm has been allocated.

Through this procedure we are not attempting to accurately predict the relative CCS for a paddock (as this is not possible with BPD only) but rather to provide a form of discrimination between the relative CCS values of paddocks and farms with different trends and at different harvest weeks.

3.4 S-plus functions for modelling approach and estimation

Instructions

1. To compute the residuals e_{iy} in model (3) representing paddock CCS relative to the mill average, use the S-plus function **millres.fun**. To save the residuals in the data frame pos**mill.df**, type the command

posmill.df$millccs.res <- millres.fun(posmill.df)

This will save the residuals in a column called **millccs.res** in the data frame **posmill.df**.

2. To fit simple linear models, as in (4), to these residuals separately by farm split number use the **lmList** command.

First you must create a column of characters in the data frame, called fmsplit, to indicate records associated with a particular farm split rather than just farm number. This can be done using the paste command as follows:

posmill.df$fmsplit <- paste(as.character(posmill.df$Farm),
as.character(posmill.df$Split)).

Then create a new data frame containing only the variables necessary for fitting the simple linear models:

lmlistmill.df <- data.frame(posmill.df$fmsplit, posmill.df$Week,
posmill.df$millccs.res)

names(lmlistmill.df)<-c("fmsplit", "Week", "millccs.res")

The modelling command is:

fmccs.lmlist <- lmList(millccs.res ~ Week|fmsplit, data=lmlistmill.df).

The results of the linear model fits for each farm split are now stored in the object **fmccs.lmlist**.

3. The function millslope.fun was at different stages of standardisation according to the format of the BPD input and the time available to generate the information. In the latest version for the Mossman example (found in directory My Documents/Packaged Computer Modelling/Example Computer Files/Mossman), the commands to generate the object fmccs.lmlist are contained in the function millslope.fun. To extract the linear trend coefficients and relevant statistics for each farm split, run the commands **millslope.df <- millslope.fun(posmill.df),** where millslope.fun is a function and millslope.df is a data frame containing the output from the function.

4. To allocate a CCS group to each farm split you need to create three additional columns in **millslope.df** called **intgrp**, **slopegrp** and **ccsgroup**. The column ccsgroup is derived from intgrp and slopegrp. The groupings are allocated according to the values intercept, pvint, slope, pvslope and records stored in this data frame.

Rules for creating ccsgroups

If intercept < 0 and pvint <= 0.1 then intgrp=1
If intercept > 0 and pvint <=0.1 then intgrp=3

If pvint > 0.1 then intercept is not significantly different from 0 so intgrp=2

If slope <= -0.2 and pvslope <= 0.1 then slopegrp=1
If slope >= 0.2 and pvslope <= 0.1 then slopegrp=3
If –0.2<slope<0.2 OR pvslope > 0.1 then slopegrp=2

But if the number of records is < 20 then we let intgrp=slopegrp=2 as we are conservatively assuming that there is not enough data to detect a trend. In general, the allocation of the value 2 indicates that there is no evidence that the trend in either the intercept or slope is significantly different from that of the mill trend.

This creates 9 intercept/slope or ccs groups denoted by

11 starts below mill and moves in a negative direction with harvest week
12 starts below mill and moves with the mill
13 starts below mill and moves in a positive direction with harvest week

21 starts level with mill and moves in a negative direction with harvest week
22 starts level with mill and moves with the mill
23 starts level with mill and moves in a positive direction with harvest week

31 starts above mill and moves in a negative direction with harvest week
32 starts above mill and moves with the mill
33 starts above mill and moves in a positive direction with harvest week

S-plus code

Populate both **intgrp** and **slopegrp** with the number 2,
e.g. **millslope.df$intgrp<-rep(2, length(millslope.df$fmsplit).**

```
millslope.df$intgrp[millslope.df$intercept<0 & millslope.df$pvint<=0.1] <- 1
millslope.df$intgrp[millslope.df$intercept>0 & millslope.df$pvint<=0.1] <- 3
millslope.df$intgrp[millslope.df$records<20] <- 2
```

```
millslope.df$slopegrp[millslope.df$slope<=-0.2 & millslope.df$pvslope<=0.1] <- 1
millslope.df$slopegrp[millslope.df$slope>=0.2 & millslope.df$pvslope<=0.1] <- 3
millslope.df$slopegrp[millslope.df$records<20] <- 2
```

Create column **ccsgroup**

```
millslope.df$ccsgroup<-as.factor(millslope.df$intgrp*10 + millslope.df$slopegrp)
```

Now that the **ccsgroup** column has been created you can add another column called ccstrend to identify whether the farm has a high early or high late relative ccs trend.

```
millslope.df$ccstrend<-rep("Neutral", length(millslope.df$fmsplit))
millslope.df$ccstrend[millslope.df$slopegrp==1]<-"Early"
millslope.df$ccstrend[millslope.df$slopegrp==3]<-"Late"
```

5. If a data frame containing harvest group numbers, farm splits and group names is available, then you can match this data through the fmsplit variable to create a harvest group name and number column in millslope.df.
e.g.

```
m <- match(millslope.df$fmsplit, harvgrp.df$fmsplit)
millslope.df$grpname<- harvgrp.df$grpname[m]
```

The data frame millslope.df can then be exported to an Excel file to be used for disseminating farm trend information to project groups and growers in participating mill regions.

6. Estimating CCS relative to the mill average for input to the optimisation model.

To fit a smooth trend with harvest week to the CCS data, run the function **millgrp.fun** to generate the residuals from model (5). Create a column of residuals, millgrp.res, in the data frame with the command

posmill.df$millgrp.res <- millgrp.fun(posmill.df).

Fit the following model to these residuals to estimate overall crop class, variety and their interaction with harvest week effects, after seasonal trends have been removed:

ccsfac.lm <- lm(millgrp.res ~ Week + Week^2 + vrcode + adjcclass + vrcode:(Week + Week^2) + adjcclass: (Week + Week^2) + vrcode:adjcclass, data=posmill.df, weights=Tonnes, na.action=na.exclude, singular.ok=T) (6)

Extract the residuals from this model fit in

posmill.df$ccsfac.res <- residuals(ccsfac.lm)

To estimate the linear trend in these residuals fit

farmgrp.lm <- lm(ccsfac.res ~ ccsgroup/Week, data=posmill.df, na.action=na.exclude, singular.ok=T) (7)

The overall mill trend for all seasons is estimated through the fit

polyccs.lm <- lm(Ccs ~ Week + Week^2 +Week^3 + Week^4 , data=posmill.df, na.action=na.exclude) (8)

The fitted values from models 6, 7 and 8 are added together to give an estimated paddock level CCS value, relevant for input to the optimisation model.

7. The coefficients and effects for these models are generated and output to a “txt” file as follows

Model (6): running the function op3models.fun(posmill.df)
NB: The fit for model 6 is contained in the latest version of op3models.fun.

Model (7): running the function op2models.fun(posmill.df)
NB: The fit for model 7 is contained in the latest version of op2models.fun.

Model (8): simply typing polyccs.lm (does not write to a file).

These model coefficient and effects are then forwarded on to Andrew Higgins for use in the CSO model. The functions op3models.fun and op2models.fun may need to be adjusted to generate the exact format required for input to the CSO model.

4. Directory Structure containing computer files

<u>Where to Find Latest S-plus Functions (developed 20 April 2002)</u>

The latest versions of the S-plus functions: millres.fun, millgrp.fun, millslope.fun, op2models.fun and op3models.fun have been developed for the Mossman data as an example. For ease of access these functions and the Mossman data are stored in the directory

D:/ My Documents / Packaged Modelling / Example Computer Files / Mossman.

You will need to access these files from within S-plus.
For other mill areas, given that the process from Sections 2 and 3 above have been followed and the data frame posmill.df and subsequently millslope.df has been created, these functions may be applied with little adjustment (if any) required.

BUT REMEMBER, at this stage the data files need to be copied to the same directory as the S-plus functions. The S-plus data files for each mill are located in the research modelling directories as described below.

Where to Find Latest Farm Slope-CCS Group Spreadsheets for Participating Mills

An Excel spreadsheet containing farm CCS group information for each of the mills in the Mossman, Mackay, Maryborough and Burdekin regions can be found in

D:/ My Documents / Packaged Modelling / Excel slope files

Directories for Research to Date

For the research process leading up to the development of modelling methods for the purpose of input to the CSO model, I created a directory called

D:/ My Documents / Michele_Mill Regions / Research by Mill Region .

Within this directory there is a sub-directory for each mill region considered, and a further sub-directory for specific mills within a region e.g. for Mackay ……/Mackay/Farleigh. Within this sub-directory there will be a range of admin type directories as well as a directory called "Research_Modelling" (in most cases, with Burdekin mills as the exception). Then there may be a further sub-directory indicating the year of the pilot study. This is where all of the modelling research was carried out and where you will find original S-plus data and functions. All of the posmill.df and millslope.df data frames are located in these directories.

Sub-directories for:

Mossman:
….. Research by Mill Region/Mossman/Research_Modelling/Stage Two 2001vs6

Mackay:
….. Research by Mill Region/Mackay/Research_Modelling/Pilot 2002/Farleigh6
….. Research by Mill Region/Mackay/Research_Modelling/Pilot 2002/Marian6
….. Research by Mill Region/Mackay/Research_Modelling/Pilot 2002/Pleystowe6
….. Research by Mill Region/Mackay/Research_Modelling/Pilot 2002/Racecourse6

Maryborough:

….. Research by Mill Region/Maryborough/Research_Modelling6

Burdekin:

….. Research by Mill Region/Burdekin/Invicta Mill6
….. Research by Mill Region/Burdekin/Pioneer Mill

NB: The 6 at the end of directory names refers to directories which needed to be converted to be compatible with version 6 of S-plus (using the "migrate" command).