INCORPORATING THE CANEGRO SUGARCANE MODEL INTO THE DSSAT V4 CROPPING SYSTEM MODEL FRAMEWORK

by

MATTHEW ROBERT JONES

Thesis submitted in fulfilment of the academic requirements for the degree of

MASTER OF SCIENCE

in Agrometeorology
School of Agricultural, Earth and Environmental Sciences
College of Agriculture, Engineering and Science
University of KwaZulu-Natal
Pietermaritzburg
South Africa

March 2013

ABSTRACT

Canegro is a leading sugarcane crop simulation model and has been used extensively in agronomic research and management. The model has been under development since the late 1980s at the South African Sugarcane Research Institute (SASRI). The Decision Support System for Agrotechnology Transfer (DSSAT) is a software package containing models for a wide range of field crops, and utilities for processing, storing and analysing model inputs and outputs. Canegro was included as part of version 3.1 of DSSAT in the mid-1990s. The SASRI Canegro model was subsequently developed further, but these changes were never integrated, nor incorporated, into DSSAT. DSSAT has also developed substantially, and as of version 4 adopted a modular Cropping System Model (CSM) structure, providing numerous scientific and practical advantages over previous non-modular versions. The DSSAT-Canegro v.3 model was not modified to use this modular structure.

Following recognition of the advantages offered by DSSAT and its modular CSM, a project was initiated to incorporate the Canegro model into the DSSAT CSM. The project entailed: (i) restructuring and integrating the current Canegro plant growth and development code into the DSSAT v4 CSM modular framework, making use of its generic modules for management, soil, weather and the energy balance; (ii) verification of DSSAT CSM Canegro model results against the current SASRI version of Canegro to ensure that the new model produced similar results to the original model, for a set of simulated situations; and (iii) evaluation of the new DSSAT CSM Canegro model against experimental datasets.

The new DSSAT v4 CSM Canegro model has been verified to behave identically to the SASRI Canegro model when the water balance is not modelled and growth can occur at climatic potential rates. When the water balance is simulated but where the crop is not stressed, near identical output is produced by both models. Under water-stressed conditions, some discrepancies appear between the two models, due to differences in the calculation of reference evaporation, soil surface evaporation and runoff. Validation of the new model against data from 16 experimental crops produced root mean squared errors of 6.62 t ha^{-1} for stalk dry mass and 3.59 t ha^{-1} for sucrose mass – very similar to published values for Canegro.

This project has yielded a functional, well-documented, maintainable and user-friendly version of the Canegro model, which is available for universal use via the official release of the DSSAT v4.5.

PREFACE

The research work described in this dissertation was carried out at the University of Florida, USA (September-December 2006), and at the South African Sugarcane Research Institute (SASRI), Mount Edgecombe, South Africa (January 2007 to May 2012), under the supervision of Dr Abraham Singels (SASRI) and Professor Michael Savage (University of KwaZulu-Natal, Pietermaritzburg).

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others it is duly acknowledged in the text.

DECLARATION 1 - PLAGIARISM

I, Matthew Robert Jones, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   a. Their words have been re-written but the general information attributed to them has been referenced
   b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed: ....

Matthew Robert Jones

Signed: ....
Supervisor: Michael J. Savage
Professor of Agrometeorology, UKZN

Signed: ....
Co-supervisor: Abraham Singels
Professor of Agrometeorology, UP

DECLARATION 2 - PUBLICATIONS

DETAILS OF CONTRIBUTION TO PUBLICATIONS that form part and/or include research presented in this thesis (include publications in preparation, submitted, *in press* and published and give details of the contributions of each author to the experimental work and writing of each publication)

Eustice, T, van der Laan, M. and van Antwerpen, R. 2011. Comparison of greenhouse gas emissions from trashed and burnt sugarcane cropping systems in South Africa. *Proceedings of the South African Sugar Technologists Association* 84:326-339.

Jones, MR, Porter, C, Jones, JW, Hoogenboom, G, Singels, A, Shine, J, Nayamuth, R, Kingston, G, Chinorumba, M, and van Den Berg, M. 2007. Incorporating the Canegro sugarcane model into the DSSAT v4 crop modelling system. *Proceedings of the International Society of Sugar Cane Technologists Congress*, 26:438-443.

Jones, MR, Singels, A and van den Berg, M. 2008. *DSSAT v4.5-Canegro Sugarcane Plant Module User Documentation*. International Consortium for Sugarcane Modelling, South African Sugarcane Research Institute, Mount Edgecombe, South Africa.

Jones, MR, Singels, A and Inman-Bamber, NG. 2011. Simulating source and sink control of structural growth and development and sugar accumulation in sugarcane. *Proceedings of the South African Sugar Technologists Association* 84:157-163.

Marin, FR, Jones, JW, Singels, A, Royce, F, Assad, ED, Pellegrino, GQ and Justino, F. 2012. Climate change impacts on sugarcane attainable yield in southern Brazil. In press.

Singels, A, Jones, MR and van den Berg, M. 2008. *DSSAT v4.5-Canegro Sugarcane Plant Module Scientific Documentation*. International Consortium for Sugarcane Modelling, South African Sugarcane Research Institute, Mount Edgecombe, South Africa.

Singels, A, Jones, MR, Porter, C, Smit, MA, Kingston, G, Marin, F, Chinorumba, S, Jintrawet, A, Suguitani, C, van den Berg, M and Saville, G. 2010. The DSSAT 4.5 Canegro model: a useful decision support tool for research and management of sugarcane. *Proceedings of the International Society of Sugar Cane Technologists Congress*, 27:438-443.

van der Laan, M, Miles, N, Annandale, JG and du Preez, CC. 2010. Identification of opportunities for improved nitrogen management in sugarcane cropping systems using the newly developed Canegro-N model. *Nutrient Cycling in Agroecosystems* 90(3):391-404.

Signed: Matthew Jones ……………………………………………………………

Matthew Robert Jones

ACKNOWLEDGEMENTS

I would like to offer my profuse and grateful thanks to:

- Dr Abraham Singels at the South African Sugarcane Research Institute (SASRI) for co-supervising this project and for his support and patience.
- Dr Maurits van den Berg (originally SASRI and now based in the Netherlands) for extremely insightful discussions, mentorship and support.
- Mrs Cheryl Porter (University of Florida) for her help, support, guidance, patience and friendship.
- Prof Jim Jones (University of Florida) for his support, guidance and encouragement.
- Prof Michael Savage (University of KwaZulu-Natal) for his supervision of this project.
- My wife, Sarah, for her support, encouragement, understanding, sympathy and love along the protracted journey this project has made.

I would also like to thank the International Consortium for Sugarcane Modelling for funding this project, and SASRI for supporting the project and funding my studies.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Background

1.1.1 Crop simulation models

Crop growth simulation models are computer programs that can estimate plant growth and development, in response to environmental conditions. They are intended as mathematical abstractions representing real plant growth processes in the soil-plant-atmosphere system. Crop models are used in two broad ways – in applied studies, where the models are used to estimate the values of agronomic variables, such as yield and water use; and as research platforms, where models are used to better understand and test plant physiology hypotheses (Jones *et al.*, 2003). To fulfil this purpose, crop models must be systematic, consistent and objective, and represent syntheses of research knowledge (Bouman *et al.*, 1993).

1.1.2 The Canegro sugarcane simulation model

The Canegro model is a sugarcane crop simulation model that has been under development at the South African Sugarcane Research Institute (SASRI) since the late 1980s (Inman-Bamber, 1991; Singels and Bezuidenhout, 2002). At SASRI, and worldwide, the model is considered a valuable research platform as well as a useful tool for crop modelling applications. Canegro draws its lineage from the CERES-Maize (Jones and Kiniry, 1986) and CropGro (Boote *et al.*, 1998) models developed during the mid-1980s. Canegro has been used in a number of studies, including analysing the effects of age and season on sugarcane growth and sucrose accumulation (Inman-Bamber, 1991), production potential assessment and benchmarking (Inman-Bamber *et al.*, 1998; Cheroo-Nayamuth *et al.*, 2000), crop forecasting (McGlinchey, 1999) and investigating agronomic practices for bio-energy production (Jones *et al.*, 2006).

1.1.3 The DSSAT CSM model

DSSAT – the Decision Support System for Agrotechnology Transfer (IBSNAT, 1989; Hoogenboom *et al.*, 1999; Jones *et al.*, 2003) is a powerful crop modelling software package for assisting with agronomic decision-making. It includes detailed plant growth models for a wide range of crops, software utilities for preparing model runs and analysing output, and reasonably comprehensive user documentation. DSSAT was initially developed within the

IBSNAT (IBSNAT, 1993) project, and the first release (v.2.1) included models for Maize, Wheat, Soybean and Groundnut (Jones *et al.*, 1998). Both the CERES-Maize and CropGro models were included in DSSAT v.2.1. The system is used worldwide and has more than 1000 registered users (Jones *et al.*, 2001; Jones 2006; Porter 2006). A new modular system, the 'Cropping System Model' (CSM), was introduced in DSSAT v4, which offers numerous advantages to compatible crop sub-models (Plant Modules) (Jones *et al.*, 2003).

A version of SASRI Canegro was included in DSSAT v3.1 (Inman-Bamber and Kiker, 1997) as a standalone model (Jones *et al.*, 2003). This model has effectively not been modified or updated since 1997. It has continued to be distributed as part of subsequent DSSAT releases (including DSSAT v4.0) as a legacy (non-CSM) model.

1.1.4 Modularisation of crop models

An integrated (non-modular) crop model is one where the software is designed such that all components are inter-dependent and cannot function without all other components. A modular system, by contrast, implements an approach to software design that divides an integrated system into programmatically- and scientifically-independent modules that perform clearly-defined tasks, represent particular physiological processes, and/or may be based on knowledge from a limited range of scientific disciplines.

As crop models advanced from scientific development platforms and became increasingly applied to help solve real-life decision-making problems, the limitations of integrated models became apparent. Several different approaches to – and interpretations of – modularity have been explored. APSIM (Agricultural Production Simulator (McCown *et al.*, 1996)) is a model where modularisation was motivated by a desire to provide a common soil model for different crops (O'Leary, 2000), and by making provision for simultaneous (different) crop instances (Jones *et al.*, 2001), it provides for modelling spatial interrelationships between fields. The FST/FSE (Fortran Simulation Translator/Environment (van Kraalingen *et al.*, 2003)) represents a more software-centric approach. In this case the focus was to develop a modular environment to enforce more structured but easier crop model development; examples of models implemented in this framework include SUCROS and ORYZA (Bouman *et al.*, 1996). The various crop models considered part of the CropGro family (Boote *et al.*, 1998) were combined into a common set of source codes, configured by text files of species and cultivar coefficients which characterised the generic model template for specific crops and cultivars thereof.

The modular approach used in the DSSAT CSM draws from all of these. The DSSAT CSM bases simulations on single hypothetical fields (Jones *et al.*, 2001), but focuses on simulating the dynamics of the plant environment (soil, atmosphere) over time, with support for features such crop rotations, sequences, and climatic modification (Jones *et al.*, 2003).

An FST/FSE-like framework is employed, and the model uses a common soil/atmosphere instance to ensure that crop sequences are consistently simulated. Modules are independent and autonomous, with well-defined inputs and outputs. They are represented as (complete) syntactic code units (Fortran subroutines), following the principles of Reynolds and Acock (1997).

Each module represents knowledge from a relatively small pool of scientific expertise, by dividing a crop model into distinct modules, each of which represents a particular bio-physical process. This principle of modularity, termed '*separation along scientific discipline lines*' (Reynolds and Acock, 1997), greatly facilitates model development and collaboration by limiting the scientific scope of any particular model development or maintenance task. In this way, modularity facilitates the widening of the scientific scope of crop model development as a whole, by clearly defining scientific discipline boundaries within modules.

Individual crops are represented either as standardised sets of genetic parameters read by a single set of program code (as in CropGro), or as dedicated plant modules which are represented by both crop-specific source code and crop-specific sets of genetic parameters.

1.2 Research Context

Canegro has been developed over nearly two decades (the first version was published in 1991 (Inman-Bamber, 1991)). The source code underlying the model reflects the incremental development history of the software. The model code is fundamentally well-structured, but the lack of a preconceived implementation plan (necessary, as it was impossible to anticipate software developments (and research outcomes on which these were based) that continued for the next two decades) for the software left the code in an unstructured state. This made continued model development extremely difficult.

A Canegro workshop was held at SASRI in Mount Edgecombe, South Africa, in 2000 (O'Leary and Kiker, 2001). As well as recognising the value of the model in the research context, a number of challenges and limitations were identified (O'Leary, 2000). Weaknesses in the simulation of water stress (Singels *et al.*, 2010b) and biomass partitioning

(Singels and Bezuidenhout, 2002), and the influence of air temperature on photosynthesis (Singels *et al.*, 2005), have since been addressed through research. The demand for these features was not so much Canegro-specific as identifying a need for these aspects to exist in any sugarcane model for research and applied use. Prior to the development of the DSSAT CSM Canegro plant module, these new features were only available to Canegro users at SASRI.

1.3 Justification for Research

The increasing complexity of the Canegro source code meant that the model became difficult to maintain, modify and use (Singels and van den Berg, 2006). In order to continue to build reliably on research and development work that had been invested in the Canegro model in the past, the need for substantial improvements to the maintainability of the Canegro source code had been identified. Restructuring of the source code into operational-, environmental- and plant process-specific modules, via a process of 'modularisation', was recognised as an appropriate means to achieve this.

The modular approach implemented by the DSSAT CSM offers a number of advantages to the crop model user, the model developer, and the programmer. These advantages make the DSSAT CSM an attractive platform for crop model development. The DSSAT system is already familiar to large numbers of crop scientists worldwide, and the generic modular approach of the CSM crops make the operation of the system for different CSM crops (and new crops, such as sugarcane) consistent and relatively easy.

For these reasons, the International Consortium for Sugarcane Modelling (ICSM, 2006a) provided funding for the development of a fully DSSAT v4 CSM-compatible implementation of the most recent SASRI Canegro model.

1.4 Hypotheses

Both the DSSAT Cropping System Model (CSM) and Canegro are derived from the CERES-Maize and CropGro crop simulation models. This project tested the following hypotheses:

(1) The similarity of pedigree of the DSSAT CSM and Canegro models meant that the nature and structure of the Canegro model made it suitable for incorporation into the DSSAT CSM.

(2) It would be possible to incorporate SASRI Canegro into the DSSAT CSM while retaining the original functionality of the SASRI Canegro model by generating near-identical simulation results for sugarcane plant growth (daily mass of plant components - roots, leaves, stalks, sucrose and trash) and development (tiller and leaf phenology, leaf area and stalk height) for equivalent simulation scenarios. The source of any discrepancies between the models could be explained, and traced to differences in the simulation of non-plant processes that were not originally part of the SASRI Canegro model.

1.5 Objectives of Project

The aim of the project was to incorporate the SASRI Canegro model (2006 version) into the modular DSSAT v4 Cropping System Model, such that:

- The DSSAT CSM Canegro model should exhibit near-identical simulation behaviour as the SASRI version of the model.
- The DSSAT CSM Canegro model should be integrated completely with the modular CSM, so as to take full advantage of the benefits offered by the modular structure in terms of use, maintainability, and suitability for further scientific development.
- The model should abide by DSSAT conventions and be compatible with DSSAT software utilities, to facilitate operation by users familiar with existing DSSAT crop models.

An additional aim was that the restructuring exercises involved in the software development would yield a clearer, more maintainable set of source code for the Canegro model.

The ultimate good can be considered to be: to create an internationally-usable version of the Canegro sugarcane model, that includes recent advances in knowledge, within the DSSAT CSM, serving as a robust platform for sugarcane model application and further scientific model development.

1.6 Structure of Thesis

The remainder of this thesis is divided into four broad sections:

- **A literature review (Chapters 2 and 3):** this section investigates the key scientific principles and technologies underlying the Canegro model, the DSSAT crop simulation system, and the modularisation of crop models in the DSSAT Cropping System Model (CSM).
- **Methodology (Chapter 4):** in this section, the process by which the Canegro model was analysed, restructured, re-implemented in the DSSAT modular format, and tested, is described.
- **Results and discussion (Chapters 5-8):** the results of this work are presented and discussed, in three sections:
    - Firstly, a physical description of the new modular DSSAT version of Canegro that was developed; this was the primary objective of the project (Chapter 6).
    - Secondly, a comparison of the performance of the new model compared with the old model, for hypothetical scenarios, to demonstrate that the models operate near-identically (Chapter 7).
    - Finally, a validation of the new model: independent graphical and statistical testing of the new model against measurements taken in real field trials (Chapter 8).
- Thereafter, a brief summary of the project results is provided (Chapter 9).

An appendix containing cultivar coefficients used for code verification, and a **CD-ROM** containing source code listings of the new model are provided.

2. THE CANEGRO MODEL

2.1 History and Background of the Canegro Model

The Canegro sugarcane model (Inman-Bamber, 1991; Singels and Bezuidenhout, 2002) has been under ongoing development at the South African Sugarcane Research Institute (SASRI) since the late 1980s (Inman-Bamber, 1991; O'Leary, 2000; Inman-Bamber, 2001; Singels and Bezuidenhout, 2002). Canegro is used as a research tool, providing an objective platform for investigating and testing plant physiology hypotheses, exploring genotype by environment interactions, yield benchmarking and optimising field management strategies (including irrigation scheduling and timing of planting/harvest). It represents a synthesis, and to some extent a benchmark, of sugarcane plant physiology knowledge at SASRI. Worldwide, the model is considered a useful tool for crop modelling applications. The origins of Canegro lie with the CERES-Maize (Jones and Kiniry, 1986) and CropGro (Boote *et al.*, 1998) models developed from the mid-1980s (Inman-Bamber, 2001).

The need for a mechanistic sugarcane model was created by questions posed to SASRI by growers and millers in the South African sugar industry. Among the earlier applications of the model, which "defied conventional field experimentation" (Inman-Bamber, 2001, pg. 5) was determining optimum harvest age for sugarcane under the different agro-climatic regions within the SA sugar industry. Farmers had been forced to rethink conventional 18-24 month seasons by the ravaging Eldana stalk borer, which appeared to favour cane older than 12 months. This study required a sensitive, mechanistic model of the sugarcane canopy and carbon balance (Inman-Bamber, 1991; 2001).

Canegro existed in two versions. SASRI actively developed and maintained a version of Canegro, which for the purposes of this thesis will be termed the 'SASRI Canegro' model. A version included in DSSAT (the Decision Support System for Agrotechnology Transfer, discussed in the next chapter), which will be termed 'DSSAT-Canegro v.3', was first released as part of DSSAT v3.1 in 1997 (Inman-Bamber and Kiker, 1997; O'Leary, 2000). The DSSAT-Canegro v.3 model read and wrote DSSAT-compatible files and adhered to DSSAT usage conventions. Development of the two versions of Canegro diverged, and prior to this project, only the SASRI Canegro version was still maintained. DSSAT-Canegro v.3 was packaged as part of DSSAT v.4 as a legacy model, and at the time of writing was the current release version of DSSAT. The 2006 SASRI Canegro model contains more advanced algorithms for simulation of many key processes. It was, however, accessible only to SASRI scientists.

Canegro is considered to be among the leading sugarcane models worldwide (O'Leary 2000). The model is calibrated for South African (Inman-Bamber, 2001) and Mauritian (Cheeroo-Nayamuth and Nayamuth, 2001) sugarcane cultivars, for sub-tropical conditions. It is believed that the DSSAT v.3 version of Canegro is used fairly widely, particularly in academia. The continued development of SASRI Canegro by a number of different researchers over the years has meant that the model has become increasingly complex and difficult to maintain (van den Berg, 2005). Despite the development of a graphical user interface for creating input files, it is also relatively difficult to use: creating the format-sensitive input files requires care and patience, and adding support for cultivars and conditions outside of the southern African region is cumbersome.

DSSAT-Canegro v.3 was the first attempt to make Canegro accessible to an international audience. The Canegro Plant Module developed during this project (referred to as 'DSSAT CSM Canegro') will be released as part of the official DSSAT v.4.5, and has been released in beta version to many users who have attended the DSSAT training courses since 2007 (Porter, 2010). The DSSAT CSM Canegro model has become the development version of Canegro within SASRI, meaning that improvements and revisions to the Canegro algorithms can be released in future versions of DSSAT with ease.

Descriptions of different model processes are provided in the following sections of this chapter. The scientific bases for the model are presented first, followed by a brief discussion of the model's simulation performance. The structure of the program itself (the Fortran code, input and output files) is presented thereafter.

2.2 Model Processes

A diagram detailing key SASRI Canegro model plant and development processes is shown in Figure 2.1. For the sake of simplicity, this diagram excludes the water balance and water stress impacts. Each process is described in detail in Sections 2.2.1 to 2.2.9.

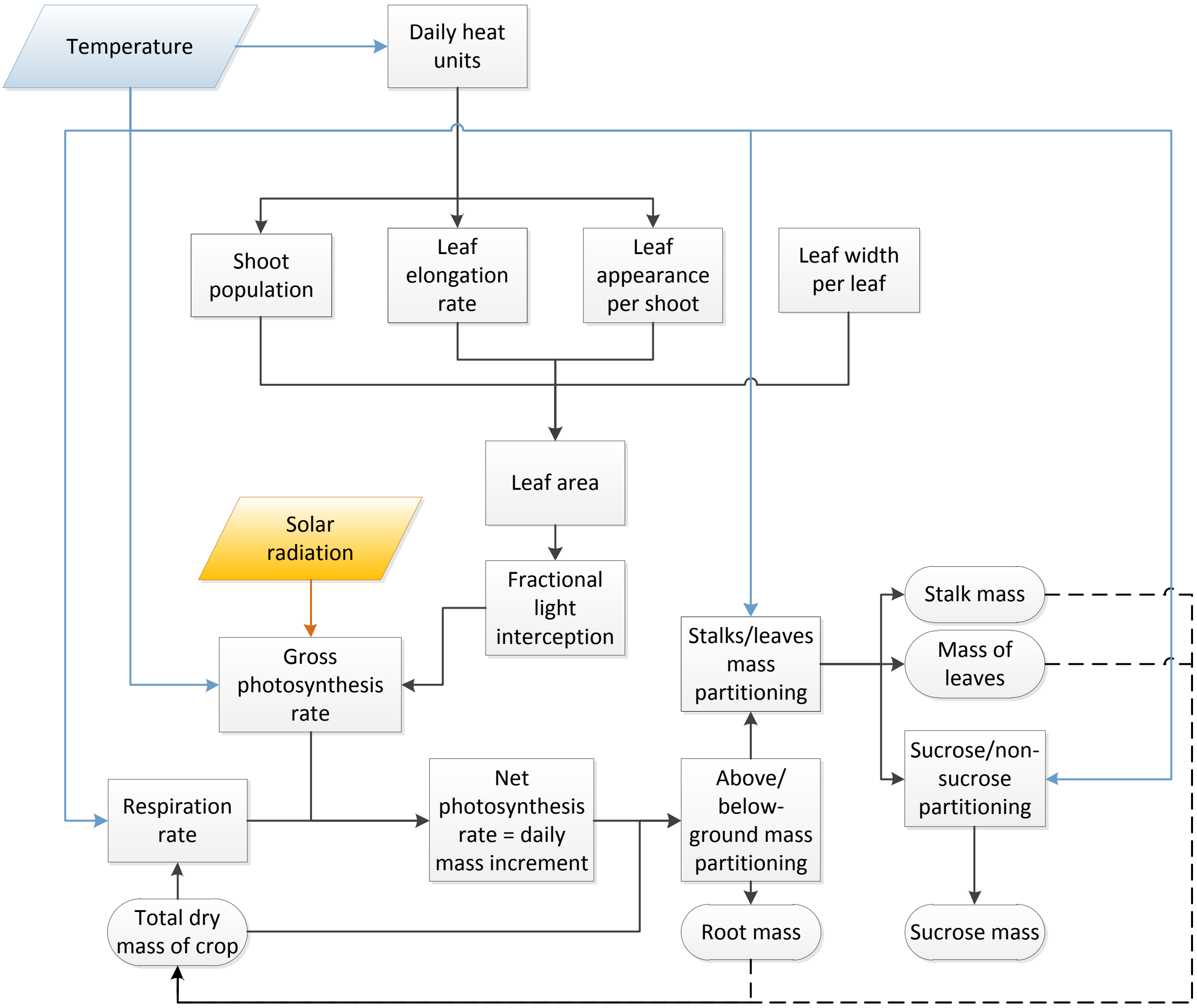


Figure 2.1 Schematic diagram of SASRI Canegro model plant growth and development processes. For simplicity, water stress impacts are excluded, but affect shoot population, leaf elongation rate and gross photosynthesis rate.

2.2.1 Germination and emergence

Canegro simulates germination and emergence as a single process. Phenological stages are not defined as such (Inman-Bamber, 1991). Emergence is a single event, and is the day that the first shoot appears. Time to emergence was originally set to 21 days (Inman-Bamber, 1991) but the model currently considers a thermal time delay from planting to emergence (°C d above a base temperature (°C) (Inman-Bamber, 2001)). Thermal time from crop start to emergence for NCo376 is 428 °C d (plant crop) and 203 °C d (ratoon crop), with a base temperature of 10 °C.

2.2.2 Tiller population

Tiller population (tillers.ha^{-1}) is described by two functions, one for population increase, and the other for tiller senescence following peak population.

Tiller population increase

A polynomial function of thermal time (HU_s, °C d) above a base temperature (T_{base}, °C) is used to describe tiller population growth. Sugarcane tiller populations reach a genotype-specific maximum value (for NCo376 this is around 500 000 tillers ha^{-1}), but this is limited in Canegro to 300 000 tillers ha^{-1}, as the shoots that grow after this were considered insignificant (code comment in the SASRI Canegro tiller population source code file Poplt3.for). Inman-Bamber (1991) described the following equation for calculating tiller population increase:

$$N_s = tc_1 + tc_2.HU_{s,n} + tc_3.HU_{s,n}^2 + tc_4.HU_{s,n}^3 \tag{1}$$

$$HU_{s,n} = \sum_{day=1}^{n} MAX(0.,T_{mean,n} - T_{base,n}) \tag{2}$$

where T_{mean} is daily mean air temperature (°C), N_s is the tiller population (tillers ha^{-1}) and $HU_{s,n}$ is thermal time (°C d) above T_{base} on day n of the simulation (Inman-Bamber, 1991). Tillers emerge in cohorts of 11 000 tillers ha^{-1}, and each cohort emerges with the appearance of each new leaf.

Inman-Bamber (1991) suggested values of 25000, 1120, -0.984 and 0.000224 respectively for tc_1, tc_2, tc_3 and tc_4 for NCo376.

The algorithm was subsequently modified as follows:

$$N_s = \frac{1.4}{rowspc} 1000(tc_2.HU_{s,n} + tc_3.HU_{s,n}^2), HU_{sn} < 600 \tag{3}$$

where 1.4 represents a reference row-spacing of 1.4 m, *rowspc* is actual row-spacing (m), and values for tc_2 and tc_3 are set to 1.826 and 0.00201 respectively for NCo376. The change in order of magnitude of the coefficients is partly explained by the change in units of the polynomial component of the equation, from tillers ha^{-1} (Eqn (1)) to 1000 tillers ha^{-1} (Eqn (3)). Rates of change of tiller population are limited by the water stress factor $SWDF_{30}$, described in Section 2.2.7.

Inman-Bamber (1994) indicated that the base temperature for tiller appearance is 16 °C for cultivars NCo376 and N12.

Tiller senescence

Tiller population decreases after canopy closure (at peak tiller population) to a relatively stable population at harvest. For NCo376 this value is 133 000 tillers ha^{-1}. After HU_s reaches 600 °C d, or if tiller population decreases below 300 000 tillers ha^{-1} following peak population, the calculation of N_s switches from Eqn (3) to Eqn (4). This is used to reduce the tiller population to the final population (Singels *et al.*, 2008):

$$N_s = N_{s,n-1} - 0.004\left(HU_{s,n} - HU_{s,n-1}\right) \tag{4}$$

Previously, the following equation was used:

$$N_s = \frac{1.4}{rowspc}1000\left(tc_3 + tc_4.HU_{s,n} + tc_5.HU_{s,n}^2\right), HU_{sn} \geq 600 \tag{5}$$

with values for tc_3, tc_4 and tc_5 of 866.7, -0.99024 and 0.0003282 respectively.

2.2.3 Canopy development and radiation interception

Canopy development

A phyllocron interval is the thermal time between successive leaf tip appearance events on a single tiller. Leaf appearance is calculated according to thermal time above a base temperature $T_{baseleaf}$ (°C). Two phyllocron intervals (°C d) are specified in Canegro: the first phyllocron interval (PI_1) is used for calculating the timing of appearance of the first N_L leaves. For subsequent leaves, PI_2 is used. Inman-Bamber (1994) suggested values of 109 °C d, 169 °C d, 14 leaves tiller^{-1} and 10 °C for PI_1, PI_2, N_L and $T_{baseleaf}$ respectively for variety NCo376.

Leaf size and area is driven by leaf elongation rate (LER, cm d^{-1}). Unstressed leaf elongation rate (LER_P) is calculated as follows (Inman-Bamber, 1991; Inman-Bamber, 1994):

$$LER_P = -1.77 + DLERDT(T_{MEAN}).\frac{24}{10} \tag{6}$$

where T_{MEAN} is daily mean air temperature (°C), $DLERDT$ is the leaf growth rate (mm h^{-1}), and the 24/10 term scales this value to cm d^{-1}. Actual leaf elongation rate (LER_A, cm d^{-1}) is limited by water stress:

$$LER_A = LER_P \,.SWDF_2 \tag{7}$$

where $SWDF_2$ is a water stress factor described in Section 2.2.7.

Maximum length ($LeafLength_{MAX,i}$, cm), width ($LeafWidth_{MAX,i}$, cm) and area ($LeafArea_{MAX,i}$, cm^2) of leaf *i* (leaf number in order of appearance on a tiller) are determined by polynomial functions of leaf number (Inman-Bamber, 1991, p. 94):

$$LeafWidth_{MAX,i} = 7.8 + 2.24i - 0.034i^2 \tag{8}$$

$$LeafLength_{MAX,i} = 21.8 + 12.2i - 0.376i^2 \tag{9}$$

$$LeafArea_{MAX,i} = -56.6 + 36.9i - 0.945i^2 \tag{10}$$

The actual length of leaf *i* ($LeafLength_i$, cm) is the sum of LER_A for each day of its growth but is limited to a maximum length given by Eqn (9). The width of the leaf ($LeafWidth_i$, cm) is proportional to the length of the leaf, as follows:

$$LeafWidth_i = \frac{LeafLength_i}{LeafLength_{MAX,i}}\, LeafWidth_{MAX,i} \tag{11}$$

Area of individual leaves can be reduced by up to 50% by a reduction factor calculated from water stress.

Leaf area index (LAI, m^2 m^{-2}) is defined as leaf upper surface area over a unit area of ground. The area of individual leaves on all tillers growing in one square metre is calculated and summed to determine LAI.

Leaf area index can be calculated for green leaves only (LAI_G), in which case only the surface area of healthy leaves capable of photosynthesis and transpiration are considered. LAI can also be calculated using the surface area of healthy and senesced leaves still attached to the shoot (LAI_T), for calculating soil surface shading.

Stalk elongation starts when a genotype-specific thermal time has elapsed (*CHUPIBASE*, 1050 °C d at a base temperature of 10 °C) after emergence. Potential (unstressed) stalk elongation rate (SER_P, cm d^{-1}) is calculated with the following equation:

$$SER_P = C_{PER}.LER_P \tag{12}$$

where C_{PER} has the value 0.16 cm cm^{-1} for NCo376.

Actual stalk elongation rate (SER_A, cm d^{-1}) is SER_P limited by $SWDF_2$:

$$SER_A = SER_P \,.SWDF_2 \tag{13}$$

Stalk height (H_{STALK}, cm) is calculated by adding SER_A to the previous day's stalk height ($H_{STALK,yesterday}$, cm):

$$H_{STALK} = H_{STALK,yesterday} + SER_A \tag{14}$$

Radiation interception

The fractional interception (LI) of the photosynthetic portion of solar irradiance (termed photosynthetically-active radiation (PAR, MJ m^{-2}) is calculated as an exponential function of LAI, according to Beer's Law:

$$LI = 1 - e^{-k.LAI} \tag{15}$$

where k is a PAR extinction coefficient that ranges from 0.58 at the first leaf tip appearance to 0.86 after 20 leaf tips have appeared (Inman-Bamber, 1991). LI calculated using LAI_G is used to determine photosynthesis and transpiration. LI calculated using LAI_T determines shading of the ground surface, the complement of which is used to estimate soil surface evaporation.

2.2.4 Photosynthesis and respiration

Photosynthesis is calculated according to a model described by Inman-Bamber and Thompson (1989). Gross photosynthesis (P_g, g m^{-2}) is determined by radiation use efficiency (RUE, g MJ^{-1}), according to (Inman-Bamber, 1991):

$$P_g = RUE.PAR.LI \tag{16}$$

Eqn (16) excludes the effects of lodging, discussed in Section 2.2.8.

RUE is influenced by air temperature:

$$RUE = RUE_{MAX}\left(1 - e^{KRUE(T_{MEAN} - B_{photos})}\right) \tag{17}$$

where RUE_{MAX} (g MJ^{-1}) is the theoretical maximum RUE (9.9 g MJ^{-1}, Singels and Bezuidenhout (2002)) and B_{photos} is the base temperature for photosynthesis (7 °C). $KRUE$ is an air temperature sensitivity coefficient and has a value of -0.08 (Singels *et al.*, 2005), and T_{MEAN} is daily mean air temperature (°C).

Maintenance respiration (R_m, g g^{-1}) is the carbohydrate cost of maintaining plant structures. R_m was initially considered a fixed proportion of total dry biomass (W, g m^{-2}) (0.3 %, Inman-Bamber, 1991, increased to 0.4 %, Inman-Bamber, 2001), and then replaced with an air temperature-determined function (Singels and Bezuidenhout, 2005):

$$R_m = R_{Ref}.C_{Rm}.e^{0.1(T_{MEAN}-10)} \tag{18}$$

where C_{Rm} is a temperature sensitivity parameter (C_{Rm} = 1.68) representing the fractional increase in respiration rate per 10 °C rise in air temperature and R_{Ref} is a reference respiration rate at 10 °C (0.0033871 g g^{-1}) (Singels *et al.*, 2008).

Growth respiration (R_{gr}, g g^{-1}) is an energy cost associated with producing new plant structures (Singels and Bezuidenhout, 2005). Growth respiration is calculated as a fixed proportion of P_g (0.242 g g^{-1} (Inman-Bamber, 1991)).

Daily biomass increment ($DWDT$, g m^{-2} d^{-1}) includes a reduction due to water stress ($SWDF_1$, discussed in Section 2.2.7), and can be expressed as:

$$DWDT = \left(P_g - R_m.W\right).\left(1 - R_{gr}\right).SWDF_1 \tag{19}$$

2.2.5 Biomass partitioning

Daily biomass increment ($DWDT_t$, $DWDT$ expressed in t ha^{-1} d^{-1}) is partitioned into above- and below-ground components according to the following equations (Singels and Bezuidenhout, 2002):

$$DWDT_A = F_A.DWDT_t \tag{20}$$

$$DWDT_B = (1 - F_A).DWDT_t \tag{21}$$

$$F_A = MAX\left(0.05, F_{A,MAX}.(1 - e^{-b.W})\right) \tag{22}$$

where F_A is the fraction (0.05-1.00) of biomass allocated to aerial parts of the crop, $F_{A,MAX}$ is a maximum fraction that can be allocated, b is a partitioning coefficient, and $DWDT_A$ (t ha^{-1} d^{-1}) and $DWDT_B$ (t ha^{-1} d^{-1}) are the daily increments in above- and below-ground dry biomass respectively. F_A should not decrease below 0.05 to ensure that very young and small crops receive some allocation to aboveground biomass.

$DWDT_A$ is integrated with the previous day's aboveground (aerial) dry biomass ($W_{A,yesterday}$, t ha^{-1}) to yield aboveground dry biomass (W_A, t ha^{-1}):

$$W_A = W_{A,yesterday} + DWDT_A \tag{23}$$

Similarly, the dry mass of roots (W_R, t ha^{-1}) is calculated from the previous day's root mass ($W_{R,yesterday}$, t ha^{-1}) and the daily increment in below-ground biomass ($DWDT_B$):

$$W_R = W_{R,yesterday} + DWDT_B \tag{24}$$

$DWDT_A$ is then partitioned into stalk and leaf components. Daily stalk ($DWDT_S$, t ha^{-1} d^{-1}) and leaf ($DWDT_L$, t ha^{-1} d^{-1}) biomass increments are calculated as:

$$DWDT_S = F_S.DWDT_A \tag{25}$$

$$DWDT_L = (1 - F_S).DWDT_A \tag{26}$$

where Singels *et al.* (2005) indicated that the stalk partitioning fraction (F_S) is calculated as a function of mean daily air temperature (T_{MEAN}, °C), and two temperature sensitivity coefficients FSK (-0.5) and FST (7.5 °C):

$$F_S = 0.65 + e^{FSK(T_{MEAN} - FST)} \tag{27}$$

The dry mass of stalks (W_S, t ha^{-1}) each day is the integration of the previous day's stalk mass ($W_{S,yesterday}$, t ha^{-1}) and $DWDT_S$:

$$W_S = W_{S,yesterday} + DWDT_S \tag{28}$$

Similarly, the dry mass of leaves (W_L, t ha^{-1}) is calculated from the previous day's leaf mass ($W_{L,yesterday}$, t ha^{-1}) and $DWDT_L$:

$$W_L = W_{L,yesterday} + DWDT_L \tag{29}$$

Partitioning of stalk biomass into sucrose and non-sucrose components originally used an empirical rather than mechanistic approach, because the mechanisms were not fully understood (Inman-Bamber, 2001). Singels and Bezuidenhout (2002) developed a more sophisticated sucrose partitioning algorithm for SASRI Canegro. This approach makes use of source-sink concepts for sucrose and non-sucrose accumulation in the stalk.

Sucrose content in a sugarcane stalk is greatest at the base of the stalk and lowest at the immature growing tip of the stalk. Stalk sucrose dry mass content cannot exceed a maximum value (SC_{MAX}, g g^{-1}). In young stalks the sucrose content at the base of the stalk will be less than the sucrose content at the base of older and more mature stalks. A stalk will have a greater sucrose content at the base than at the top of the stalk. The distribution of sucrose content over the stalk, from the base to the tip, is termed the "ripening gradient" (M_R, Singels and Bezuidenhout (2002)). Under conditions that promote vigorous growth, the ripening gradient will be reduced as more of the stalk will be immature, compared with cool and/or water-stressed conditions under which stalk elongation rates are limited. Under limited growth rate conditions, a greater proportion of the stalk is considered mature and so would have an increased ripening gradient. For the purposes of calculating sucrose content, the crop is represented as a single large stalk.

The ripening gradient is calculated in Canegro each day by a genotype-specific maximum value (M_{MAX}, 0.07 ha t^{-1} for NCo376) limited by water stress (F_W) and air temperature (F_T) factors:

$$M_R = M_{MAX}(F_W + F_T - F_W . F_T) \qquad (30)$$

and

$$F_T = \frac{1}{1 + e^{FTCON(T_{MEAN} - T50)}} \qquad (31)$$

$$F_W = (1 - SWDF_2)^{FWCON} \qquad (32)$$

where $FTCON$ and $FWCON$ are genotype parameters, with values of 0.32 and 0.5 respectively for NCo376. $T50$ is also a genotype parameter, defined as the air temperature (25 °C for NCo376) at which the ripening gradient is half of M_{MAX} (Singels and Bezuidenhout, 2002). $SWDF_2$ is a water stress index explained in Section 2.2.7.

The theoretical maximum sucrose mass (SUC_{EQ}, t ha^{-1}) of the stalk each day is the integral of the mature and immature sections of the stalk. This integration is the sum of the area of a triangle for the un-ripened section and the area of a rectangle for the ripened section, which reduces to:

$$SUC_{EQ} = SC_{MAX}.W_S - \frac{{SC_{MAX}}^2}{2M_R} \tag{33}$$

where W_S is stalk mass (t ha^{-1}). The complement of SUC_{EQ} is FNS_{EQ} (t ha^{-1}), the maximum theoretical mass of fibre and non-sucrose in the stalk:

$$FNS_{EQ} = W_S - SUC_{EQ} \tag{34}$$

The "sink capacity" (*FCAP*, t ha^{-1}) for fibre and non-sucrose (*FNS*) each day is the difference between actual *FNS* (FNS_A, t ha^{-1}) at the end of the previous simulated day and FNS_{EQ}. The change in *FNS* (*DFNSDT*, t ha^{-1}) each day is the product of *FCAP* and *SSR* (0.99, Singels and Bezuidenhout (2002)), which represents "sink activity". The sucrose mass increment each day ($DWDT_{SUC}$, t ha^{-1} d^{-1}) is simply the difference between $DWDT_S$ and *DFNSDT*:

$$DWDT_{SUC} = DWDT_S - DFNSDT \tag{35}$$

Sucrose mass (W_{SUC}, t ha^{-1}) is the sum of the previous day's sucrose mass ($W_{SUC,Yest}$, t ha^{-1}) and the current day's change in sucrose mass:

$$W_{SUC} = W_{SUC,Yest} + DWDT_{SUC} \tag{36}$$

Stalk fresh (wet) mass (W_F, t ha^{-1}) is calculated as

$$W_F = 4.607W_S - 2.078W_{SUC} \tag{37}$$

based on the findings of Martine and Lebret (2001).

2.2.6 The water balance and root growth

The southern African sugarcane industry is generally characterised by shortages of water (Inman-Bamber, 2001), rather than nutrients. Canegro thus has a well-developed water balance subroutine (Inman-Bamber, 2001). Figure 2.2 shows a schematic diagram of the model's water balance processes.

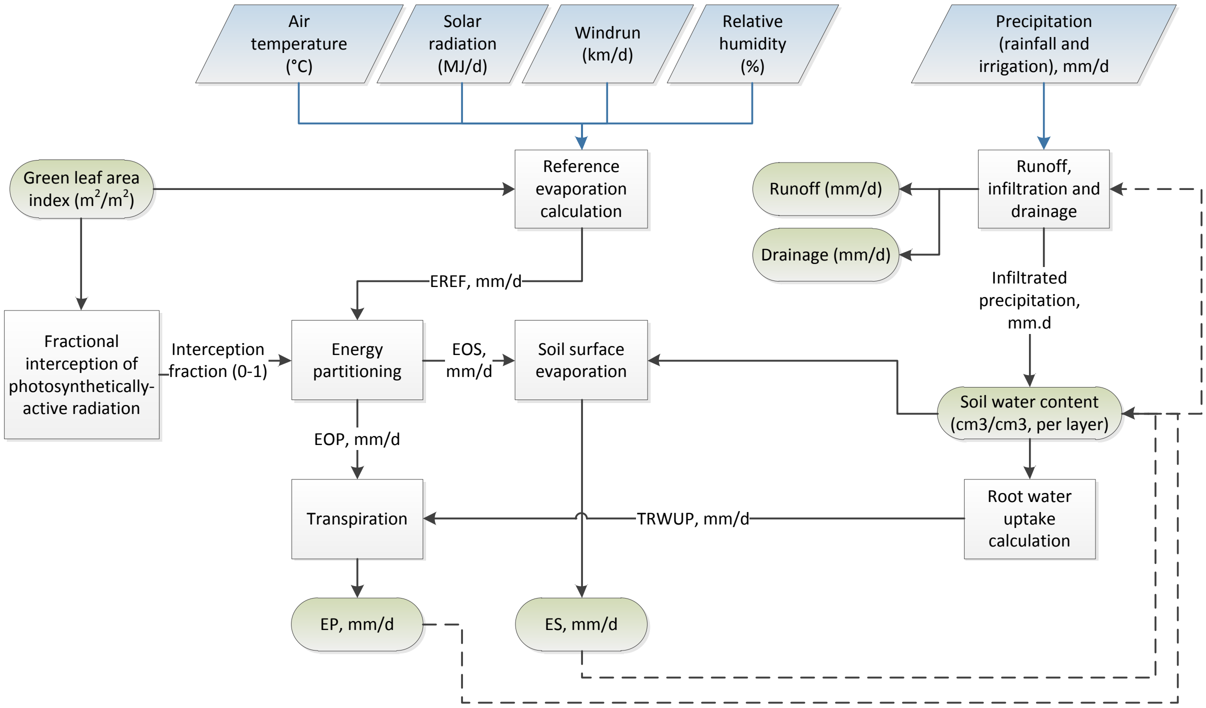


Figure 2.2 Schematic diagram of water balance processes in the SASRI Canegro model. Daily weather data variables are represented by blue diamonds. The green rounded rectangles are state variables, and unshaded rectangles represent model calculations or processes. Solid lines indicate data flows during one day of simulation, while dashed lines indicate effects on the next day's calculations. Acronyms are explained in the text.

Daily reference evaporation (E_{REF}, mm d^{-1}) is the amount of water demanded by the atmosphere of a full-canopied, well-watered crop. Potential plant transpiration (EOP, mm d^{-1}, defined as the amount of that water would be transpired if soil water/roots were not limiting (van Antwerpen, 1998)) is set to a fraction of E_{REF} proportional to canopy cover:

$$EOP = E_{REF}.LI \qquad (38)$$

The complement of this fraction drives potential soil surface evaporation (EOS, mm d^{-1}):

$$EOS = E_{REF}(1 - LI) \qquad (39)$$

Three options for calculating E_{REF} are available. The most accurate is the Penman-Monteith (Monteith, 1965) algorithm modified to take account of the height, LAI_G and canopy resistance of a reference sugarcane crop (McGlinchey and Inman-Bamber, 1996; McGlinchey, 2001). In addition to solar radiation, and air temperature, this method requires relative humidity and wind-speed information, which may not be available. Alternatively, the model can use a method based on the Priestley-Taylor (Priestley and Taylor, 1972) approach (requiring only solar radiation and air temperature), or can read pre-calculated reference evaporation or recorded pan evaporation values.

Soil water balance and root water uptake algorithms in Canegro are derived from the CERES-Maize model (Jones and Kiniry, 1986) (van Antwerpen, 1994).

For the soil as a whole, the daily water balance equation is as follows:

$$Rainfall + Irrigation = EP + ES + Runoff + Drainage + \Delta SW \qquad (40)$$

where EP (mm d^{-1}) represents transpiration of water from the roots to the atmosphere via the leaves, ES (mm d^{-1}) is the loss of water from the soil via soil surface evaporation, and ΔSW (mm d^{-1}) is the change in soil water content. The sum of EP and ES are limited by E_{REF}. EP is limited by EOP on the demand side and by potential total root water uptake ($TRWUP$) on the supply side. ES is demand-limited by EOS and supply-limited by moisture availability near the soil surface. Net precipitation comprises rainfall and irrigation, less runoff (soil surface runoff or stormflow, water that does not infiltrate the soil surface). Water is lost to deep drainage (mm d^{-1}), when the soil water content of the deepest layer (that can be reached by the roots) exceeds its drained upper limit (DUL) of plant-extractible water, (cm^{3} cm^{-3}, defined as the volumetric soil water content at a soil water potential of -10 kPa).

Canegro simulates a layered soil, described by the depth of the lower boundary of each layer, and for each layer: the lower limit (LL) of plant-extractible water (cm^{3} cm^{-3}, defined as the volumetric soil water content at a soil water potential of -1500 kPa), DUL, bulk density (g cm^{-3}) and saturated hydraulic conductivity (mm h^{-1}). Water drains vertically between layers when the soil water content of the upper layer exceeds its DUL. Root growth is directed by a root distribution function described as part of the soil definition.

Soil surface evaporation uses the Ritchie (1972) two-stage approach. Three runoff curve numbers (82 for up to 50% canopy cover, 75 for 50-75% canopy cover and 70 for when the plant has greater than 75% canopy cover (Schmidt *et al.*, 1998)) are used to more subtly simulate runoff as the canopy closes.

The calculation of daily root mass increment ($DWDT_B$) was described in Section 2.2.5. $DWDT_B$ is transformed into root length (cm) at a fixed rate. This rate can be set by the user but the default value is 500 cm g^{-1} (Inman-Bamber, 2001). Maximum root water use is set to 0.007 cm^{3} mm^{-1} d^{-1} (van Antwerpen *et al.*, 1994; Inman-Bamber, 2001). Root length density

(RLV, cm cm^{-3}) in each layer is calculated according to root length allocated on the basis of layer thickness and water content.

$TRWUP$ is the sum of individual soil layers' potential root water uptake ($RWUP_i$, mm d^{-1} per layer i). $RWUP_i$ is given by the CERES equation:

$$RWUP_i = C_1 e^{\frac{MIN(C_{2,i}(SW_i - LL_i),4)}{(C_3 - LOG(RLV_I))*RLVF}} \tag{41}$$

where C_1 is a constant (0.00132), C_3 is a constant (6.67), $C_{2,i}$ is calculated per layer i as:

$$C_{2,i} = SWC_{120} - 250 * LL_i \tag{42}$$

where SWC_{120} is an input (typical value 120) and LL_i is the lower limit of plant extractible water for layer i (cm^3 cm^{-3}), and $RLVF$ is an input factor affecting root length density, with a typical value of 1.

Actual total root water uptake ($TRWU$, mm d^{-1}) is limited by either $TRWUP$ or EOP, whichever is smaller. If EOP is limiting, root water uptake per layer (RWU_i) is allocated proportionally to the relative total root length in each soil layer. SW_i is then updated as follows:

$$SW_i = SW_{i,yest} - RWU_i + INF_i - Drainage_i \tag{43}$$

where INF_i is the amount of water infiltrating into this layer from the one above, $SW_{i,yest}$ is the previous day's soil water content (cm^3 cm^{-3}), and $Drainage_i$ is the amount of water draining from this layer to the one below. This calculation is performed in a cascade from the topmost layer downwards.

2.2.7 Water stress

Water stress occurs if atmospheric demand for water exceeds the plant's ability to supply water via the roots ($TRWUP$) (Inman-Bamber, 2001). Root water uptake is limited by water availability in the soil and the size and depth of the root system.

Water stress factors are calculated from EOP and RWU. The approach used in the CERES-Maize model is followed in Canegro. Two water stress factors are calculated, namely $SWDF_1$ and $SWDF_2$. These stress factors are calculated in Canegro as:

$$SWDF_1 = MIN\left(1, f_1 . {}^{RWU}/_{EOP}\right) \quad (44)$$

$$SWDF_2 = MIN\left(1, \left(f_2 . {}^{RWU}/_{EOP}\right)\right) \quad (45)$$

where $f_1 = 1$, $f_2 = 0.5$, and RWU is the root water uptake (mm d^{-1}).

$SWDF_2$ limits leaf extension rate and stalk extension rate (Eqns (7) and (13)). $SWDF_1$, the less sensitive index of water stress, affects biomass accumulation (Eqn (19)). Under conditions of mild water stress (such as a cool dry winter period), therefore, expansive growth slows more than photosynthesis rate. This naturally favours sucrose accumulation under these conditions, a phenomenon observed in reality.

$SWDF_{30}$ is a soil water stress factor affecting tillering. It is calculated as follows:

$$SWDF_{30} = 3\left(\frac{SW_i - LL_i}{DUL_i - LL_i}\right) DL_i, \sum_{i=1}^{L} DL_i \leq 30 \quad (46)$$

where DL_i is soil layer thickness (cm), and L is the number of soil layers whose lower boundary does not exceed 30 cm depth. $SWDF_{30}$ is therefore an index of soil water availability in the top 30 cm of the soil profile.

2.2.8 Lodging

The calculation of lodging in SASRI Canegro has not been published. It is based on work by Singh *et al.* (1999). The lodging calculation consists of two components: factors that cause lodging, and the impacts lodging has on *PAR* interception, and therefore evapotranspiration, photosynthesis and yield. Lodging is assumed to be affected by three risk factors: soil water content, wind speed and above-ground dry mass. In all cases, larger values lead to a greater extent of lodging. Lodging is a cumulative process, so that with each lodging event the proportion of the crop that is lodged increases.

The soil water content impact on lodging (L_{SWC}) is given value of 2 if runoff exceeds 0 mm d^{-1} (indicating a saturated topsoil), otherwise it has a value of 0.

The wind risk factor (L_{WIND}) is calculated as:

$$L_{WIND} = MIN(2, \frac{wind_{km} - 200}{100}) \tag{47}$$

where $wind_{km}$ is the daily wind-run in km. This implies that if wind-run is below 100 km d^{-1}, there is no risk of lodging from wind, and the impact of wind cannot exceed 2 on a contribution scale of 0-8 (i.e. 25% of the total contribution to lodging).

The mass risk factor (L_{MASS}) is given by:

$$L_{MASS} = -7.3, W_S \leq 1 \tag{48}$$

$$L_{MASS} = 8\frac{\frac{W_{AW}}{W_F.W_S} - 220}{30}, W_S > 1 \tag{49}$$

Essentially, the fresh mass of the above-ground components of the crop, plus the mass of intercepted precipitation (W_{AW}, t ha^{-1}), is calculated (if stalk dry mass (W_s, t ha^{-1}) exceeds 1 t ha^{-1}). If this mass exceeds a critical mass (hard-coded to 220 t ha^{-1}), the mass risk for lodging starts. W_F is fresh stalk mass (t ha^{-1}).

The risk of lodging (L_{TOT}) is the sum of these risk factors, limited to between 0 and 8, (corresponding with the SASRI plant breeder's variety lodging rating scale of 1-9):

$$L_{TOT} = MAX(0, MIN(8, L_{SWC} + L_{WIND} + L_{MASS})) \tag{50}$$

The fraction of the crop lodged (F_{LODGE}) is then calculated:

$$F_{LODGE} = \frac{L_{TOT}}{8} \tag{51}$$

Lodging is allowed to reduce the rate of gross photosynthesis (P_g, g m^{-2} d^{-1}) *PAR* interception by up to 28%. Hence:

$$P_g' = P_g * \big((1 - F_{LODGE}) * 0.28\big) \tag{52}$$

where P_g' is gross photosynthesis rate reduced by lodging (g m^{-2} d^{-1}).

2.2.9 Nutrients

SASRI Canegro does not model nutrient effects. An initial attempt to model nitrogen in the DSSAT v.3.5 Canegro model, based on concepts from CERES-Maize, was made by O'Leary *et al.* (2001b). Van der Laan *et al.* (2010) developed a nitrogen module for the DSSAT CSM Canegro model.

2.3 Canegro Model Performance

Table 2.1 presents SASRI Canegro performance indicators for aerial (above-ground) dry mass (t ha^{-1}), stalk dry mass (t ha^{-1}) and sucrose mass (t ha ^{1}) as reported by O' Leary (2000) and Singels and Bezuidenhout (2002). Values are shown for root mean squared error (RMSE), and linear regression R^{2} and slope, for these variables for the 24 sugarcane treatments (in five experiments) described in Singels and Bezuidenhout (2002), as well as RMSEs reported by O'Leary (2000). These values present the performance of Canegro model prior to the inclusion of the temperature-sensitive photosynthesis and respiration algorithms (described in Section 2.2.4).

Table 2.1 Performance of the Canegro sugarcane model as indicated by root mean squared error (RMSE) values, and linear regression R-squared and slope values. Values as reported in O'Leary (2000) and Singels and Bezuidenhout (2002).

Variable	**RMSE[a] (t ha^{-1})**	**RMSE[b] (t ha^{-1})**	**R^{2} [a]**	**Slope[a]**
Aerial (above-ground) dry biomass	6.94		0.70	0.84
Stalk dry mass	5.48	11.10	0.82	0.84
Sucrose mass	3.18	6.07	0.86	0.95

[a] Singels and Bezuidenhout (2002)

[b] O'Leary (2000)

Based on these validation results, the authors concluded that Canegro gave satisfactory simulation of above-ground biomass, stalk mass and sucrose mass.

2.4 Canegro Computer Program Organisation

Canegro is a computer program that encapsulates scientific concepts and algorithms described in Section 2.2. Knowledge of the structure of the Fortran code, and the organisation of the computer program, was essential for the restructuring task that was the main objective of this project. The description that follows is based on an analysis of the code and operation of the model.

2.4.1 Program operation

The Canegro model is executed on the command line with the name of the INP file (see below) as a command line parameter. It is also possible to run the model from a rudimentary graphical interface.

Several formatted-text data input files are required to run a simulation using Canegro. These are:

- **The input configuration file (INP)** contains a list of paths to other files (DTE, PRN, SOL, RAM, IRI) that Canegro requires for a simulation run. Each line of this file contains a file path. Canegro expects these in a standard order.
- **The date file (DTE)** contains a list of dates for the start and harvest of crops to be simulated. The first column in this file indicates plant/ratoon status. A sequence of crops can be simulated, although the values of state variables are not maintained between crops.
- **The simulation settings file (PRN)** contains an extensive list of switches that configure the model itself (e.g. choice of reference evaporation calculation), crop management parameters (e.g. row-spacing, irrigation), initial conditions (e.g. initial soil water content), references to variety information (including the path to the variety file in use and the index number of hard-coded variety parameter sets in the code), and a file reference to the output settings (SET) file in use.
- **The output settings file (SET)** instructs the model as to which state/rate variables should be output.
- **The variety file (VAR)** contains some genetic information on the variety currently in use.
- **The soil file (SOL)** contains information on the depth of the soil profile, soil albedo, drainage properties, and runoff curve numbers. Layer thickness, water retention characteristics (drained upper limit, lower limit, saturated water capacity (all cm^3 cm^{-3}); saturated hydraulic conductivity (cm h^{-1})), bulk density (g cm^{-3}, optional), initial soil water content (cm^3 cm^{-3}, optional) and relative root distribution are listed per soil layer.
- **The daily weather data file (RAM)** provides daily solar radiation (MJ m^{-2}), maximum and minimum air temperatures (°C), rainfall (mm d^{-1}), wind-run (km d^{-1}), maximum and minimum relative humidity (%), and 8 a.m. and 2 p.m. wet and dry bulb temperatures (°C).
- **The irrigation application file (IRI)** is optional and contains irrigation applications on particular dates. This file is only read if the model is configured to irrigate.

The model generates a single formatted-text output file (OUT). The relationship between these files is illustrated in Figure 2.3.

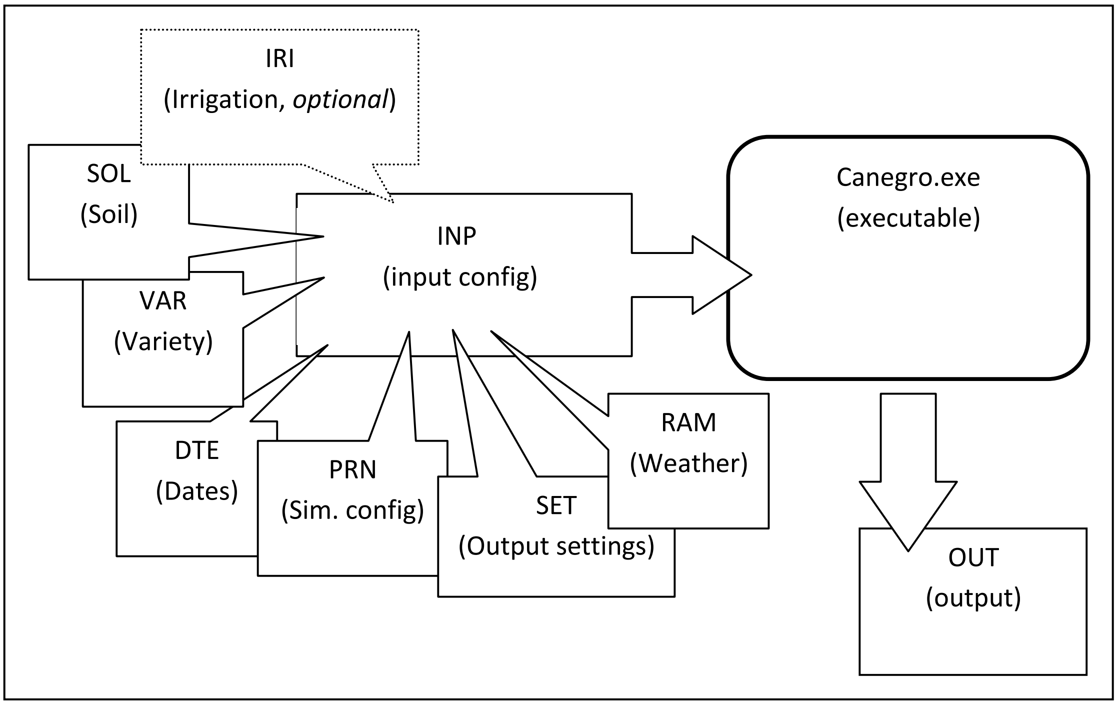


Figure 2.3 Organisation of the Canegro computer program, showing model executable reading inputs from variety, date, simulation configuration, output settings, daily weather and soil files, and an optional irrigation application file, and producing an output file.

2.4.2 Program organisation

Canegro consists of a single executable file implemented in the Fortran computer language. The code follows the 1977 Fortran version's syntax style (fixed format source code), with some language features from the 1990 version of Fortran.

Subroutines and source code files

Table 2.2 lists the name and purpose of each of the major subroutines in the Canegro model, and the source code files in which they are defined.

Table 2.2 Names and purposes of major Canegro subroutines and the main program, and their source code filenames. Items are listed in alphabetical filename order.

Subroutine name	**Source file name**	**Main purpose**
CANINT	CANINT.for	Initialise with choice of hard-coded variety parameters
CANOP3	Canop3.for	Calculates leaf and canopy development: leaf area index and radiation interception
DOYEAR	DOYEAR.for	Convert day, month year values into year and day of year

INITIAL	Initial.for	Initialisation of global variables
IRRIGT	IRRIG.for	Calculates irrigation applications
IRRMAN	Irrmanc.for	Manage irrigation settings and initialise irrigation-related variables
LIMITS	LIMITS.for	Checks weather data
(main program)	Mainnewc.for	Entry point into program. Coordinates operation of program (opens files, calls initialisation procedures, runs daily loop, calls growth and development procedures)
METMAN	Metmana.for	Initialisation of weather data reading variables
NOTICE	Hgtetal.for	Subroutines for checking weather data, generating frost and lodging warnings
OUTPUT	OUTPUT.for	Process and write output
PARTIT	Partit.for	Biomass partitioning (above-ground biomass partitioned to stalks and tops; then stalks to sucrose and non-sucrose)
PENMON	Penmon.for	Calculate reference evaporation with the Penman-Monteith model
PHOTOS	Photos.for	Calculate gross and net photosynthesis, and growth and maintenance respiration; above- and below-ground biomass partitioning
PHYSOP	Physoptc.for	Read PRN file and check/prompt for parameter values
POPLT3	Poplt3.for	Calculate stalk population
PRIEST	PRIEST.for	Calculate sugarcane reference evaporation using a method based on the Priestley-Taylor evaporation model
SOILRD	Soilmet.for	Read soil file and initialise soil profile
SOLAR	SOLAR.for	Calculate incident solar radiation from sunshine hours, latitude and date
SUCPDM	SUCPDM.for	Alternative biomass partitioning function
WATBAL	Watbal.for	Calculate the water balance (runoff, infiltration, soil surface evaporation, transpiration, root growth and water stress).

Canegro is a state variable-based model that operates on a daily time-step. The calculations performed for each day of the simulation represent the core steps in the simulation of the crop. The daily loop formed the basis of the analysis of the code for the incorporation of Canegro into DSSAT as a Cropping System Model (CSM) module. The relationship between the source code files listed in Table 2.2 and the daily loop calculations is illustrated in Figure 2.4.

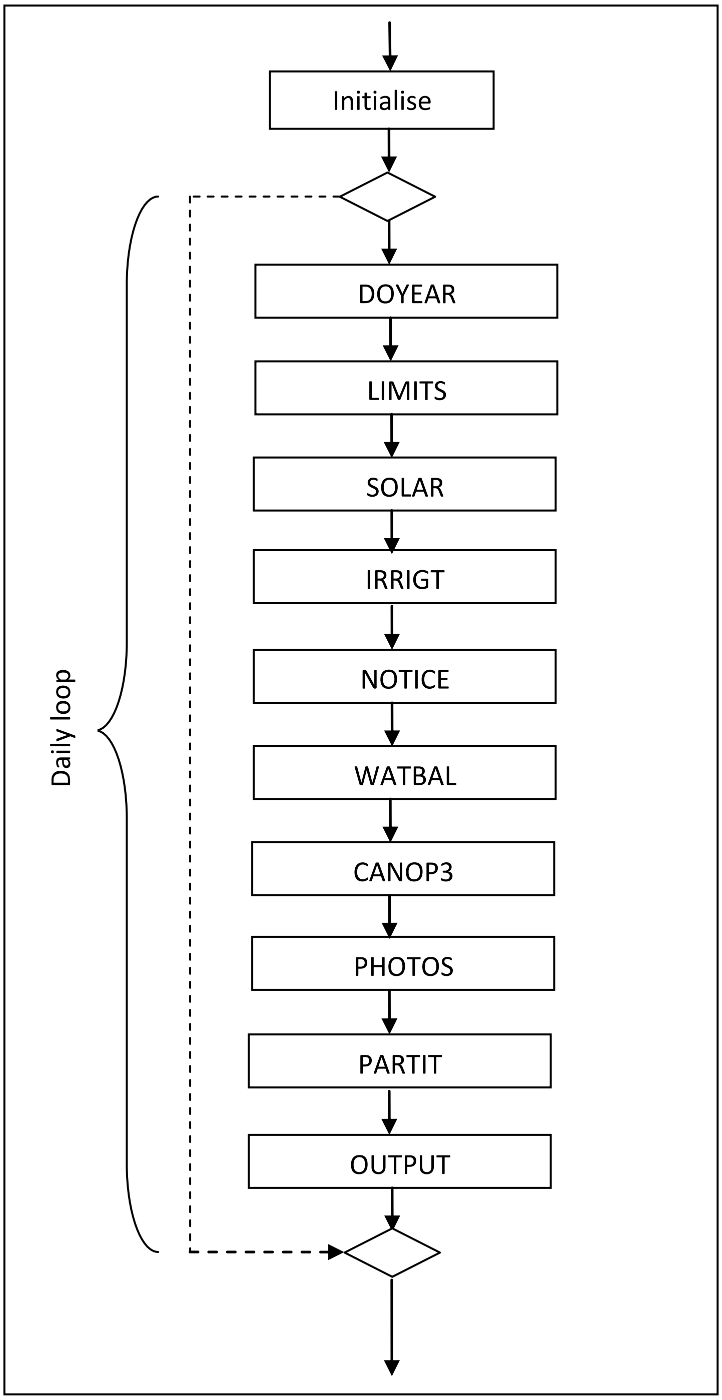


Figure 2.4 Order of subroutine calls executed for each day of a Canegro crop simulation, following initialisation.

Source code issues

A number of shortcomings in the approach to Canegro were identified. These are discussed here because they were addressed in the creation of the new DSSAT CSM module version of Canegro.

Variable type declarations

If a variable is not explicitly declared, Fortran will implicitly set its type (REAL, INTEGER, CHARACTER, etc.) according to the first letter of the variable name. In most Canegro source code files and subroutines, variables were not explicitly typed. This was problematic because variables could be implicitly wrongly-typed, or misspelt variable names would be implicitly created as new variables, affecting the program logic.

Subroutine parameters

Fortran subroutine parameters (or ‘arguments’) are variables that are passed into a subroutine from the code block from which the subroutine was called. They are passed by reference, which means that a subroutine can change the value of a subroutine parameter variable. Fortran INTENT statements, which can be optionally used to specify if parameters are inputs or outputs, were not used in Canegro. This forced the programmer to examine the code to determine if subroutine parameter variables were inputs, outputs or both.

Common statements

The Canegro source code makes extensive use of Fortran COMMON statements to mark sets of variables to have program-global scope. This means that those variables are accessible from anywhere within the program. These sets of variables are generally referred to as ‘common blocks’, a term which will be used in this thesis.

The values of the variables defined in a common block will be the same in all subroutines that declare the same common block. While these constructs make initial programming easier, they obscure the flow of information in the program. They bypass the subroutine parameter list, meaning that the programmer must examine the subroutine parameter list and the common block definitions to determine where the value of a particular variable comes from.

Common blocks are not always defined identically in all instances where they are used, potentially leading to a situation where one variable might be interpreted as another. Such programming errors can be nearly impossible to diagnose or identify.

Source code comments

The Canegro code was not consistently documented. The few code comments made the source code difficult to follow.

3. THE DECISION SUPPORT SYSTEM FOR AGROTECHNOLOGY TRANSFER (DSSAT)

3.1 Introduction to DSSAT

The intention of this project was to incorporate the Canegro model into the Decision Support System for Agrotechnology Transfer (DSSAT) Cropping System Model. This chapter sets out the background knowledge and understanding that was essential for achieving this effectively and appropriately.

The history and purpose of the system are described, followed by a discussion of the various components of DSSAT. The concept of modularity in crop models is then introduced in the context of the DSSAT Cropping System Model (CSM) and the discussion progresses from abstract concepts of modularity to concrete programming specifications for DSSAT CSM modules. Plant-related and non plant-related modules of the DSSAT CSM are discussed in detail. DSSAT input and output conventions are then described.

The chapter concludes with brief discussion of the DSSAT v.3 version of the Canegro model.

3.2 Overview of DSSAT

3.2.1 History

DSSAT, the Decision Support System for Agrotechnology Transfer (IBSNAT, 1989; Jones *et al.*, 1998; Jones *et al.*, 2003) is a software package consisting of crop models and related software tools (Jones *et al.*, 1998), and was developed as part of the International Benchmarking Sites Network (IBSNAT) project (IBSNAT, 1993). The initial intention was to provide a mechanism for using a systems approach to support agricultural production decision-making in resource-constrained situations (Jones *et al.*, 1998). The systems approach focussed on understanding the underlying relationships between weather, crops and soil, and then integrating this knowledge into models for predicting production and resource use (Jones *et al.*, 1998; Jones *et al.*, 2003). Jones *et al.* (1998) describe DSSAT as a system that allows a user to "estimate production, resource use and risks associated with different crop production practices".

Several DSSAT versions have been released. The initial release (version 2.1) in 1989 (IBSNAT, 1989) provided crop models for maize, wheat, soybean and groundnut. Version 3 added models for eight additional crops: rice, drybean, sorghum, millet, barley, potato, aroid and cassava. Version 3.1 included the Canegro sugarcane model (Inman-Bamber and Kiker, 1997). Version 4 was the first release of the modular Cropping System Model (Jones *et al.* 2003), and v.4.0.2 remains the current release version at the time of writing (2012). Version 4.5 has not yet been officially released, although it will include the modular version of the Canegro model (Porter, 2010) as a product of this project. DSSAT is relatively easy to use, and enjoys worldwide uptake as a crop modelling platform (more than 1000 users in 90 countries (Jones *et al.*, 2001; Jones 2006; Porter 2006)).

3.2.2 DSSAT functionality

DSSAT provides the ability to (1) manage, display and analyse weather, crop and soil data; (2) calibrate and evaluate crop growth models and (3) evaluate different crop management practices (Jones *et al.*, 1998). Point (2) makes provision for users to improve simulations for a particular site by allowing them to change parameters to better describe local cultivars. Point (3) refers to the execution of crop models to evaluate different agronomic management practices. This allows users to simulate plant growth and development for a wide range of scenarios – these include a large number of supported crops, different permutations of management choices for experiments, and crop sequences and rotations.

3.2.3 Structure and components of the DSSAT system

The structure of the DSSAT v.3.5 system is described by Jones *et al.* (2003), and consists of five broad components (see Figure 3.1):

1. Databases for soil, weather, genetic, experiment, pest and economic information.
2. Support software for tasks such as creating experiment files, importing weather data, and graphing and analysing data.
3. Applications for performing analyses of simulated and observed data for validation, seasonal runs and crop rotations/sequences, and linking to geographical information systems for spatial analyses.
4. The set of crop simulation models.
5. A user interface for linking the various components.

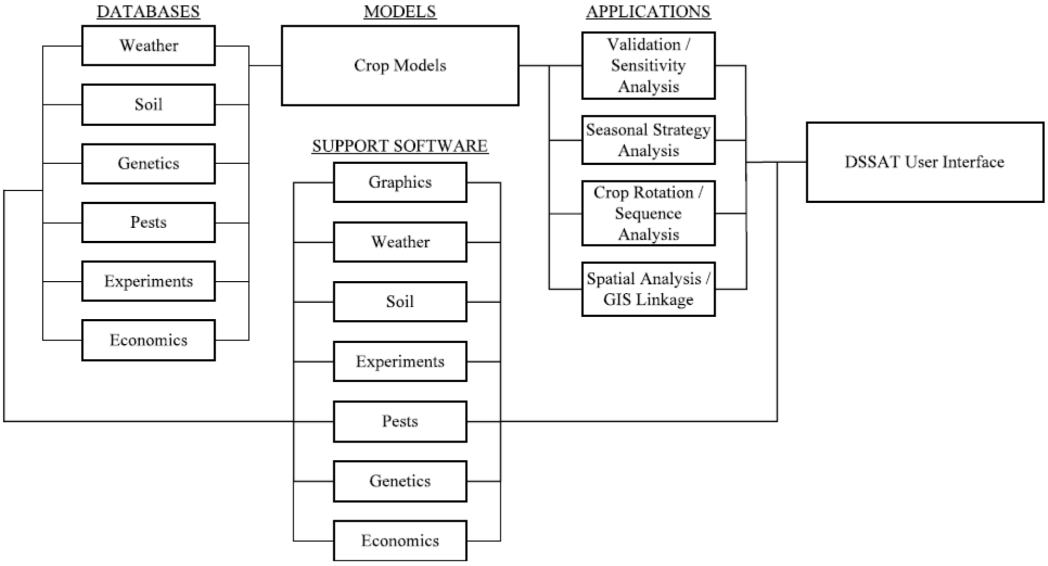


Figure 3.1 The structure of DSSAT v.3.5; taken from Jones et al. (2003).

In versions prior to v.4, the crop modelling component of DSSAT consisted of a set of individual, separate crop model programs. These were integrated via a uniform user interface and common input/output file formats (Jones *et al.*, 2001). The modular DSSAT Cropping System Model (CSM) (Jones *et al.*, 2003) was developed for version 4 of the software, following recognition of a number of shortcomings that non-modular systems suffer, and additional advantages that modular systems afford. The benefits to modularising the DSSAT crop models are significant from both software development/maintenance and user perspectives. The structure of the CSM and a discussion of its advantages are presented in Section 3.3.2. The other significant difference between versions 3.5 and 4 of DSSAT is that the latter is now Windows- rather than DOS-based, but the basic structure (as illustrated in Figure 3.1) is unchanged.

3.2.4 Operation of crop simulations in DSSAT

Each DSSAT crop simulation run is represented in the DSSAT system as what is termed a 'treatment'. A treatment consists of information about the crop species, cultivar, soil, weather data source, planting and harvest details, irrigation settings, other field management settings, and simulation settings. Treatments are grouped into what are termed 'experiments', even if the simulation represents a set of completely hypothetical crops. An experiment is defined in a 'FILEX' (Hunt and Boote, 1998). FILEX files are formatted ASCII text files. They are compact and not easily understood, and so these files can be created and edited using a graphical desktop program, XBuild.

The Input program (in v.4.0.2) or module (v.4.5) (Porter, 2006) reads and interprets the FileX and generates an intermediate and temporary INP ('input') file for each separate simulation run. The INP file is fed to the crop model for execution of a single model run. This intermediate step allows users to override experiment settings and parameter values at runtime, by changing one or more parameter values after they have been read from the FileX but before model execution. This facility is provided via an interactive user mode and is intended to allow users to explore simulation behaviour without the encumbrance of having to modify several input files.

Output is generated in formatted-text 'OUT' files. OUT files are grouped into sections for each treatment, and different files contain output from each model process. For example, output from plant growth and development routines is captured in 'Plantgro.OUT', and evapotranspiration output is captured in 'ET.OUT'. Some input data are echoed to output files: weather data, for example, are output in WEATHER.OUT.

3.3 Modularity in the DSSAT v.4 Cropping System Model

This project involved the creation of a DSSAT CSM modular version of the Canegro model. To do this most efficiently and appropriately required not only a clear understanding of the programming aspects of the CSM, but also an understanding and appreciation of the motivation and purpose behind modularity in the DSSAT CSM.

The discussion below explores modularity in the DSSAT CSM. A comparison of integrated and modular approaches to crop model development is presented first, as somewhat abstract concepts. These ideas are then reduced into a set of guiding principles, which in turn feed into more specific programming specifications and requirements for the DSSAT CSM.

3.3.1 Integrated vs. modular approaches

The traditional software architecture approach to crop modelling can be considered integrated and procedural. This is characterised by a linear sequence of source code, executed in a linear fashion. Modular models, on the other hand, divide the source code of the model into separate and self-contained units. Ideally, this separation is done according to scientific discipline (e.g. soil water uptake processes separated from root growth and development processes). Modules should also be represented as discrete syntactic units (e.g. subroutines, classes) in the source code, meaning that such crop models are simultaneously modular from both scientific and software engineering perspectives.

The incorporation of Canegro into the DSSAT CSM involved restructuring the procedural (integrated) Canegro source code into the DSSAT CSM modular format. The discussion below attempts to highlight some of the advantages and disadvantages of these different approaches to model development. The discussion opens with a very short overview of how state variable-based crop models generally operate.

Crop model operation

State variable-based crop models operate by calculating changes in magnitude of several state variables per unit of time ('time-step'), and updating the values of the state variables at the end of each time-step by integrating these rates of change. State variables represent the ‘state’ of the crop, and might include the mass of different plant components (e.g. roots, stalks, leaves and grain), physical characteristics of the crop (such as height, leaf area and shoot population), and the physical state of the environment in which the plant grows (e.g. the volume of water in the soil). Mathematical equations capture relationships between plant state variables, environment state variables (e.g. soil water content) and rates of change of the environment (usually weather data variables such as daily rainfall, maximum and minimum air temperatures and solar radiation) to determine rates of change of plant state variables. The time-step for most crop models is a single calendar day, although some models (e.g. Boote and Pickering, 1994) use hourly time-steps for certain processes. In general, the size of the time-step is limited by the temporal resolution of the input weather data (Jones and Luten, 1998), which is usually one day. The greatest shortcoming of daily time-steps is rainfall: a daily time-step cannot differentiate between an intensive downpour, in which a substantial proportion of precipitation runs off the soil surface, and a more even distribution of gentle rainfall across the day, most of which would infiltrate the soil and become available to the plant.

A crop model must read daily weather data from a file, database or other source of data. It must also produce output. Crop models need to be configured for specific crop situations and genotypes by reading in genetic, environmental and field management information.

The basic structure of the state variable approach to crop model design implies a certain logical organisation of computer programs that represent crop models. A crop model program starts by reading configuration information and initialising state variables. It then reads inputs for each time-step in a loop, calculates rates of change for state variables and then integrates them, until the season is complete or an error occurs. Output is produced daily and/or aggregated over several time-steps (weekly, monthly, or possibly the entire season). When the time-step loop (often termed the 'daily loop') is complete, the crop model program will close input and output files and exit.

Ideally, the calculation of rates of change for all state variables (those representing the plant and its environment) are simulated simultaneously, and then integrated simultaneously. This is not possible with a sequential computer program, but strict separation of rates calculations and the integration of these rates of change approximates this.

Integrated and modular models necessarily differ only in the structuring of the program source code. It is technically possible that an integrated model could perform an identical simulation to a modular one, because the difference lies within the structure of the source code, not the sequence of execution of statements.

The following discussion attempts, however, to argue that the modular approach, particularly as implemented in the DSSAT Cropping System Model, has practical benefits over the integrated approach.

Integrated models

In integrated models, source code statements appear in the source code set in the order in which they are executed. Plant growth- and development-related processes and environmental processes are simulated together. Source code for initialisation of plant and environmental processes, calculation of rates of change of plant- and environment-related state variables, and their integration, are not formally separated and appear in the same logical sections of the program.

Integrated models are relatively simple to develop. Source code statements are – logically – written in the same sequence in which they are executed. Well-structured integrated models are not inherently better in terms of simulation accuracy than well-structured modular models. Poorly-structured modular models are likely to have all of the disadvantages of poorly-structured integrated models superimposed onto non-linear source code structures.

New scientific knowledge drives improvements and changes to crop models, and models are often used as test-beds for investigating theories in the fields of crop physiology, changes in soil condition, interactions between the environment and the plant, climate change, and many others, which require ongoing modification (rather than simply application) of the models.

After a few iterations of development, release and redevelopment, integrated models can become unwieldy to manage: code becomes increasingly complex and difficult to understand and debug. Inputs become more complex as new options are added, and extra information is required for more advanced calculations. A well-managed and disciplined software development process will safeguard against such eventualities. Crop models are, however, generally written by scientists often unfamiliar with these software development processes.

Collaborative work on integrated models can be difficult, as working on a single scientific aspect of the model often requires an understanding of the whole model and code changes from different contributors must be merged into a single entity. The incremental development of integrated models is often so difficult that new models are frequently developed in preference to building on existing ones (Reynolds and Acock, 1997). It is, however, quite possible to write integrated models in a structured manner that negates many of these possible shortcomings.

In cases (such as DSSAT v.3.5 and earlier) where a change is required to a non crop-specific aspect of a suite of several models (for example, the addition of a new potential evapotranspiration algorithm), it is necessary to re-implement the change in each model. This can increase the scale and complexity of the operation enormously (Jones *et al.*, 2001).

Where the simulation of plant and non-plant processes is integrated, there is a danger that a shortcoming in the simulation of a plant process is inappropriately compensated for by a complementary over- or under-estimation in the simulation of one or more environment variables. Modularity itself offers no safeguard against this. However, where a single module is used by several crop models, it is tested by a larger and more diverse group of model developers, users and simulation scenarios. This tends to ensure that common modules are more robust and crop-specific modules are calibrated appropriately. It must be acknowledged that this approach can also reduce flexibility in the model development process.

Integrated models typically need to be implemented in a single language (Jones *et al.*, 2001). Given the advances made in computing technology, knowledge of a different programming language ought not to be a barrier to contributing to a scientific software development project.

Modularity in crop models

As crop models have advanced – as increasingly complex scientific development platforms and as applied tools for helping to solve real-life decision-making problems – the limitations of integrated crop models have become apparent. Jones *et al.* (2001) presented a comprehensive overview of modularity in crop models as an alternative.

Modularity has been approached independently by a number of modelling groups in the past. The 'School of de Wit' (Bouman *et al.*, 1993) models, including SUCROS (van Laar *et al.*, 1997) and ORYZA (Kropff *et al.*, 1994) have used modular approaches from very early stages of development, based originally on the IBM CSMP (Continuous Simulation Modeling Program, IBM, 1975) and subsequently the FST/FSE (van Kraalingen *et al.,* 2003) modelling environments. APSIM (McCown *et al.*, 1996) makes use of a sophisticated object-oriented modular structure to perform whole-farm simulations. DSSAT v.3.5 (Hoogenboom *et al.*, 1999) employed a modular approach following recognition of the similarities between several models based on the CROPGRO model (Boote *et al.*, 1998). The DSSAT v.4 Cropping System Model approach to modularity draws from all of these (Jones *et al.,* 2003).

The key feature of modular models is the organisation of the source code. The source code in a modular model is arranged into groups, and each group contains code statements that are closely related or perform some particular well-defined function in the program. Modules are typically self-contained, and conform to standard programming interfaces. This allows the addition or removal of modules without any other modification to the software source code.

This approach to source code organisation has numerous benefits, discussed in detail later in this chapter.

Modularity can be considered from "systems" and "software" perspectives (Jones *et al.*, 2001). The systems perspective divides the crop model source code into modules based on scientific discipline – these might include soil water, nutrition, surface residue, soil surface evaporation and transpiration. The software view considers issues of efficient and manageable software design; for example, providing for routines to be defined in dynamically-linked libraries, issues of code re-use (to prevent potentially-inconsistent repetition of code) and standard mechanisms for communicating information within the program.

The systems perspective

Modularity allows scientists to focus on scientific, rather than software-related, aspects of crop model development (Jones *et al.,* 2001). When code is grouped logically and independently into modules along scientific discipline lines, it is possible for different sections of models to be independently developed and maintained, as long as input and output parameters are well-defined and described. Modules can be added and replaced without any code outside the module being affected.

The software perspective to modularity

Modularity is implemented in software, and it is manifested in the structure of the source code. The primary feature is that all source code representing a scientific process or concept is captured in a single syntactic unit (Reynolds and Acock, 1997). Syntactic units differ between programming languages, but may include subroutines, procedures, data structures and objects. The sequence of execution is very likely to differ from the sequence in which the source code statements appear in the source code. This non-linear organisation of source code is one of the defining features of modularity.

The implementation of modularity in a crop model is subject to practical considerations of the realities of the crop model software development process. An understanding of the modular framework – in the software sense – is necessary before model development within a modular system can proceed. All software frameworks involve a learning curve of some kind, and the more complex the software, the higher the level of programming expertise required to work within it.

To put the discussion of the software engineering implications of the DSSAT CSM into context, two modular modelling frameworks are briefly discussed: ACRU (Agricultural Catchments Research Unit (Schulze, 1989, 1995)) and APSIM (Agricultural Production Simulator (McCown *et al.*, 1996)). These are examples of systems where a very high level of programming expertise is required for model development and maintenance.

ACRU is a daily-timestep catchment-scale hydrological model developed at the University of Natal, Pietermaritzburg. ACRU model development began in the 1970s, with the initial implementation in Fortran. Over the years, with successive software development by a large number of scientists and students, the Fortran code became increasingly difficult to maintain and further develop. Between 1999 and 2002 the ACRU model was re-implemented in the

Java programming language, and adopted a highly-structured object-oriented software framework. The core software is based on three types of objects, representing the physical components of the simulated system (such as catchments and dams), data, and hydrological processes. Additional objects are used for input and output. Model development requires the addition of new process objects, or the modification of existing ones. After implementing the Java object-oriented structure, it was found that the original text file inputs did not allow for the potential level of flexibility offered by the new software structure. XML (Extensible Markup Language)-based file formats were introduced to enrich input sufficiently to match the model functionality. These XML files are more difficult to work with, however, and this necessitated the development of user interface utility software for creating and editing input files (Kiker *et al*., 2003).

Detailed knowledge of both the object-oriented software structure and how it represents the hydrological system modelled, along with the more complex XML-based file format, is essential for effective continued development of the ACRU model. Ongoing scientific development of the ACRU model accordingly requires a relatively high-level of software engineering skills (Clark, 2011).

APSIM, like DSSAT, is a crop modelling and decision support system consisting of cropping system models, and utility software for managing model inputs/outputs and for additional software development. The crop model itself consists of modules for the growth and development of different types of crops, the soil water balance, soil organic matter and nitrogen, crop residues, soil phosphorous, soil pH, erosion, and crop management. The crop 'Manager' module is perhaps unique in that it allows the user to interact directly with program variables to effect field management operations (sowing, applying fertiliser, irrigation, harvesting) and also to track the values of system variables and calculate additional derived values. These modules are assumed to be entirely self-contained, and can be written in any programming language. Modules communicate with each other via a central simulation engine, using a messaging system. The independence of modules is enforced by only permitting inter-module communication via the central engine. Module development is facilitated in two ways: the APSIM Explorer software provides a programmer-centric approach, and is similar in structure to an integrated development environment (IDE) used for software development (e.g. products such as Microsoft Visual Studio and Borland C++ Builder). Users can also develop modules from a more logical and less technical perspective using APSFront, which provides access to a large library of pre-built functions for weather, soil, crop and management processes and operations. APSIM source code is automatically backed-up and versioned, and each code revision is subjected to validation tests. Software problems are logged and tracked via a web-based tool, and all development time is logged to assist in predicting how long new development tasks will take. The APSIM software development and distribution process is closely managed and centrally monitored. A group of scientists, software engineers and programmers are responsible for APSIM code development and maintenance. Users wishing to develop or modify modules can form partnerships with the APSIM developers. Source code was initially released only following the establishment of a joint development project (Keating *et al.*, 2003), but is now more freely available.

Despite the provision of APSIM Explorer and APSFront, and the support for modules written in any programming language, the development of APSIM modules presents significant technical challenges. These are addressed primarily via a dedicated team of software developers, through whom all model development is conducted. While this approach is undoubtedly effective, it also presents significant costs and is impractical as a general model of crop simulation software development. Certainly, the scope for an individual scientist to experiment with developing and modifying APSIM modules is limited.

The DSSAT v.4 Cropping System Model (CSM) needs to provide for a wide community of scientific contribution, with varying levels of software engineering support. It is likely that the CSM presents a practical compromise between modularity in the software engineering sense and accessibility to scientists wishing to develop models within the CSM framework. This means that the DSSAT CSM is an easier modular framework in which to maintain existing models and develop new ones, than modular models such as APSIM or ACRU.

It can be argued that the Fortran language flourishes in the scientific world ("*FORTRAN continues to be the predominant programming language of simulation modelling in agriculture*" (McCown *et al.*, 1996)) partly, at least, because it is procedural (rather than object- and/or event-based), simple, and has a limited and easily-learnt standard library of functions. This principle could be extended to support and justify the relatively simple implementation of modularity in the DSSAT CSM.

3.3.2 Implementation of modularity in DSSAT 4: the Cropping System Model

The implementation of modularity in DSSAT 4 emerged from the systems and software engineering perspectives previously described.

Fully-featured modularity in DSSAT is implemented via the CSM in DSSAT versions 4 onwards. The immediate impetus to modularise the DSSAT crop simulation models was the desire to develop a system where several crops could use the same simulated soil instance, to facilitate (better) simulated crop rotations (Jones *et al.*, 2003).

The DSSAT journey to modularity can be followed in Porter *et al.* (2000) and Jones *et al.* (2001), and is described in a summarised form in Jones *et al.* (2003). The modules in the CSM adhere to guidelines for modular biological simulation systems suggested by Reynolds and Acock (1997), and follow the initial approach described by Porter *et al.* (2000).

The CSM has a number of formal aims (Jones *et al.*, 2003):

1. Modelling the growth cycles of single crops, and crop sequences and rotations, taking into account weather, genetic traits, and soil water, carbon and nitrogen.
2. Easy incorporation of new software modules that extend DSSAT's crop modelling capabilities, including "other biotic and abiotic factors, such as soil phosphorous and plant diseases" (Jones *et al.*, 2003).
3. Alternative module selection and comparison.
4. Structured integration/interfacing with external programs.
5. Better facilitation of documentation.

The DSSAT CSM draws a particularly strong inspiration from the Fortran Simulation Translator (FST) (van Kraalingen *et al.*, 2003), which provides for a structured and modular simulation environment. The principles underlying the design and structure of the CSM and its modules are discussed in Section 3.3.3. These principles led to the formulation of prescriptive requirements for CSM modules, presented in Section 3.3.4.

3.3.3 Principles underlying DSSAT CSM modules

Narrow criteria for DSSAT modules might refer to those features of DSSAT modules which are purely implementation details – these are listed in Section 3.3.4. Underlying these, and in addition to them, are matters of principle that have been observed during the construction of the CSM and ought to be observed when developing new modules. These are as follows:

Separation of source code along scientific discipline lines

Models represent syntheses of knowledge of aspects of several scientific disciplines including plant physiology, soil physics, atmospheric science, and hydrology. Sections of source code in crop models represent bio-physical processes that fall within one or more of these scientific disciplines. Any particular bio-physical process in a model is usually based on knowledge from only a small number of these disciplines, although some processes and associated calculations span disciplines (such as irrigation).

Separating the source code into separate modules for each plant process in a modular model means that (1) source code statements that represent these distinct bio-physical processes are organised into discrete units, and (2) because each bio-physical process is based on scientific knowledge from a limited number of scientific disciplines, there is a close correspondence between modelled bio-physical processes and scientific discipline areas. This has the effect, as far as possible at least, of separating the source code across scientific discipline lines as well. This allows specialists in those fields to develop sub-models and processes with little or no confusion or distraction from source code belonging to other disciplines (i.e. this approach can "*facilitate interaction of specialists by allowing them to critique and contribute to the development of the modules from their discipline without requiring them to* [fully] *understand all parts of a larger model*", Reynolds and Acock (1997)). This supports collaborative work across discipline boundaries

Explicit definition and measurability of inputs and outputs

It is scientifically attractive and robust for inputs (including constants and input parameters, as well as variable inputs) and outputs to be defined with clear units. In the case of input parameters (such as, in DSSAT, cultivar and species parameters, see Section 3.5.1), the values for such parameters could be either measured directly or in some way derived from measurements. If values can be measured directly, this is considered particularly strong. This means that other scientists can test the algorithm and the underlying hypotheses, and it also means that parameters can be modified in explicit terms for different situations (crops, regions, cultivars, etc.). Where inputs and outputs are well-defined (and measurable), other algorithms with the same inputs and outputs can be tested as alternatives. The significance of measurability of inputs and outputs in modular models is described in Reynolds and Acock (1997).

Use of common components

Common components of modular crop models allow several crop models to share functionality. Simulated processes that are very similar in a number of crop models can be consolidated into single entities (modules) and made available for use by all models. This has two advantages.

Firstly, the processes represented by common modules are always simulated identically in all crop models, enforcing consistency and allowing for meaningful comparison. In the DSSAT CSM, this allows the continuous and consistent simulation of the interaction between the simulated atmosphere, plant and soil, between crop species and across crop sequences and rotations, by using a single soil water balance instance (i.e. a single representation in computer memory) for all cropping seasons.

Secondly, making use of common components can reduce the amount of time and effort spent developing and maintaining models, by avoiding repetition of work for several crops. In crop sequences, it also avoids the effort required in communicating state variable information between each crop's water balance. This aspect is only valuable where a crop modelling system needs to make provision for several crops, such as the DSSAT CSM (and other systems such as APSIM).

Reynolds and Acock (1997) present the common module principle in formal terms, although their principles of additivity and transferability strive toward a somewhat loftier ambition (the development of a generic model for biological systems) than that of DSSAT (the development of a field-level model for a range of different crops).

Adherence to DSSAT conventions

The design and operation of DSSAT modules is subject, informally, to a number of conventions. These are listed in Section 3.5; while not enforced in any way by the system, diverging from these runs the risk of losing some of the advantages of the CSM, such as ease of software maintenance and consistency of common module calculations. Following DSSAT conventions, users familiar with one DSSAT crop will be able to adapt almost effortlessly to using a new DSSAT crop because the operation of the model will be almost identical, potentially saving enormous effort in training time and materials. Not following conventions would make a model instantly less usable, and many of the advantages associated with DSSAT's large and diverse user base would be lost.

3.3.4 Module requirements and implementation

Adherence to basic specifications is what defines a subroutine as a CSM module. Module specification is as follows:

Module independence

DSSAT CSM modules are arranged in a tree-like hierarchical structure in the source code. The requirement that all DSSAT modules should be independent and self-contained means that modules cannot directly call or depend upon other modules at the same or higher hierarchical level in the software. They may call modules at a lower hierarchical level, however.

This is meant strictly in programming terms of calls to modules. A module 'B' may depend on the outputs of another module 'A' further up the program execution hierarchy, but B's outputs are provided as interface variable inputs to A, rather than A calling B directly.

Communication by interface

All communication between modules and the CSM is performed via the subroutine interface. Each module may in turn call sub-modules, which in turn must conform with this requirement. Modules may call non-module subroutines, but the scope of the effects of these subroutines must be limited to the calling module. The structure and operation of the CSM should not be compromised by calling a non-module subroutine. This aspect of modularity complements the separation of modules along scientific discipline lines, as it allows a module to be independently developed and tested by specialists prior to inclusion in the model.

For a new module to replace or exist as an alternative to another module, the interface variables need to be identical as well.

Variables in Fortran are passed by reference (Beebe, 2001), so subroutines have the ability to read and change the value of interface variables. Where a module calculates a value for a parameter variable that is then used elsewhere in the CSM, the module must assign a value to the interface variable. Leaf Area Index, for example, is calculated by a plant module and used to determine water uptake in the 'Soil-Plant-Atmosphere-Module' (SPAM) module.

Persistent self-maintenance of state variables

Each module is responsible for calculating rates of change of, and integrating, its own state variables, the values of which need to persist between subroutine (module) calls.

In a Fortran subroutine, the SAVE keyword ensures that the subroutine will store the values of its local variables between calls.

Use of RUNMODE variable for operating in specific simulation modes

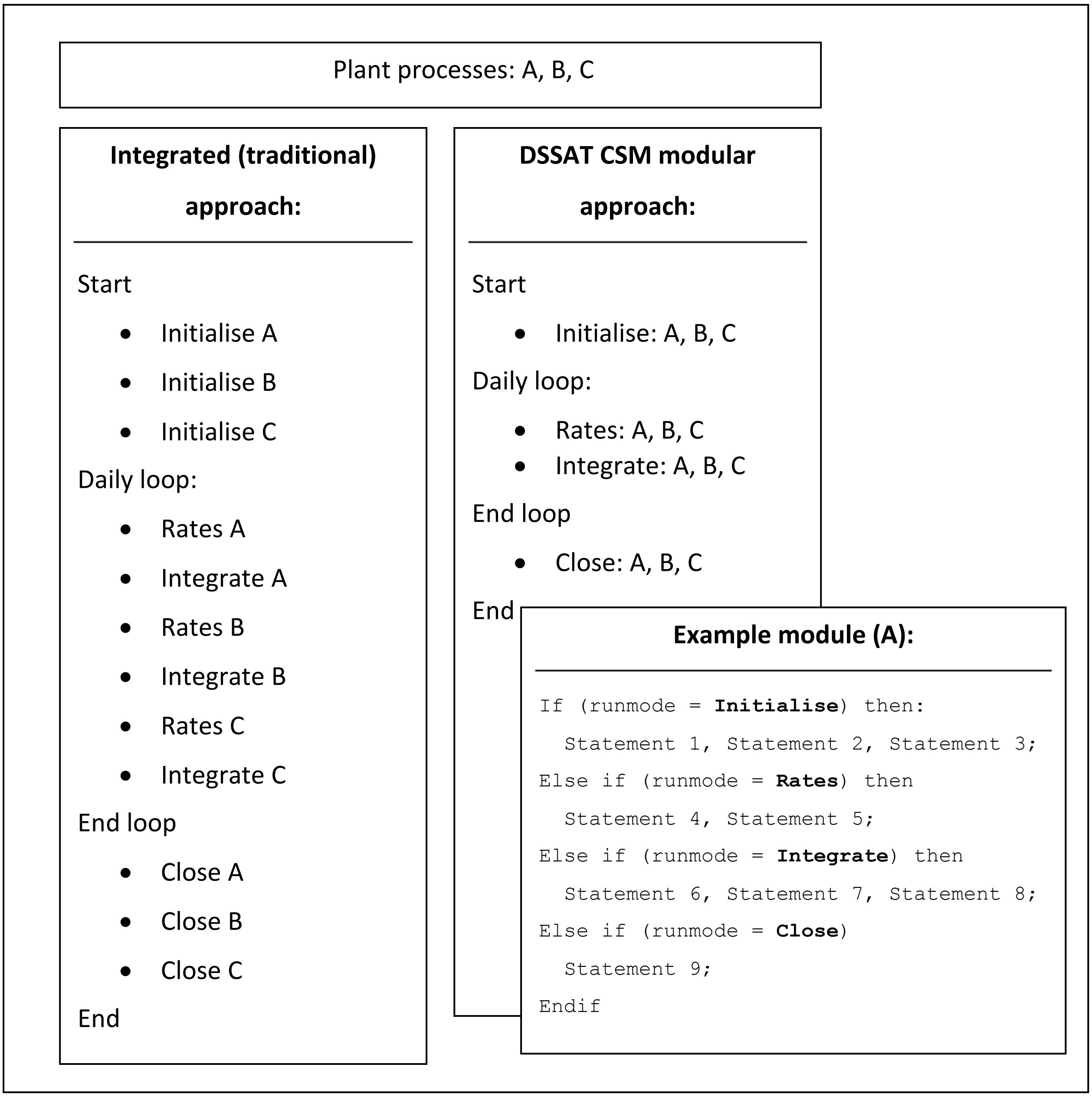
Extending the modularity ideas used in the FST system (van Kraalingen *et al.*, 2003), each module runs in modes which correspond with well-defined stages in model operation. These modes, in DSSAT v.4 are, in order:

1. Simulation initialisation
2. Seasonal initialisation
3. Calculation of rates of change of state variables
4. Integration of state variables
5. Writing of outputs
6. Finalisation

Steps 2-5 are repeated for each season (crop) in the simulation. Steps 3-5 are repeated for each time-step (usually one day) in the simulation. The main (driver) process calls the land module in each of these modes.

Every module is provided with an interface variable called 'RUNMODE'. The values of this variable correspond with the simulation steps described above. When a module is called, it examines the value of RUNMODE and executes the corresponding source code statements. For example, when the module subroutine is called with RUNMODE set to 'SEASINIT', the module is expected to perform appropriate steps to prepare the module for a seasonal initialisation. This might include resetting certain state variables to zero and reading cultivar information.

Figure 3.2 illustrates this method of operation and contrasts it with a non-modular approach.



Plant processes: A, B, C

Integrated (traditional) approach:

```
Start
    • Initialise A
    • Initialise B
    • Initialise C
Daily loop:
    • Rates A
    • Integrate A
    • Rates B
    • Integrate B
    • Rates C
    • Integrate C
End loop
    • Close A
    • Close B
    • Close C
End
```

DSSAT CSM modular approach:

```
Start
    • Initialise: A, B, C
Daily loop:
    • Rates: A, B, C
    • Integrate: A, B, C
End loop
    • Close: A, B, C
End
```

Example module (A):

```
If (runmode = Initialise) then:
  Statement 1, Statement 2, Statement 3;
Else if (runmode = Rates) then
  Statement 4, Statement 5;
Else if (runmode = Integrate) then
  Statement 6, Statement 7, Statement 8;
Else if (runmode = Close)
  Statement 9;
Endif
```

Figure 3.2 Diagram illustrating the difference between an integrated model code structure and the modular DSSAT Cropping System Model (CSM) structure. Three plant processes, A, B, and C, need to initialise, perform rates and integration calculations, and close. In the integrated approach, these are represented as separate groupings of code distributed across the program. In the CSM approach, all code belonging to a plant process is stored within a single module subroutine (for all plant processes), with the 'runmode' indicating whether that module should execute initialisation, rates, integration or closing code (more detail on an example module A is shown in pseudo-code).

Use of the Fortran programming language

Fortran is currently used throughout the DSSAT CSM. Modules are typically implemented as conditional statements that choose which statements to execute based on the value of RUNMODE.

It should be possible to write DSSAT modules in other languages, provided that the module is compiled as a library and that the subroutine calling convention (the way in which variables are stored and translated in the computer's memory between programs) is the same as Fortran. This is certainly the intention of the DSSAT developers (Jones *et al.* 2003). Communication between Fortran and other languages appears to be somewhat standards-based, but also subject to individual Fortran compiler implementations (Beebe, 2001). Use of alternative languages was not a matter of concern within this project, as both DSSAT v.4 and SASRI-Canegro are implemented in Fortran. Code conventions (in both models) follow a Fortran-77 fixed format structure making use of Fortran-90 and -95 features where appropriate.

One of the Fortran-95 features used extensively in the DSSAT CSM is the composite or derived variable. Composite variables were used extensively in the DSSAT CSM Canegro model that was developed during this project, so warrant explanation:

Fortran supports a number of basic variable types – examples include REAL, INTEGER and CHARACTER, corresponding with floating point numbers, whole numbers and character strings.

A composite variable type is simply a collection of these basic types, similar in concept to an array. Unlike arrays, composite variables can store any combination and number of different variable types, in any order. Each element – termed a 'field' – is referenced using the name of the field as a suffix to the name of the instance of the composite variable, separated by a percentage sign as a delimiter. Box 3.1 demonstrates the definition of a composite variable and how it might be used.

Definition of the DSSAT Weather variable type:

```
      TYPE WeatherType

!       Weather station information
        REAL REFHT, WINDHT, XLAT

!       Daily weather data.
        REAL CLOUDS, CO2, DAYL, PAR, RAIN, RHUM, SNDN,
     &    SNUP, SRAD, TAMP, TA, TAV, TAVG, TDAY, TDEW,
     &    TGROAV, TGRODY, TMAX, TMIN, TWILEN, WINDSP

      END TYPE WeatherType
```

Example declaration – creation of an instance of WeatherType, called '*Wth*':

```
      TYPE (WeatherType) Wth
```

Example usage – assigning the value of daily rainfall from *Wth* to a local variable named 'RAINFALL':

```
      RAINFALL = Wth%RAIN
```

Box 3.1 Example definition and usage of the DSSAT CSM Weather variable type, as a demonstration of definition and usage of a Fortran composite variable.

Composite variables facilitate a more object-oriented approach than previously possible with Fortran. Variables common to a particular plant or environmental process in a model can be grouped logically together. Sets of variables, of different types, can be passed to subroutines and functions as a single entity. As only the reference to the composite structure is passed, it is more efficient than passing numerous individual variables.

The DSSAT CSM makes extensive use of composite variables. Examples include variables for storing daily weather data, soil properties and simulation controls.

3.3.5 Additional advantages of modularity in the CSM

The modular approach implemented by the DSSAT CSM offers a number of advantages, to the crop model user, the model developer, and the programmer. These advantages make the DSSAT CSM an attractive platform for crop model development. The DSSAT system is already familiar to large numbers of crop scientists worldwide, and the generic modular approach of the CSM crops make the operation of the system for different CSM crops (and new crops, such as sugarcane in this case) relatively easy. Many general advantages to modularity have already been discussed. This section highlights some key advantages of the DSSAT CSM, in addition to the advantages it enjoys as a modular system.

Ease of software development

The CSM is particularly valuable from a software development and maintenance point of view. It provides a suite of common modules for the soil water balance and related calculations, routines for reading inputs, running simulations and handling errors, as well as a comprehensive set of utility software for creating input files and analysing results. It is also well-documented and -tested. As the CSM provides a framework replete with all simulation controls and complete environmental simulation, it allows the crop scientist to focus almost exclusively on modelling plant processes, rather than being side-tracked by re-implementing necessary auxiliary processes.

Compatibility with future CSM improvements

As new techniques are developed for the modelling of non-plant processes, all CSM crops can take advantage of these without having to implement the change separately for each crop.

Ease of standardisation

Adherence to DSSAT standards and conventions is generally to the benefit of the model user, as the value of DSSAT is not just the crop simulations it can do but also the analysis and visualisation of simulation outputs using the software tools, utilities and documentation that make up the rest of the DSSAT system. Alignment with DSSAT standards is made easier using the CSM as many of the DSSAT conventions are enforced by the CSM, and all functionality (e.g. output) provided by common modules is fully compatible with the complete DSSAT package.

Following the DSSAT CSM conventions means that new models effectively have the same user interface as other DSSAT crops, which dramatically reduces the training burden for new users (and the developers).

Established user community and training facilities

The large number of DSSAT users worldwide, good documentation and an active emailing list make DSSAT an accessible and well-supported system to use. By virtue of having a common set of user interfaces for all crops, experts in one crop can often assist users simulating other crops. Training courses are held annually in the USA, and less frequently in other countries.

3.4 Architecture of the DSSAT CSM

Figure 3.1 provides a diagrammatic representation of the DSSAT system as a whole (for DSSAT v.3.5). Although DSSAT v.4 introduces the CSM with its new modular structure, the architecture of the DSSAT system as a whole remains very similar. The significant difference is that where DSSAT v.3.5 had a separate computer program for each of its crop models, the DSSAT v.4 has, at its heart, one computer program model containing modules for each crop.

The architecture of the CSM is described in this section, but the reader is reminded that the CSM represents only one component of the DSSAT system ('crop models', in Figure 3.1) as a whole.

At its core, the CSM has a 'Land Unit' module. This represents a unit of land which is subject to weather, has a soil and on which a crop may be grown; it can be considered to represent a field. The Land Unit represents a field and interfaces with sub-modules for simulating each component of the field – weather, plant, soil, the soil-plant-atmosphere interface and management (Jones *et al.*, 2003).

The operation of the Land Unit module (and indirectly, all sub-modules) is coordinated by the main program. The main program is responsible for controlling the simulation – including reading the initial experiment parameters from file, starting and ending seasonal simulation runs and setting up a time-step loop within which the land module is called in each run mode (discussed in Section 3.3.3).

3.4.1 Non-plant modules

Non-plant modules in the CSM simulate the environment in which the plant is grown. Included in these are:

Weather module

This module provides weather data to the CSM. It either reads values from files or generates weather data. Weather data are stored as daily values for maximum and minimum air temperature (°C), solar radiation (MJ m^{-2}) and rainfall (mm) as a minimum set; other variables including maximum relative humidity (%) (Jones, 2007) and wind-run (km) can be stored and are read by this module. The weather module can also modify weather data values each day (for climate change/modification investigations) via the ‘weather modification’ module (Jones *et al.*, 2003).

Soil module

The soil module is responsible for simulating various aspects of the modelled soil profile. It consists of four sub-modules: soil water, soil temperature, soil nitrogen and carbon, and the ‘Soil-Plant-Atmosphere-Module’ (SPAM) interface. Soil water and SPAM are discussed here, as DSSATv.4.5 Canegro does not yet use soil nitrogen, soil carbon or soil temperature.

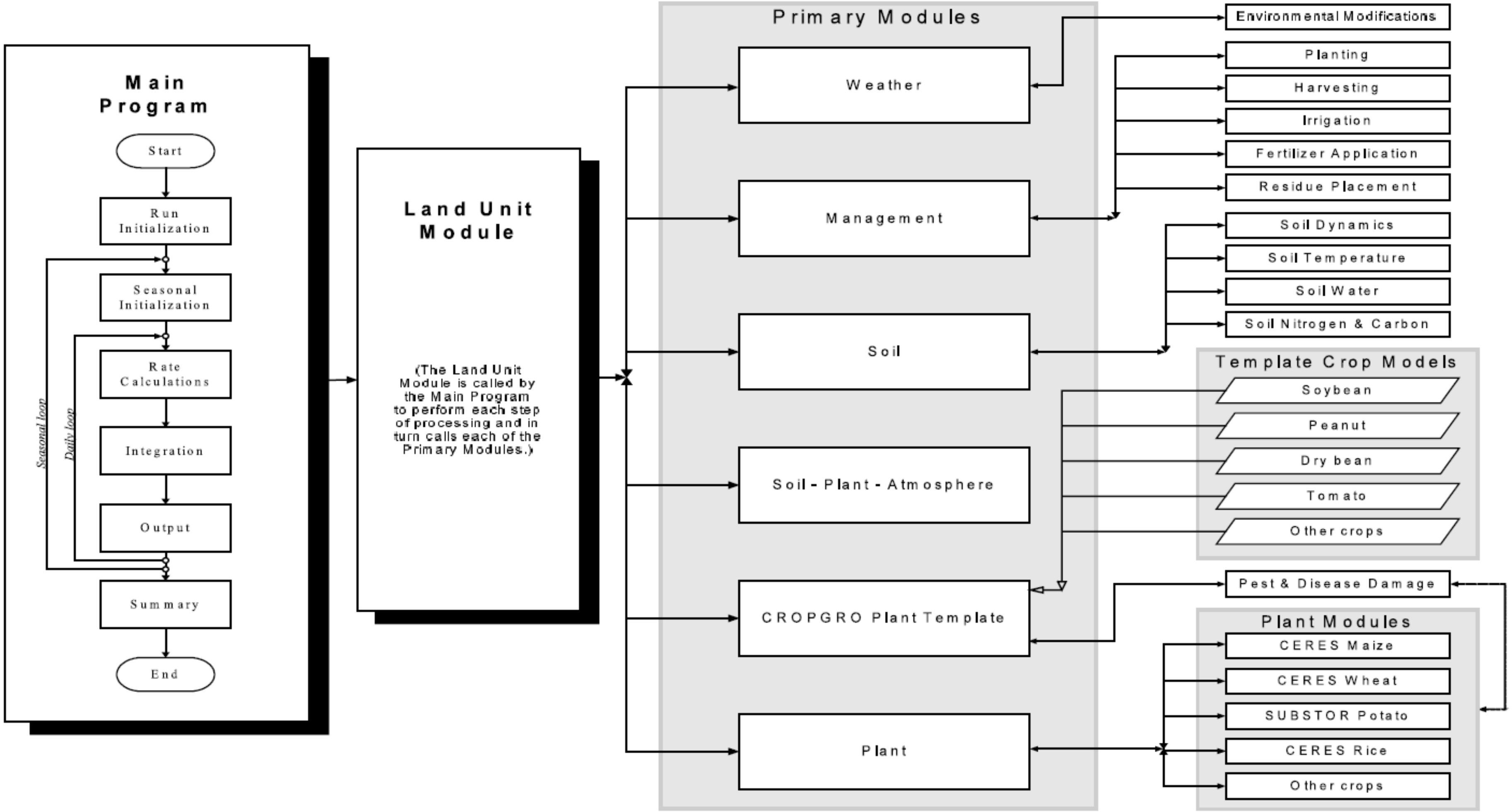


Figure 3.3 Structure of the DSSAT v.4 Cropping System Model (from Jones et al., 2003).

Soil water module

The soil is represented as a profile of homogenous layers. Each layer has attributes associated with its water-holding and water conductivity characteristics, as well as cation exchange capacity, bulk density, clay and silt fractions and soil organic matter content. The soil profile also contains information on soil albedo (surface reflection coefficient) and SCS (Soil Conservation Service, 1972) curve number. For each simulated day, rainfall (and irrigation, if applied) is partitioned into runoff and infiltration. A 'tipping bucket' approach to soil water infiltration is applied: saturated hydraulic conductivity rates are examined to determine the rate of infiltration from a shallower layer into a deeper one, when soil water content in the upper layer exceeds the drained upper limit of that layer (Jones *et al.*, 2003). When the soil water content of the deepest layer exceeds its drained upper limit, drainage may occur if the saturated hydraulic conductivity of the deepest layer permits this. A perched water table can be simulated by defining a soil layer with a very low saturated hydraulic conductivity, effectively preventing water from shallower layers from draining through it. Upward unsaturated flow is also conservatively calculated. Rates of water removal are calculated by the SPAM module and the water content of each soil layer integrated by adding infiltration and removing evapotranspiration (Jones *et al.*, 2003).

Soil-plant-atmosphere module

This, the 'SPAM' module, represents the interface between the soil, the plant, and the atmosphere. It calculates soil surface evaporation and plant transpiration. Potential (maximum unstressed) plant evapotranspiration (*EOP*) is calculated using one of several methods. These include the Priestley-Taylor method (Priestley and Taylor, 1972) and the FAO-56 (Allen *et al.*, 1998) method. *EOP* is partitioned into potential (unlimited) rates of soil surface evaporation and plant transpiration, based on a negative inverse function of leaf area index (m^2 m^{-2}) and a canopy *PAR* extinction coefficient. Actual soil evaporation is limited by the potential soil evaporation and soil water content, and is calculated based on a two-stage process described by Ritchie (1972), which takes into account the wetness of the top layer of soil and the time since the last wetting event. Maximum water uptake from each soil layer is calculated as a function of root length density (cm cm^{-3}) and layer water content, using 'an approximation to the radial flow equation' (Jones *et al.*, 2003). Actual plant transpiration is the minimum of the potential transpiration rate and the maximum root water uptake rate. Plant stress indices are calculated as ratios of actual and potential transpiration, and potential root water uptake and potential transpiration (Jones *et al.*, 2003). Canegro calculates its stress indices on a similar basis.

Management module

The Management module performs such functions as simulating irrigation, controlling planting and harvesting, applying fertiliser and organic material. The user can specify that these operations be performed automatically given certain criteria (e.g. irrigate a certain amount when soil water content falls below a certain threshold) (Jones *et al.*, 2003).

3.4.2 Plant modules

Each plant module is responsible for simulating the growth and development of a crop. These are the only parts of the CSM where source code is crop-specific. Plant modules communicate with the CSM via certain interface variables. These take the form of input variables supplied to the plant module and output variables that are expected to be calculated by the plant module and which are necessary for CSM SPAM and management calculations. Inputs include daily weather data (maximum (*TMAX*) and minimum (*TMIN*) air temperature (°C), daily rainfall (*RAIN*, mm), etc.), soil water content per layer (*SW*, $cm^3\ cm^{-3}$) and simulation settings. Outputs include leaf area index (*XLAI* (green and senesced leaves) and *XHLAI* (green leaves only), $m^2\ m^{-2}$) and root length density (*RLV*, $cm\ cm^{-3}$) for calculating transpiration, crop growth stages for calculating harvest dates (*STGDOY*), and soil nitrogen uptake (*UNH4* and *UNO3*, $kg\ ha^{-1}\ day^{-1}$) (Jones *et al.*, 2003).

Plant modules are defined in one of two ways: as individual crop modules, or as CROPGRO Template modules. The focus in this thesis is on individual crop modules, as this is how Canegro was implemented within the CSM.

CROPGRO Template plant modules

The CROPGRO template modules are based on earlier work described by Boote *et al.* (1998); essentially, as the BEANGRO (Hoogenboom *et al.*, 1994) and PNUTGRO (Boote *et al.*, 1987) models were developed from SOYGRO (Wilkerson *et al.*, 1983), the similarities in model structure were such that it was possible to use a single software program to model all of these crops. Each individual crop is represented as a collection of parameters which are used in equations inside the model to configure the program to model a particular crop. CROPGRO became a ‘generic’ crop model (Jones *et al.*, 2003). In DSSAT, this approach is extended: species coefficients are defined in a Species file (see Section 3.5.1); and cultivar coefficients are defined in Cultivar and Ecotype files. The CSM chooses the appropriate set of parameters with which to configure the CROPGRO Template module before simulation commences (Jones *et al.*, 2003). The development of a sugarcane model using the CROPGRO template approach is feasible (CASUPRO (Villegas *et al.*, 2005) is implemented this way). Converting Canegro to use this approach is not, however (it would be a different model), and as the SASRI intention was to use the CSM version of Canegro model as a development platform, a dedicated plant module is the more flexible and more appropriate choice.

Individual Crop plant modules

The use of individual crop plant modules involves having a dedicated set of (autonomous and independent) modules that simulate the crop growth and development aspects of each crop. Each crop is represented at the top level by a single plant module. This module can then call sub-modules for performing crop-specific simulation calculations. A number of DSSAT crops are implemented this way; examples include Maize, Wheat and Barley. The CSM communicates with the 'Individual Crop Module Interface'. This module then chooses which specific plant module to call in each RUNMODE (see Section 3.3.4); the choice is based on Crop Code (e.g. 'MZ' for Maize) and model name (in DSSAT v.4.5 onwards, Porter, 2007).

This approach is well-suited for conversion of existing standalone models to the modular format, and was used in the construction of the DSSAT CSM Canegro module. The process necessary to achieve this is described in the methodology chapter.

3.5 DSSAT v.4 Conventions

DSSAT v.3 allowed consistent access to a wide range of crops and functionality by ensuring that each model adhered to certain input and output file formats and conventions (Jones *et al.*, 1998). In DSSAT v.4, these conventions are partially enforced automatically because common subroutines are used by all crops for reading most input files (for example, the experiment file, weather data files and soil description files) and generating most output files.

Adherence to certain DSSAT conventions was expected of the DSSAT-Canegro CSM Plant Module. These are discussed below. This is not an exhaustive list of DSSAT conventions, but rather those that were necessary for this project.

3.5.1 Genetic files

Plant modules in DSSAT (the crop-specific modules) perform calculations with respect to different plant processes; these include phenology, photosynthesis, biomass partitioning and so on. Each of these sub-processes relies on one or more numeric parameters which have been derived through experimentation to apply to that crop for that process. Certain of these parameters have been found to apply to all cultivars of a crop species. DSSAT considers these 'species coefficients'. Other parameters are known to be different between cultivars of that species, and these are termed 'cultivar coefficients'.

Species parameters are stored in the 'Species file' in the DSSAT 'Genotype' directory. The file naming convention is 'crop-code version model.spe'. Species coefficients for Maize, for example, will, for DSSAT version 4, be found in the file 'MZCER040.SPE' ('MZ' = Maize, 'CER' = CERES-Maize, 040 = 'version 4.0'); this file, in a typical DSSAT installation, would be found in C:\DSSAT4\Genotype\. Parameters are stored in 'parameter_name = parameter_value' pairs, one on each line of the file.

Cultivar files have the extension 'CUL' rather than 'SPE', and each cultivar is represented by a single line of parameters. Each line contains the cultivar's unique code, name, its associated 'ecotype' name (discussed shortly), and a set of cultivar-specific coefficient values. Data are stored as formatted text.

DSSAT also makes use of 'ecotype' coefficients. Ecotypes represent sets of parameters that are the same for groups of cultivars. They were, initially, introduced to simplify the process of working with cultivar parameters (Jones, 2006). The file extension is 'ECO' and each line of coefficient values represents an ecotype definition, identified by an ecotype code. The cultivar file references a set of ecotype parameters by association with the ecotype name.

3.5.2 Variable codes

DSSAT input and output files are characterised by formatted-text columns of data headed by data codes. The meaning (definition and units) of each of these codes is expected to be available in the file 'DATA.CDE', located in the DSSAT root directory (usually C:\DSSAT4\).

As a sugarcane model already existed in DSSAT as of v.3.1, much of the DSSAT infrastructure for supporting this crop was already in place. This included data definitions and some sugarcane-specific inputs in the experiment file and Xbuild user interface (e.g. ‘Ratoon’ as a planting type option).

Some of the sugarcane-specific variables already defined in DSSAT v.4 (for the Canegro v.3) include ‘SUAD’ for sucrose mass (kg ha^{-1}), ‘SUID’ for sucrose harvest index (kg kg^{-1}), ‘SUW_1’ for sucrose weight of the first stalk (g stalk^{-1}) and ‘T#AD’ for tiller population (stalks m^{-2}).

3.5.3 Crop codes

Crop codes are associated with species. The file DSSATPRO (with the extension matching the DSSAT version number, e.g. v40 for version 4.0) in the DSSAT root directory lists the association of crops (by code) with executable program files. This makes provision for support of legacy models (such as DSSAT-Canegro v.3 and the Pineapple model). As of DSSAT v.4.5 (Porter, 2007), crops can be associated with multiple models/plant modules, to facilitate the inclusion of several modelling approaches for a single crop.

A crop code (‘SC’) had been established for sugarcane with DSSAT Canegro v.3.

3.5.4 Crop sequences and rotations

DSSAT simulates the soil in a field, and simulates a crop on that field when instructed to do so. One of the great strengths of the DSSAT CSM is that soil processes are simulated continuously for a sequence of crops. In the case of crop sequences or rotations (possibly involving more than one crop species), the water and nutrient status of the soil at the end of one crop is carried to the start of the next crop. Any intermediate days are simulated as well, with a ‘fallow’ crop.

3.6 DSSAT-Canegro v.3

Canegro was modified to be compatible with the DSSAT system and became the sugarcane model provided with DSSAT versions 3.1 to 4.0.3 (Inman-Bamber and Kiker, 1997). This version of Canegro included a rudimentary nitrogen model (O’Leary *et al.*, 2000), and a small number of cultivar inputs. The changes made to this version of Canegro were never carried back to the original version of the model. DSSAT-Sugarcane was not developed further, whereas substantial additions have since been made to the standalone SASRI Canegro. The DSSAT CSM has also been developed since; the DSSAT v.3 sugarcane model cannot make use of the wide (and increasing) range of functionality offered by the modular

DSSAT CSM. A workshop was held in 2000 to train users in the operation of Canegro and to discuss different aspects of sugarcane modelling and Canegro in particular (O'Leary and Kiker, 2001). A number of shortcomings were identified in the DSSAT version of Canegro and several suggestions were put forward for improving and upgrading of the DSSAT version of Canegro, many of which were addressed in this project.

4. METHODOLOGY

4.1 Introduction

This chapter describes the steps and processes that were followed in creating a DSSAT CSM version of the Canegro simulation model. To simplify the explanation of the process, Section 4.2 presents an overview of the methodology followed.

Thereafter, from Section 4.3 onwards, the methodology followed is described in detail.

4.2 Overview of Methodology

The process of creating a DSSAT v4 Cropping System Model (CSM) version of Canegro required several distinct steps:

- Statements relating to plant growth and development in the SASRI Canegro source code were identified. These source code statements were consolidated into prototype CSM modules representing plant growth and development processes. These modules were tested in a modified version of SASRI Canegro.
- A new DSSAT CSM plant module was created, and the Canegro plant process modules were added to this as sub-modules.
- The new code in the Canegro CSM module was verified (tested), and simulation output discrepancies between this and the standalone Canegro were analysed and explained using sensitivity analysis.
- Support for crop sequences was added to the Canegro CSM plant module and DSSAT genotype files were created.
- The new model was validated by statistically and graphically comparing Canegro CSM simulations with observations from experiments.

The remainder of this chapter describes in detail the methodology followed in performing these tasks and thereby creating the DSSAT CSM Canegro model.

4.3 Preparation of the Canegro Source Code for Modularisation

The Canegro source code was prepared by adding explicit declarations and definitions for variables, replacing common code blocks (discussed in Section 2.4.2) with Fortran composite variables (Section 3.3.4), and adding code comments and explanations. The code was then analysed to identify and isolate code statements belonging to each plant growth and development process. Prototype DSSAT CSM modules representing these plant processes were created from these code statements, which were then tested in a modified version of the SASRI Canegro model before being incorporated as sub-modules into the DSSAT CSM as part of the Canegro CSM Plant Module.

Verification of the integrity of the Canegro CSM source code required that the simulation of the water balance could be disabled in the SASRI Canegro model, and such a facility was developed.

The sub-sections below discuss these steps in detail.

4.3.1 Variable identification and declaration, and removal of common code blocks

Declaration of variables

Very few variables in the SASRI Canegro source code had been explicitly assigned a data type – REAL, INTEGER, LOGICAL, etc. An Excel spreadsheet was developed that used the Fortran implicit typing rules (e.g. variables starting with A-H and O-Z are considered REAL values, and those starting with I-N are of type INTEGER) to explicitly declare all variables in the SASRI Canegro model..

Definition of variables

The meanings of variables were determined using the following approaches:

1. By examining literature (such as recognising terms in equations that appeared in the code), model documentation and through discussion with SASRI colleagues.
2. From code comment explanations and by inferring from comment explanations of how variables were used to model biophysical processes.
3. By inferring from the units and meanings of adjacent variables in equations using variables defined using the steps described above.
4. By examining DSSAT definitions of variables with similar names. This was generally successful, because of the similar pedigree (CERES-Maize and CropGro models) of both DSSAT CSM and Canegro.

All rate and state variables were identified using these techniques, and extensive explanations were added as code comments.

Replacement of common code blocks with Fortran composite variables

It was necessary to remove the use of common code blocks, because this conflicts with the requirement that DSSAT modules must communicate by interface only (discussed in Section 3.3.4). Fortran composite variables were created to represent each of the common code blocks used in the SASRI Canegro model.

4.3.2 Plant growth and development process code identification

The identification of plant growth and development code statements in SASRI Canegro was started by examining the daily loop, where new values of state variables are calculated for each day of the simulation.

The following plant growth and development processes were identified, listed in order of execution:

1. Root growth and development (defined in the water balance code)
2. Stalk population
3. Canopy development
4. Photosynthesis and respiration
5. Biomass partitioning

Each of these plant processes was originally represented by a single subroutine in the SASRI Canegro source code.

4.3.3 Identification of input and output variables in subroutines

Key input and output variables were identified in each plant process subroutine by examining the subroutine interfaces and use of common blocks.

4.3.4 Identification of initialisation source code for each plant process

Initialisation statements of variables used in plant process subroutines appeared before the daily loop section in the SASRI Canegro source code. By tracing the inputs and outputs of each plant process subroutine through the program source code, corresponding initialisation statements were identified.

It was noted that in some instances a variable initialised once was used by several subroutines. In such cases, the initialisation of the variable was considered to be the responsibility of the first subroutine to use that variable.

4.3.5 Creation of plant process modules

A DSSAT CSM plant process sub-module was created from each of the plant growth and development subroutines identified in Section 4.3.2. The following actions were required:

- The calculation of thermal time and other phenological variables was performed in the main body of the SASRI Canegro daily loop. These code statements were moved into a new sub-module in the DSSAT CSM (SC_PHENOLOGY).
- The IMPLICIT NONE statement was added to the module subroutines to force the compiler to require explicit variable declarations.
- Where a plant process subroutine in SASRI Canegro made use of a variable defined in a common block, the block's equivalent Fortran composite variable was added to the CSM plant process module version of that subroutine, and made an input/output interface parameter.
- Initialisation statements identified in each plant process subroutine (described in the previous step, Section 4.3.4) were transferred to the initialisation sections of the equivalent new plant process CSM modules.

As each original subroutine was called in the daily loop, the remainder of the code in each module needed to be allocated to 'rates', 'integration' or 'output' modes (model operations that happen for each day of the simulation). It was noted that in most cases that the calculation of rates and states was not separated in SASRI Canegro. Upon the advice of Jim Jones (2006), where this was the case, these code statements were simply allocated to 'rates'.

According to the specification for DSSAT CSM modules (see Section 3.3.4), these modules:

1. Saved the values of their own state variables between subroutine calls.
2. Operated in initialisation and rates/integration modes.
3. Explicitly-declared and defined all of their variables.
4. Communicated by interface only, meaning that they could not use common blocks, but instead used composite variable equivalents passed via the subroutine interface.

4.3.6 Creation of an intermediate version of SASRI Canegro for testing modules

Given the complexity of the SASRI Canegro source code, and the interconnected nature of dynamic models, it was decided that the functionality of each plant process module should be verified before adding the modules to the DSSAT CSM. An intermediate version of the SASRI Canegro model was created that was able to use these modules in a limited way to facilitate this testing.

The new CSM modules could not make use of common block variables. As only the plant growth and development aspects of SASRI Canegro were to be modularised and tested, it was unnecessary to replace common blocks with composite variables throughout the model. Doing so might also have increased the chances of a coding error. Instead, the intermediate SASRI Canegro model was modified to transfer the values of common block variables to equivalent composite variable fields before calling a module. The composite variables were passed into the modules as subroutine variables. After calling a module, the values of fields within these composite variables were then copied back to the common block variables after the call to the module. This process is illustrated in Figure 4.1.

The calls to original subroutines representing plant processes in the daily loop, and the initialisation code statements corresponding with those subroutines, were removed from the intermediate model's source code.

At least two calls to each plant process module were inserted into the intermediate model. The first was inserted into the pre-daily loop initialisation section, and this was called in 'initialisation' mode. The second was inserted into the daily loop section of the Canegro code, such that the module subroutine was called in 'rates' mode for each day of the loop. If rates and integration were obviously separated, an additional call was added in 'integration' mode.

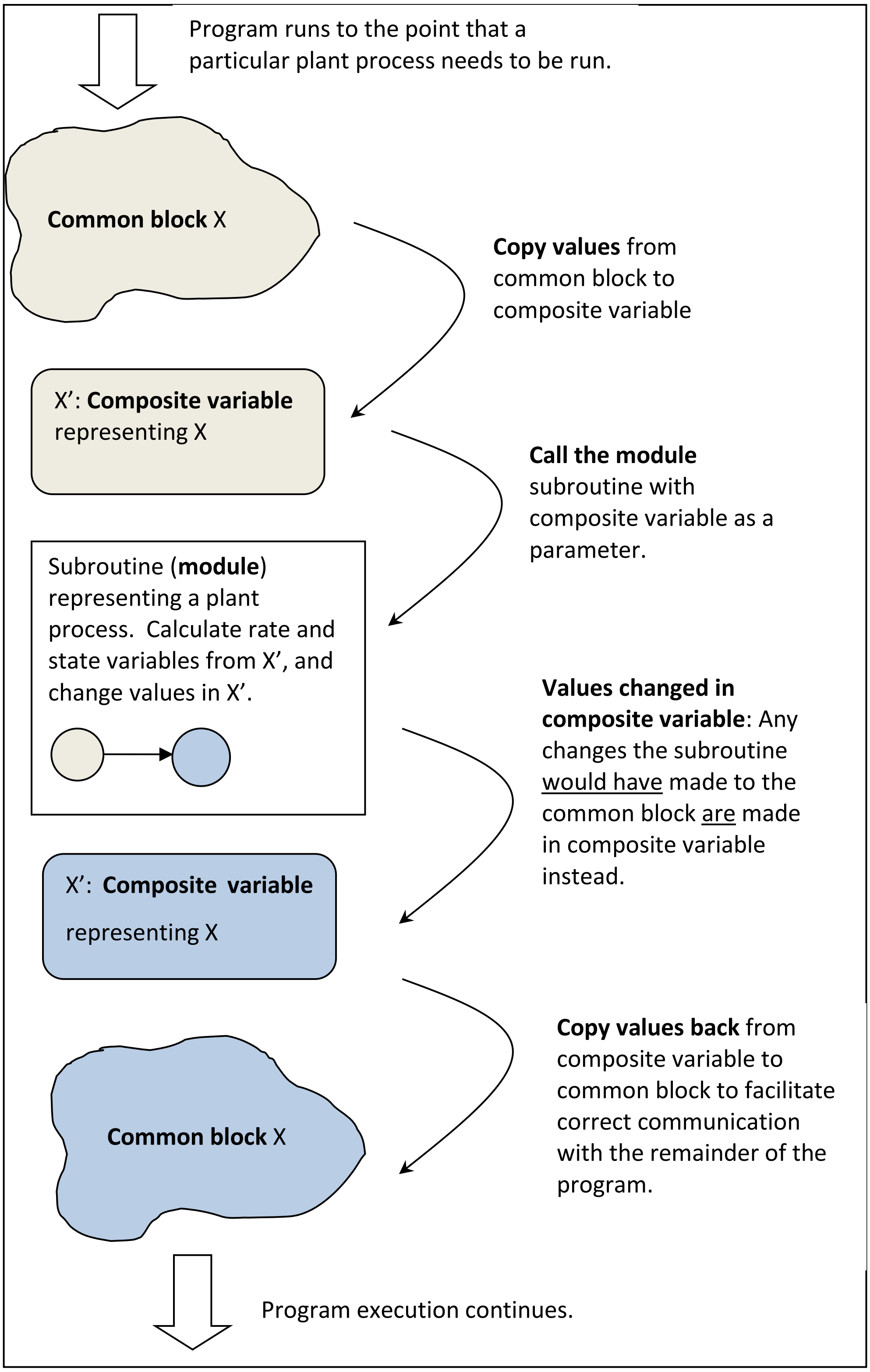


Figure 4.1. Diagrammatic representation of the process by which the values of common code blocks are copied to composite variables prior to the execution of a module, after which the values within the common blocks are updated with any changes that might have been made to the composite variables within the module. This process allows an intermediate module to be tested in SASRI Canegro without common blocks being used in the module subroutine.

This approach allowed an entity equivalent to a DSSAT CSM module to be tested within SASRI Canegro itself, one module at a time, keeping the rest of the SASRI Canegro functionality identical. By comparing the output from several runs performed with SASRI Canegro and this intermediate version of the model, it was possible to verify whether or not an intermediate module behaved exactly the same as the original subroutine and its initialisation statements. This verification process also served as a trouble-shooting process, occasionally highlighting initialisation code that had been omitted, and similar inconsistencies.

When each module appeared to be functioning correctly, it was copied to the CSM-Canegro plant module (the development of which happened in parallel and is described in Section 4.4).

4.3.7 Identification and correction of software bugs

The process of analysing Canegro and explicitly typing its variables led to the discovery of logical inconsistencies and other bugs in the code. Corrections were made both in the DSSAT CSM-Canegro and in the (otherwise unchanged) version of SASRI Canegro that was used for code verification purposes. This ensured that the same entity was being compared during verification. For model validation, however, the uncorrected version of the SASRI Canegro model was used.

4.3.8 Disabling the water balance

It was necessary to provide a mechanism for disabling the water balance processes in SASRI Canegro to facilitate verification of the Canegro code in the DSSAT CSM. While the DSSAT CSM provides an input simulation switch to turn the water balance on or off, an equivalent feature needed to be implemented in SASRI Canegro. The influence of water in SASRI Canegro was disabled by:

- Initialising the soil water content to drained upper limit (DUL, $\text{cm}^3\ \text{cm}^{-3}$) for each layer, at the beginning of the simulation.
- Initialising stress variables ($SWDF_1$, $SWDF_2$, and $SWDF_{30}$) to 1 at the beginning of the simulation.
- Preventing the call to the WATBAL (water balance) subroutine from the main daily loop process.

An input configuration switch was added to instruct the model when to enable/disable the water balance.

4.4 Development of the DSSAT CSM Canegro Plant Module

The DSSAT CSM provides a programming interface to single-crop models implemented as CSM Plant Modules. Several crops use this approach (this was discussed in detail in the DSSAT chapter (Section 3)). Each plant module represents only the plant growth and development aspects of the simulation. Simulation of processes such as transpiration, soil surface evaporation and the soil water balance is performed by CSM modules common to all crops.

This section describes the creation of the DSSAT CSM Plant Module version of the SASRI Canegro model.

4.4.1 Creation of DSSAT CSM Canegro plant module framework

A large number of variables are passed from the CSM to each plant module. These include inputs, such as daily weather variables and simulation configuration information (e.g. whether or not the water balance is simulated; treatment/rotation numbers; whether to operate in ‘initialisation’ or ‘rates’ mode). Certain variables are changed within the plant module and communicated back to the CSM – these primarily relate to the water balance, and include state variables such as root length density and leaf area index, as well as configuration variables such as the crop factor for transforming short grass reference evaporation into sugarcane reference evaporation, and the canopy *PAR* extinction coefficient.

Writing a plant module interface from scratch would have been possible but rather difficult, particularly for a first-time developer of a plant module. Rather than attempting to do so, the subroutine interface of the CERES-Maize module was copied and then modified for sugarcane.

The top-level CERES-Maize plant module code file (MZ_CERES.FOR) was opened and the MZ_CERES subroutine code copied and pasted into the new DSSAT CSM Canegro file (SC_CNGRO.FOR). The subroutine was renamed to SC_CNGRO. Most of the code was then deleted from the body of the new subroutine, leaving only the variable declarations and the RUNMODE structure (i.e. the conditional statements to operate the module in initialisation/rates/etc. modes) in place. The SC_CNGRO.FOR file was added to the build list in the compiler software (Compaq Visual Fortran originally; then Intel Visual Fortran).

A call to the SC_CNGRO subroutine was added to the CSM PLANT module. This module, defined in PLANT.FOR, is responsible for linking CSM to an individual crop's plant module. The call to the CERES-Maize module was copied and modified to call the SC_CNGRO subroutine when the 'SC' crop code is matched.

When executed, simple WRITE statements in each of the RUNMODE sections of the SC_CNGRO module verified that the new plant module was being called.

4.4.2 Linking to Canegro-equivalent input/output variables in the DSSAT CSM

The MZ_CERES module provided ready-declared variables and subroutine arguments for key inputs required by the plant module and key outputs required by other DSSAT CSM modules. Inputs included DSSAT CSM composite variables describing the soil (*SOILPROP*), simulation control (*CONTROL*), simulation configuration (*ISWITCH*) and weather data (*WEATHER*), and an array for soil water content per layer (*SW*). Outputs included canopy height (*CANHT*), canopy PAR extinction coefficient (*KCAN*), crop reference evaporation coefficient (*KEP*), root length density (*RLV*) and green (healthy) leaf area index (*XHLAI*).

The Canegro plant module required total root water uptake (*TRWU*) and total potential root water uptake (*TRWUP*) in order to calculate its stress indices. These variables were available in the PLANT module, and were added to the subroutine parameter list.

The intermediate version of Canegro made use of composite variables to store key input and state variables. The definitions of these composite variables were placed in SC_CNG_mods.FOR and added to the compiler build list. Declarations for these composite variables were added to SC_CNGRO. Statements were added at the top of the SC_CNGRO module to copy values from the DSSAT CSM composite variables to the Canegro ones that are used internally. The last lines of the SC_CNGRO subroutine copy the values from the composite variables back to their CSM equivalents. This was easier than attempting to rewrite the modules to use the DSSAT CSM composite variables directly.

Additionally, where required, variables were transformed if units in Canegro were different to those of DSSAT. For the vast majority of variables, no such transformation was necessary.

4.4.3 Adding a reference evaporation crop coefficient (*KEP*)

The DSSAT CSM required that *KEP* be set within the Canegro plant module. *KEP* is the canopy evaporation coefficient to convert short-grass reference evaporation into sugarcane reference evaporation.

4.4.4 Adding plant process sub-modules to the CSM-Canegro framework

Calls to each of the plant process modules developed in the intermediate SASRI Canegro model were added into SC_CNGRO for each RUNMODE – run initialisation, seasonal initialisation, rates, integration and output.

4.5 Support for Cropping Sequences

Adding support for sequential model runs involved giving consideration to variable initialisation, ratoon crops, and also required additional specific verification steps.

The DSSAT CSM runs sequences within a single execution of the program, so rate and state variables (and some local subroutine variables) in each module needed to be explicitly re-initialised in the module's seasonal initialisation section.

This was verified by instructing the DSSAT CSM to run two or more identical simulations as independent treatments as a batch (rather than as a sequence). This should produce identical result sets, so any discrepancies in output could be attributed to variables not be re-initialised.

Where necessary (in sequence and crop rotation runs), appropriate variable values needed to be carried over from the end of one simulation to the beginning of the next. In the case of sugarcane, the primary issue was identifying whether or not a ratoon crop followed a previous (plant or ratoon) crop. It was decided, in consultation with SASRI colleagues, that root depth needed to be carried over from end of the previous crop to the start of the new crop in these circumstances.

The following five conditions were identified for the DSSAT CSM Canegro model to be convinced that a crop is a ratoon crop following a previous crop in a sequence:

1. The simulation type must be 'sequence' or 'rotation'. This corresponds with values of the RNMODE variable (a field of the CSM Control composite variable) of 'Q' and 'F' respectively.

2. There must have been a previous crop run, and its treatment number must be the same as this one.
3. The previous crop must be sugarcane, and the same cultivar of sugarcane.
4. The crop must be a ratoon.
5. The crop start date must be within two days of the harvest date of the previous crop.

The *CARRYOVER_RATOON* subroutine sets the logical value *CARRY_OVER*, in the *CaneCrop* composite variable, to TRUE if these conditions are met and FALSE otherwise.

Additionally, it was assumed that if three of these conditions are met, then there is a reasonable likelihood that the user intended the crop to follow on. A function was added to the Canegro CSM module to issue a warning into the standard DSSAT run warning file (WARNING.OUT) in such situations. This warning indicates that the ratoon information was not carried over, and also lists the reasons why not. This is to enable a user to modify the run setup if a consecutive ratoon crop was intended but one of the conditions was not met.

4.6 Creation of Species, Cultivar and Ecotype files

Species, cultivar and ecotype files were created for the CSM Canegro model. This involved identifying coefficients in input files and hard-coded in model equations. These were then assigned to species or non-species categories. The non-species parameters were then divided into cultivar parameters and ecotype parameters. Details of the identification and categorisation of parameters are discussed in the next sub-sections.

These parameters were then copied into files following the format of DSSAT genotype files. Two modules were written for reading cultivar/ecotype coefficients (GET_CULTIVAR_COEFF) and species coefficients (GET_SPECIES_COEFF). These modules read the genetic files at seasonal initialisation and the store the parameters in arrays. When another module requests a genetic coefficient by name, the module searches for the coefficient name and returns the matching value from the array.

A full list of cultivar, species and ecotype parameters is provided in the results section of this thesis.

4.6.1 Identifying genetic parameters/coefficients

SASRI Canegro used a large number of coefficients, in every sub-process within the model. Many of these parameters were provided via configuration files. The remainder were hard-coded in model equations.

SASRI Canegro read a settings file which referred to a variety (cultivar) number, and the name of a variety coefficient file. The variety number referenced sets of hard-coded values for a selection of cultivars. The variety parameter file contained additional variety-specific coefficients.

Identification of parameters/coefficients was done as a two-pass process. First, the parameters referenced from the settings file were listed, associated with internal variables and equations in the model, and defined. Second, the CSM Canegro source code was examined for remaining numeric coefficients in equations. These coefficients were replaced with variables, and were similarly listed and defined. All coefficients were initially considered cultivar parameters.

4.6.2 Differentiating between different categories of coefficients

The author's colleagues at SASRI were consulted as to which parameters ought to be considered in each category. Initially, a single division between cultivar and species coefficients was made. Parameters were associated with plant processes, and those processes known to be different between sugarcane varieties were assigned to the cultivar parameter category. The remaining parameters were assigned to the species parameter category.

The sheer number of coefficients eventually demanded that the cultivar file be divided into separate cultivar and ecotype parameter files. Ecotype parameters were identified as those parameters whose values could be shared between groups of related cultivars.

4.7 Code Verification

Code verification refers to the process of ensuring that the computer code is logically consistent and operating as intended (Jones and Luyten, 1998). Two levels of verification were performed.

The first was to ensure that the Canegro algorithms in the CSM Plant Module were operating identically to the algorithms in SASRI Canegro model. The first level of verification required that both models ran with their water balances disabled, to verify that the plant growth and development processes were operating identically, and involved:

1. Ensuring that the individual plant process sub-modules in the CSM plant module version of Canegro operated identically to the corresponding collection of initialisation and plant process subroutine code statements in Canegro. This is termed 'per-module' verification and this testing was conducted during the development of each plant process module.
2. Ensuring that the behaviour of the DSSAT CSM version of Canegro as a whole matched that of Canegro. This is termed 'complete verification' because it tested the relationships between the plant process modules.

The second level of code verification examined the performance of DSSAT CSM Canegro simulations compared to SASRI Canegro simulations when the non plant-related aspects of the models were operational. This explored the influence of differences in the calculation of variables related to the water balance – reference evaporation, infiltration of rain and irrigation, saturated and capillary flow between soil layers, runoff, soil surface evaporation, transpiration/root water uptake and drainage.

The process followed to identify and explain (and where possible, correct) differences is described in Section 4.7.6.

4.7.1 Statistical measures of similarity

Similarity of model estimation behaviour was quantified by root mean squared error (RMSE), average prediction error (APE), and R^2, slope and intercept of linear regression between corresponding sets of output simulated by SASRI Canegro and DSSAT CSM Canegro. RMSE, R^2, slope and intercept were chosen because they were used to evaluate the performance of the SASRI Canegro model by Singels and Bezuidenhout (2002).

The root mean squared error (RMSE) and average prediction error (APE) for a particular variable were calculated according to Eqns (53) and (54):

$$RMSE = \frac{1}{N}\sqrt{\sum_{event=1}^{N}(DSSAT_{event} - Canegro_{event})^2} \tag{53}$$

$$APE = \frac{1}{N} \sum_{event=1}^{N} (DSSAT_{event} - Canegro_{event}) \tag{54}$$

where *event* refers to the simulated value of a given variable on a given day as simulated by DSSAT CSM Canegro ($DSSAT_{event}$) and SASRI Canegro ($Canegro_{event}$), and N is the number of events (i.e. all days in all simulated treatments).

In all cases, the SASRI Canegro result set was considered the independent (x-axis) variable.

4.7.2 Variables examined for each of the subroutines during ongoing verification

Variables examined during the development and on-going verification of the DSSAT CSM Canegro are listed in Table 4.1.

Table 4.1 Description of variables used in the ongoing verification of DSSAT CSM Canegro.

Variable name and meaning	Calculated in (plant process and subroutine name)	Used in (plant process and subroutine name)
STKPOP (shoot population, shoots ha^{-1})	Shoot population (POPLT3)	Canopy (CANOP3)
LAI (green leaf area index, m^2 m^{-2})	Canopy (CANOP3)	Water balance (WATBAL)
TLAI (living + dead leaf area index, m^2 m^{-2})	Canopy (CANOP3)	Water balance (WATBAL)
LI (fractional interception of photosynthetically-active radiation)	Canopy (CANOP3)	Photosynthesis (PHOTOS)
PER (plant extension rate, mm d^{-1})	Canopy (CANOP3)	Canopy (CANOP3)
PARCE (radiation use efficiency, g MJ^{-1})	Photosynthesis (PHOTOS)	Photosynthesis (PHOTOS)
GROSSP (gross photosynthesis rate, t ha^{-1} d^{-1}	Photosynthesis (PHOTOS)	Photosynthesis (PHOTOS)
DWDT (daily change in net photosynthesis, t ha^{-1} d^{-1})	Photosynthesis (PHOTOS)	Biomass partitioning (PARTIT)
GRORT (daily change in root mass, t ha^{-1} d^{-1})	Photosynthesis (PHOTOS)	Root growth allocation (within WATBAL in Canegro)
RLV (root length per volume of soil, termed 'root length density', per layer, cm cm^{-3})	Root growth (WATBAL)	WATBAL (for water uptake calculations)
ROOTDM (root mass, t ha^{-1})	Biomass partitioning (PARTIT)	Respiration (PHOTOS) and output (OUTPUT).
STKMAS (stalk mass, t ha^{-1})	Biomass partitioning (PARTIT)	Respiration (PHOTOS) and output (OUTPUT)
AERLDM (above-ground	Biomass partitioning	Respiration (PHOTOS) and

dry biomass, t ha^{-1})	(PARTIT)	output (OUTPUT)
SUCDM (sucrose mass, t ha^{-1})	Biomass partitioning (PARTIT)	Output (OUTPUT)

4.7.3 Implementation of per-module verification

Similarity was established first visually via graphs of the variables examined and then statistical measures of similarity were calculated. Very small discrepancies were attributed to differences in rounding of the input and output variables.

4.7.4 Complete model verification with the water balance disabled

The next step was to verify that all the plant process modules working together operated correctly. The water balances were disabled in both models and after executing the verification simulation runs, all verification variables (listed in Table 4.1) were compared.

4.7.5 Complete verification with the water balance enabled

The second level of verification was to assess and explain the differences in simulation behaviour between Canegro and DSSAT CSM Canegro, with the water balances enabled. The purpose of this verification was to ensure that differences in simulation behaviour between the models could be attributed solely to differences in the water balance simulations, and not to differences in the behaviour of the Canegro code itself.

Verification runs performed during this process are listed in Table 4.2.

Well-irrigated scenarios

The first stage in the water balance-enabled complete verification was assessing the differences between the models under a well-irrigated scenario. This was done to ensure that all linkages with DSSAT CSM variables had been made and that the model – as a computer program, rather than necessarily in the sense of a biophysical model – was functioning largely as it should.

Running the well-irrigated scenario (Run 2 in Table 4.2) revealed differences in the calculation of daily reference evaporation (potential evapotranspiration, ET_P, mm) between the models. This facilitated the calculation of an appropriate crop evapotranspiration adjustment coefficient, required for the FAO-56 option in the DSSAT CSM for calculating ET_P.

SASRI Canegro calculates a sugarcane ET_P using sugarcane parameters (McGlinchey and Inman-Bamber, 1996), whereas the DSSAT CSM calculates a short grass reference ET_P value which is multiplied by a crop coefficient to give a specific crop species' ET_P.

The crop coefficient chosen for the remainder of the verification exercises was the one that gave the best statistical match between the SASRI Canegro-simulated and DSSAT CSM-simulated ET_P values.

Water-stressed scenarios

After completing the iterative process of verifying, correcting and re-verifying DSSAT CSM Canegro under well-irrigated scenarios, the model was then verified under water-stressed scenarios. The intention of this aspect was to assess, explain and quantify differences between the models' performance. This turned out to be by far the most challenging aspect of this project, and as such the methodology for approaching this is described in detail in the next section.

4.7.6 Sensitivity analysis for error identification

A sensitivity analysis methodology was developed to identify and quantify the influences from outside the plant growth and development processes that caused discrepancies in output between the models when the water balances were enabled. The key feature of this process was to selectively force the values of one or more rate variables to be the same in both models, and to evaluate the effects this had on discrepancies in related output variables of interest.

4.7.7 Verification runs

Seven runs were performed, each forcing the values of different variables (or sets thereof). These runs and their specific intentions are described in Table 4.2. For each of these, a ten-year sequence was run, the details of which is described in the next section.

Table 4.2 List and descriptions of verification sensitivity runs.

Run	Intervention	Objective
1	Disabled water balance.	Establish similarity between DSSAT CSM Canegro and SASRI Canegro output, without the possible confounding

		effects of the water balances.
2	Well-irrigated.	Establish similarity between DSSAT CSM Canegro and SASRI Canegro when water stress was not present. Served as an indication that linkages between the plant and environment simulation modules operated correctly. Assessed the impact of two reference evaporation calculation methods.
3	None (water stressed scenario).	Establish similarity between SASRI Canegro and DSSAT CSM Canegro when water stress was apparent.
4	Reference evaporation set to 4 mm d^{-1} in both models for entire duration of simulations.	Demonstrate the contribution of different calculation methods of reference evaporation (sugarcane reference for SASRI Canegro and crop coefficient-adjusted FAO-56 short-grass reference for DSSAT CSM Canegro) between the two models.
5	Soil evaporation set to 0 mm d^{-1} in both models.	Assess the impact of different methods of calculating soil surface evaporation on output discrepancies.
6	Reference evaporation set to 4 mm d^{-1} and soil evaporation set to 0 mm d^{-1} in both models.	Demonstrate the impacts of the combined effects of differences in reference evaporation and soil surface evaporation calculation methods.
7	Reference evaporation set to 4 mm d^{-1}, soil evaporation set to 0 mm d^{-1}, and runoff set to 0 mm d^{-1} in both models.	Verify that the different methods of calculating runoff and drainage methods in the two models compensate for each other.

Verification experiment details

Verifications were performed using hypothetical experiments. For per-module verifications, weather data from the SASRI Mount Edgecombe weather station (29° 42' S, 31° 2' E, altitude 96 m) were used (automatic station data where available, supplemented with manual station data). Characteristics of the soil chosen for the simulations, based on the soil in the rainshelter facility at SASRI (described by Singels *et al.*, 2010b) are listed in Table 4.3. Two crops were defined: one that started on 1 April 2002, the other on 1 October 2002. These planting dates are six months apart and thus represent different weather patterns, considered a good test for the model.

For the sensitivity analysis verification runs, the plant/ratoon/harvest dates are listed in Table 4.4.

Cultivar coefficients used for these validation simulations are listed in Appendix 1.

Each simulation run was set up in SASRI Canegro using standard weather and soil file formats. The run was set up exactly the same in DSSAT, using the graphical experiment editor program, XBuild.

Table 4.3 Hypothetical verification experiment soil description (based on the soil described by Singels et al.*, 2010b).*

Layer boundary depth (cm)	**Lower limit (cm^3 cm^{-3})**	**Drained upper limit (cm^3 cm^{-3})**	**Saturated water content (cm^3 cm^{-3})**	**Soil water conductivity (mm h^{-1})**	**Bulk Density (g cm^{-3})**
5	0.110	0.255	0.33	0.40	1.44
15	0.127	0.255	0.33	0.40	1.52
30	0.145	0.255	0.33	0.40	1.60
45	0.124	0.255	0.33	0.40	1.60
60	0.094	0.255	0.33	0.40	1.64
78	0.099	0.255	0.33	0.40	1.65
95	0.100	0.255	0.33	0.11	1.65

Table 4.4 Crop types and start/harvest dates for the hypothetical ten-year sensitivity analysis verification simulation runs.

Crop type	**Plant/ratoon date**	**Harvest date**
Plant	01/04/1995	31/03/1996
Ratoon	01/04/1996	31/03/1997
Ratoon	01/04/1997	31/03/1998
Ratoon	01/04/1998	31/03/1999

Ratoon	01/04/1999	31/03/2000
Ratoon	01/04/2000	31/03/2001
Plant	01/04/2001	31/03/2002
Ratoon	01/04/2002	31/03/2003
Ratoon	01/04/2003	31/03/2004
Ratoon	01/04/2004	31/03/2005
Ratoon	01/04/2005	31/03/2006
Ratoon	01/04/2006	31/03/2007

4.8 Changes to the CSM Outside of the Canegro Plant Module

Three direct changes needed to be made to the DSSAT CSM to correctly accommodate the Canegro CSM module. These are described below.

Special permission was needed from the DSSAT team, as these changes required modifications outside of the Canegro module, and so potentially affected other crops in the DSSAT CSM. The DSSAT team tested the changes to ensure that they did not alter the simulation behaviour of other CSM crops.

- **Call to SC_CNGRO module**
  The first, and most necessary change, was adding a call to the SC_CNGRO subroutine from the DSSAT CSM Plant module.

- **Dew-point temperature calculation**
  Verification revealed certain inadequacies with the calculation of reference evaporation. The sensitivity analysis runs revealed that the DSSAT assumption of equality between minimum (T_{min}, °C) and dew-point temperatures (T_{dew}, °C), even when relative humidity data were provided (as these were ignored), was not acceptable for South African sugarcane growing conditions. This was replaced with an empirical equation described by Campbell and Norman (1998), which estimates T_{dew} from air temperature and maximum relative humidity. DSSAT makes provision for input of only a single daily relative humidity value, which was assumed to represent maximum daily relative humidity. If maximum relative humidity is not provided as an input, the equation reverts back to the assumption that $T_{dew} = T_{min}$. The DSSAT weather data import program (WeatherMan, Pickering *et al.* (1994)) needed to be modified to allow the correct input of (maximum) relative humidity. This was implemented by the DSSAT development team.

- **Total leaf area index**

  Sugarcane produces a substantial quantity of senesced leaf material ('trash'), which has a significant effect of suppressing soil surface evaporation. The SPAM module in DSSAT calculates transpiration from green leaf area index (*XHLAI*, $m^2\ m^{-2}$), by deriving fractional *PAR* interception (*LI*) using a PAR extinction coefficient. Daily potential soil surface evaporation (ES_P, mm) is determined by the complement ($ES_P = f(1-FI)$) of this. A modification was made such that total (living + dead) leaf area index (*XLAI*) is passed through the Canegro CSM module interface to the SPAM module. There it is used to calculate fractional interception of incident solar radiation by all leaf material, not just the green leaves. It was necessary to introduce modifications to the DSSAT CSM that set *XLAI* equal to *XHLAI* for all crops that do not calculate a value for *XLAI* (i.e. all crops other than sugarcane).

4.9 Model Validation

Model validation is the process of comparing model simulations with observations from real experiments. Graphical and statistical comparisons were made. The same statistical methods were used for validation as for code verification (already described in Section 4.7.1).

4.9.1 Validation experiments

Validation was performed using two datasets. These are referred to in this text as (1) the 'A/Growth/HR' experiment and (2) the 'A/Growth/07' experiment. Cultivar coefficients used for these validation simulations are listed in Appendix 1.

A/Growth/HR experiment

The A/Growth/HR experiment is described by Rostron (1972a), and represents the first ratoon of a well-irrigated crop grown under very favourable conditions (deep soil, high air temperatures, high solar radiation) in Pongola, in KwaZulu-Natal in South Africa (27° 24' S, 31° 35' E, altitude 308 m). The experiment ran from 1968 until 1971. The intention was to assess seasonal effects on potential growth of the reference sugarcane cultivar NCo376. Each treatment represented a different crop start date. Treatment details are listed in Table 4.5. The crop was planted in single rows with 1.5 m spacing. The crop was fully irrigated and adequately fertilised. The soil (USDA: Ultisol Haploxerults) is described in Table 4.6 (Rostron, 1972b).

Table 4.5 A/Growth/HR validation experiment treatment details.

Treatment	Crop class	Start	Harvest	Age at harvest (days)
1	Ratoon	17-Dec-68	05-May-70	504
2	Ratoon	11-Feb-69	30-Jun-70	504
3	Ratoon	08-Apr-69	25-Aug-70	504
4	Ratoon	03-Jun-69	20-Oct-70	504
5	Ratoon	29-Jul-69	15-Dec-70	504
6	Ratoon	23-Sep-69	09-Feb-71	504
7	Ratoon	18-Nov-69	06-Apr-71	504
8	Ratoon	13-Jan-70	29-May-71	501

Table 4.6 A/Growth/HR validation experiment soil characteristics (Rostron, 1972b).

Layer thickness (cm)	Lower limit (cm^3 cm^{-3})	Drained upper limit (cm^3 cm^{-3})	Saturated water content (cm^3 cm^{-3})	Relative root distribution	Bulk Density (g cm^{-3})
5	0.101	0.261	0.368	1.00	1.39
17	0.101	0.261	0.368	0.82	1.39
32	0.101	0.261	0.368	0.64	1.39
47	0.160	0.282	0.371	0.47	1.43
62	0.160	0.282	0.371	0.35	1.43
92	0.151	0.304	0.399	0.22	1.34
122	0.151	0.304	0.399	0.12	1.34
152	0.151	0.304	0.399	0.07	1.34
182	0.151	0.304	0.399	0.03	1.34
272	0.151	0.304	0.399	0.01	1.34

A/Growth/07 experiment

The A/Growth/07 experiment is described by Inman-Bamber (1994). The experiment was run from 1989 until 1991 at the SASRI experiment farm at La Mercy, KwaZulu-Natal, South Africa (29° 37’ S, 31° 5’ E, altitude 100 m). The crop was planted on a sandy clay Swartland soil (FAO: Brunic luvisol; USDA: Alfisol Rhodoxeralf). Physical soil characteristics read by the model are listed in Table 4.8; this experiment was included in the DSSAT v3.1 release of Canegro and the soil description provided was used. The objective of the trial was to assess seasonal effects on sugarcane growth in a rainfed, water-limited scenario. Treatment details are listed Table 4.7. Although both N12 and NCo376 varieties were planted, only the NCo376 data were considered for this validation.

Table 4.7 A/Growth/07 validation experiment treatment details.

Treatment	Crop class	Start	Harvest	Age at harvest (days)
1	Ratoon	01-Jun-89	02-Oct-90	488
2	Ratoon	01-Aug-89	05-Dec-90	491
3	Ratoon	01-Oct-89	05-Feb-91	492
4	Ratoon	01-Dec-89	03-Apr-91	488
5	Ratoon	01-Feb-90	04-Jun-91	488
6	Ratoon	01-Apr-90	31-Jul-91	486
7	Ratoon	01-Jun-90	01-Oct-91	487
8	Ratoon	01-Aug-90	03-Dec-91	489

Table 4.8 A/Growth/07 validation experiment soil characteristics (provided with DSSAT v3.1 software; described by Inman-Bamber, 1994).

Layer thickness (cm)	Lower limit (cm^3 cm^{-3})	Drained upper limit (cm^3 cm^{-3})	Saturated water content (cm^3 cm^{-3})	Relative root distribution	Bulk Density (g cm^{-3})	Saturated water conductivity (mm h^{-1})
5	0.102	0.255	0.387	1.00	1.30	0.80
10	0.102	0.255	0.329	0.80	1.61	0.80
15	0.102	0.237	0.316	0.70	1.63	0.80
15	0.131	0.228	0.319	0.60	1.59	0.80
15	0.132	0.238	0.345	0.50	1.48	0.70
15	0.142	0.258	0.359	0.45	1.46	0.60
15	0.221	0.329	0.390	0.40	1.48	0.50
15	0.307	0.349	0.385	0.37	1.56	0.50
15	0.346	0.375	0.391	0.35	1.60	0.50
45	0.357	0.405	0.413	0.32	1.56	0.05

5. RESULTS OVERVIEW

Results of this project are made up of three components:

- **Description of the DSSAT CSM Canegro Plant Module**
  The development of a modular DSSAT Cropping System Model (CSM) version of the SASRI Canegro model, termed the 'DSSAT CSM Canegro' model, was the primary objective of this project. The first part of the results (Sections 6.1-6.8) describes the software and its implementation.
- **Verification of the DSSAT CSM Canegro plant module source code**
  The second part of the results (Sections 7.1-7.3) presents the outcomes of verification exercises that establish that the Canegro plant and development code-sets behave identically in the SASRI Canegro and in the DSSAT CSM Canegro models. This verification exercise facilitated the calibration of an FAO-56 (Allen *et al.*, 1998) reference evaporation multiplier for sugarcane. The results also show that the small simulation discrepancies between the models' outputs in water-stressed scenarios is a consequence of differences in the calculations of water balance variables in the two models. The specific causes of these discrepancies are explained and described.
- **Validation of the CSM Canegro plant module**
  The performance of the DSSAT CSM Canegro plant module, in terms of its ability to simulate real sugarcane crops, was evaluated using data from 16 crops in two experiments. The statistical similarities between the simulations and the values measured during the experiments were calculated. Scatter plots and time series graphs of simulated and observed values are also shown and discussed. These results are presented in Sections 8.1-8.5.

6. THE DSSAT CSM CANEGRO PLANT MODULE

6.1 Introduction

A DSSAT Cropping System Model (CSM) plant module was created from the plant growth and development aspects of the SASRI Canegro sugarcane simulation model. The functioning of these plant growth and development algorithms were verified to operate identically within the plant module as in the SASRI Canegro model.

The DSSAT CSM Canegro Plant Module consists of several parts:

1. Source code
2. Genotype parameter files
3. Data code definitions
4. User and scientific documentation

Each of these is described separately in this section.

6.2 Source Code Features

The DSSAT CSM Canegro source code consists of several DSSAT CSM modules that represent plant processes, utility subroutines (such as for reading genetic parameter files and for searching files), and definitions of composite data types (discussed in Section 3.3.4). A full source code listing of the DSSAT Canegro CSM plant module is provided on the accompanying CD-ROM disc.

6.2.1 Composite data types

Six Fortran composite data types were defined that replace the common blocks that were used in SASRI Canegro. A list of these types and the concepts they represent, the name that each instance of these variables was given in the source code, and the meanings of key variables that they contain, is shown in Table 6.1. These composite types are defined in the file SC_CNG_MODS.FOR.

Table 6.1 Details of Fortran composite data types defined and used within the DSSAT CSM Canegro plant module, how they are named in the source code, which concepts they represent and the key variables contained within each type.

Composite type name	**Variable name in the source code**	**Concepts represented**	**Key variables contained**
CNG_SoilType	Soil	Characteristics of the soil profile	Bulk density, soil water content, layer depths
ClimateType		Weather	Daily rainfall, solar radiation, maximum and minimum air temperatures
WaterType	WaterBal	Water balance	Water stress variables and related coefficients. Root depth
CaneCropType	CaneCrop	Canopy	Leaf length and area, tiller cohort information, leaf and tiller development parameters
GrothType	Growth	Plant growth	Root mass increment, leaf area index, interception of photosynthetically-active radiation
PartType	Part	Biomass partitioning	Above- and below-ground biomass, partitioning-related genetic parameters, daily dry mass increments
OutType	Out	Output	Major growth- and development-related rate and state variables
RatoonCarry-OverType	RatCarryOver	Ratoon carryover information	Root depth

6.2.2 Plant module subroutines

The CSM Canegro plant module is called by the DSSAT Individual Plant Interface module (PLANT.FOR). This module calls the entry-point Canegro plant module subroutine SC_CNGRO in each mode (initialise, rates, integrate, etc). The source code excerpt is listed in Box 6.1.

Box 6.1 Source code of the call to the Canegro module in the DSSAT Individual Plant Interface module. The 'CONTROL' parameter is a composite variable containing information to control the simulation steps and ISWITCH contains information on simulation settings. The remaining variables are input variables passed into the Canegro module (e.g. SOILPROP – soil profile information, TMAX and TMIN – maximum and minimum air temperatures respectively) and output variables from Canegro (e.g. RLV – root length density, CANHT – canopy height, XHLAI – green leaf area index).

```
CASE('SCCAN')
    CALL SC_CNGRO (
 &    CONTROL, ISWITCH,                                   !Input
 &    CO2, DAYL, EOP, EP, EO, HARVFRAC, NH4, NO3, SNOW,   !Input
 &    SOILPROP, SRAD, SW, TMAX, TMIN, TRWUP, TRWU, EOS,   !Input
 &    RWUEP1, TWILEN, YREND, YRPLT, WEATHER,              !Input
 &    CANHT, HARVRES, KCAN, KTRANS, MDATE, NSTRES,        !Output
 &    PORMIN, RLV, RWUMX,SENESCE, STGDOY, UNH4,           !Output
 &    UNO3, XLAI, XHLAI, EORATIO)                         !Output
```

Not all of the interface variables are actually used. It was decided to leave these in place, to facilitate future model development (e.g. the development of a nitrogen uptake sub-model).

The DSSAT CSM calls the CSM Canegro module when the model name for a simulation set to 'SCCAN'. This is unique to sugarcane and also differentiates it from the CASUPRO ('SCCSP') sugarcane model which is also under development and uses the DSSAT CSM framework.

The structure of the DSSAT CSM Canegro plant module is illustrated in Figure 6.1. The sequence of calculations and the direction and flow of data are very similar to the SASRI Canegro model. Instead of this sequence only operating within the daily loop, however, these modules are all called in this order to initialise (once) and then calculate/integrate rates of change of state variables and (where necessary) produce output on a daily basis. All modules are called to finalise at the end of each simulation run.

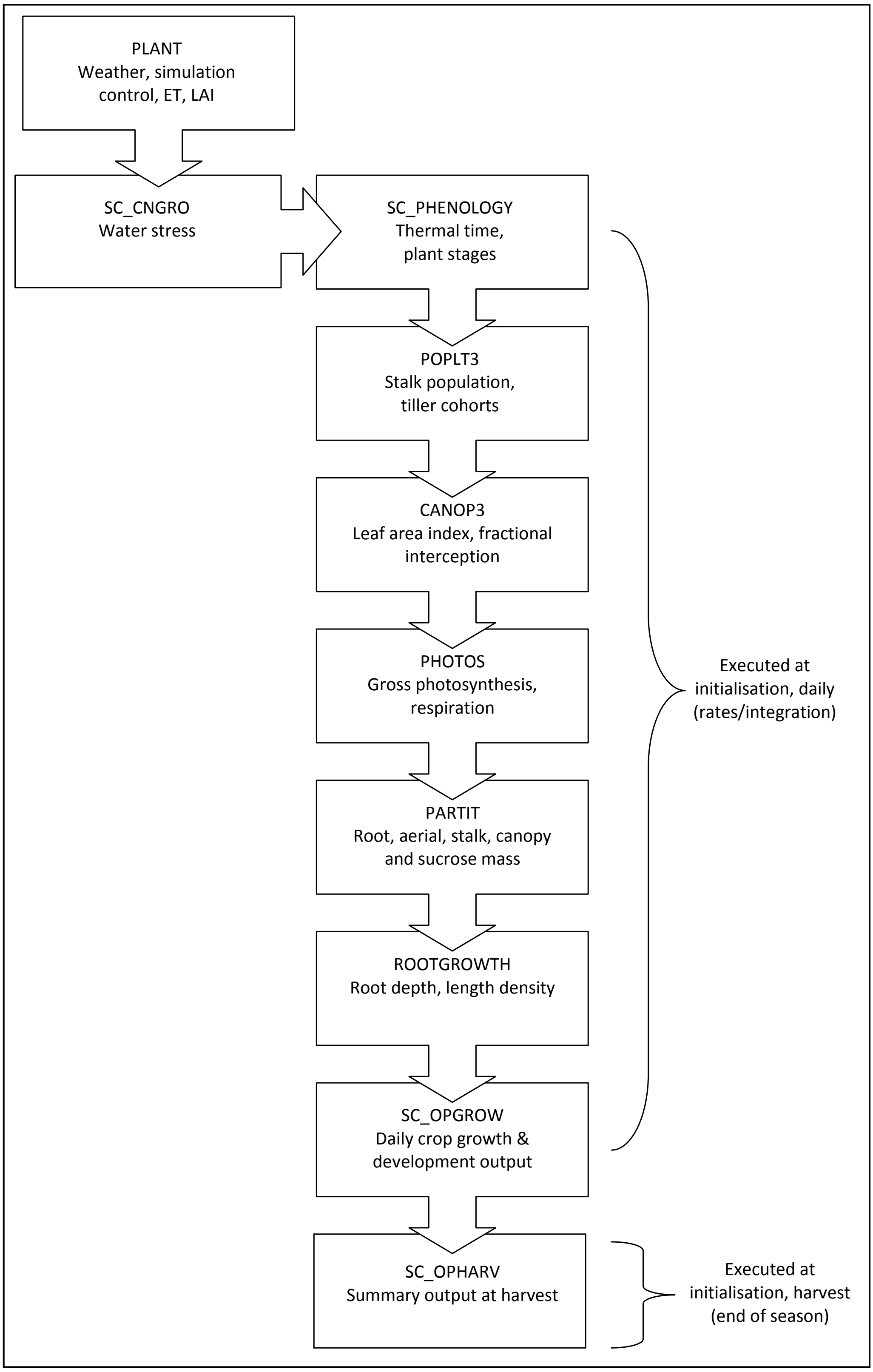


Figure 6.1 Structure of the DSSAT CSM Canegro plant module source code, showing sub-module names (represented by blocks) and key variables calculated by each of these. The CSM Canegro plant module is called by the DSSAT CSM individual plant interface module, which is itself called by the Land module, which represents a field on which the plant is grown.

Table 6.2 Name and purpose of modules in the DSSAT CSM Canegro plant module, along with their key input and output variables.

Module/subroutine name, *file name*	**Purpose**	**Key input variables**	**Key output variables**
SC_PHENOLOGY, *SC_PHENOL.FOR*	Calculate phenology: thermal time and phenological stage names	Max and min air temperature	Thermal time at different base temperatures
POPLT3, *SC_Poplt3.for*	Stalk population	Mean air temperature, soil water content	Stalk population, cohort number
CANOP3, *SC_Canop3.for*	Canopy development: leaf number, length, width, area, stalk length.	Stalk population, mean and max air temperature, water stress indices	Leaf area index, fractional interception of photosynthetically-active radiation, stalk length, canopy height
PHOTOS, *SC_PHOTOS.for*	Photosynthesis, growth and maintenance respiration.	Leaf area index, row-spacing, mean air temperature	Gross photosynthesis, growth and maintenance respiration, daily biomass increment
PARTIT, *SC_PARTIT.for*	Partitioning of daily biomass increment	Mean air temperature, total biomass increment, root partitioning fraction	Aerial biomass, stalk mass, sucrose mass, canopy mass, root mass increment
ROOTGROWTH, *SC_ROOTG.for*	Root growth and development	Mean air temperature, root biomass increment	Root depth, root length density per layer
SC_OPGROW, *SC_OUTPUT.for*	Daily output of growth and development-related variables	All state and some rate variables	Growth output file (usually named ‘Plantgro.out’)
SC_OPHARV, *SC_OPHARV.FOR*	Output of summary of growth and development-related variables, at harvest	A limited number of state variables	Contribution to ‘INFO.OUT’

The purpose and function of each sub-module of the DSSAT CSM Canegro plant module is listed in Table 6.2.

6.2.3 Non-module subroutines

DSSAT CSM modules are subroutines that fulfil a set of specific requirements (Section 3.3.4). In addition to the modules listed in Table 6.2, four utility subroutines were added to assist calculations in the model. It was not necessary, nor appropriate, for these subroutines to be implemented as CSM modules and so are listed separately, as follows:

GET_CULTIVAR_COEFF

This subroutine searches the cultivar and ecotype files for a parameter with the specified name, and returns the matching parameter value for the cultivar and ecotype currently in use. It is defined in the file SC_COEFFS.FOR.

GET_SPECIES_COEFF

This subroutine locates and searches the species file for a parameter with the specified name, and returns the matching parameter value. It is also defined in the file SC_COEFFS.FOR.

FINDNEXT

A subroutine that breaks text strings containing values separated by spaces into separate values. This subroutine is defined in the file SC_COEFFS.FOR.

FIND_IN_FILE

A subroutine that reads a file (unit number specified as a subroutine parameter), until it finds a line starting with the search keyword. The search keyword is specified as a subroutine parameter. This subroutine has since been converted, by and at the request of the DSSAT team, into a logical function and included as part of the general DSSAT CSM, defined in the CSM file READS.FOR.

6.3 Genotype Files

Genotype information is stored in three files, for species, ecotype and cultivar parameters respectively.

These files are stored in the Genotype directory of the DSSAT v.4 installation, typically `C:\DSSAT45\Genotype`. The files have the 'SCCAN045' prefix, a naming convention used in DSSAT which identifies the crop ('SC', sugarcane), the model ('CAN', Canegro) and the DSSAT CSM version to which the file belongs ('045', version 4.5). Each file is identified by its suffix – so SCCAN045.SPE is the species parameter file, SCCAN045.ECO is the ecotype file and SCCAN045.CUL is the cultivar file. Genotype parameters are listed in Table 6.3, Table 6.4 and Table 6.5.

Table 6.3 Names, values and descriptions of DSSAT CSM Canegro species parameters. The process that is affected by each parameter is also given.

Parameter name	Value	Process	Description and units
Tbasephotos	7.0	Photosynthesis	Base air temperature (°C) for photosynthesis.
Critsw	0.2	Photosynthesis	Water stress threshold for prolonged impact from severe water stress on photosynthesis.
HuRecover	150	Photosynthesis	Thermal time required for photosynthesis to recover fully after a severe water stress event (°C d).
RespQ10	1.68	Photosynthesis	Fractional increase in respiration rate per 10 °C rise in air temperature (Q10 coefficient).
RespGcf	0.242	Photosynthesis	Fraction of gross photosynthesis lost to growth respiration
PCB	0.6	Biomass partitioning	Partitioning coefficient: extinction coefficient of fraction of dry mass increments allocated to above ground biomass.
Max_rootpf	0.95	Biomass partitioning	Maximum partition fraction of daily mass increments to roots.
FTCON	0.32	Sucrose accumulation	Air temperature response shape parameter.
SURCON	0.99	Sucrose accumulation	Sucrose partitioning parameter that determines the response time of shifts in partitioning between sucrose and fibre in the stalk due to environmental changes (varies between 0 and 1).
RTcmpg	500	Root growth	Root length per mass of roots (cm g^{-1}).
RLVmin	0.02	Root growth	Minimum root length per soil volume (cm cm^{-3}).
SenesF	5	Canopy	Number of leaves per shoot senesced per 100 stress days.
Reset	5	Canopy	Rainfall required to reset stress day counter (mm).
Percoeff	0.16	Canopy	Fraction of plant elongation attributable to stalk elongation.
CHTCoeff	0.864	Canopy – height	Coefficient determining canopy height as a function of stalk height and number of leaves (cm cm^{-1}).
EORATIO	1.15	Water balance	Ratio of potential ET from fully canopied unstressed sugarcane canopy

			to grass reference ET (Kc from FAO-56).
RWUEP1	1	Water balance	Soil water supply/potential evaporation ratio threshold below which evaporation and photosynthesis are limited.
RWUEP2	2	Water balance	Soil water supply/potential evaporation ratio threshold below which expansive growth is limited.
RWUMX	0.07	Water uptake	Maximum root water uptake per unit length of root (cm^3 water/cm RLV).

Table 6.4 Names, values (for cultivar NCo376) and descriptions of DSSAT CSM Canegro ecotype parameters. The process that is affected by each parameter is also given.

Parameter name	**Value**	**Process**	**Description and units**
DELTTMAX	0.07	Sucrose accumulation	Max. change in sucrose content per unit change in stalk mass in the unripenened section of the stalk (t^{-1}).
dPERdT	0.176	Canopy – height	Change in plant extension rate (mm h^{-1}) per unit change in effective air temperature (°C).
EXTCFN	0.84	Canopy - light extinction	Maximum canopy photosynthetically-active radiation extinction coefficient.
EXTCFST	0.58	Canopy - light extinction	Minimum canopy photosynthetically-active radiation light extinction coefficient.
LFNMXEXT	20	Canopy - light extinction	Leaf number (including dead leaves still attached) at which maximum light extinction occurs.
AREAMX_CF(1)	0	Canopy – leaves	Cultivar parameter for quadratic equation defining maximum leaf area per leaf.
AREAMX_CF(2)	27.2	Canopy – leaves	Cultivar parameter for quadratic equation defining maximum leaf area per leaf.
AREAMX_CF(3)	-20.8	Canopy – leaves	Cultivar parameter for quadratic equation defining maximum leaf area per leaf
WIDCOR	1	Canopy – leaves	Parameter affecting the width of leaves.
WMAX_CF(1)	-0.0345	Canopy – leaves	Cultivar parameter for quadratic equation defining max. leaf width per leaf number.
WMAX_CF(2)	2.243	Canopy – leaves	Cultivar parameter for quadratic equation defining maximum leaf width per leaf number.
WMAX_CF(3)	7.75	Canopy – leaves	Cultivar parameter for quadratic equation defining maximum leaf width per leaf number.

LMAX_CF(1)	-0.376	Canopy – leaves	Parameter for quadratic equation defining maximum leaf length per leaf number.
LMAX_CF(2)	12.2	Canopy – leaves	Parameter for quadratic equation defining maximum leaf length per leaf number.
LMAX_CF(3)	21.8	Canopy – leaves	Parameter for quadratic equation defining maximum leaf length per leaf number.
MAXLFLENGTH	100	Canopy – leaves	Absolute maximum leaf length (cm) (overrides LMAX_CF-calculated values).
MAXLFWIDTH	3.5	Canopy – leaves	Absolute maximum leaf width (cm) (overrides LMAX_CF-calculated values).
POPCF(1)	1.826	Tiller population	Stalk population coefficient, in ideal conditions (no stress), as a function of thermal time.
POPCF(2)	-0.00201	Tiller population	Stalk population coefficient, in ideal conditions (no stress), as a function of thermal time.
POPDECAY	0.004	Tiller population	Fraction of shoots above the future mature shoot population (at a thermal time of 1600 °C d), that senesce per unit thermal time.
TTBASEEM	10	Phenology	Base air temperature (°C) for emergence and start of stalk elongation.
TTBASELFEX	10	Phenology	Base air temperature for leaf phenology (°C).
TTBASEPOP	16	Phenology	Base air temperature for shoot phenology (°C).
TBASEPER	10.57	Phenology	Base air temperature for plant extension (°C).
LG_AMRANGE	30	Lodging	Above-ground fresh mass range (above LG_AMBASE) over which lodging occurs (t ha^{-1}).
LG_GP_REDUC	0.28	Lodging	Maximum decrease in gross photosynthesis rate caused by lodging (0-1).
LDG_FI_REDUC	0.1	Lodging	Maximum decrease in fractional photosynthetically-active radiation interception caused by lodging (0-1).

Table 6.5 Names, values (for cultivar NCo376) and descriptions of DSSAT CSM Canegro cultivar parameters. The process that is affected by each parameter is also given.

Parameter name	**Value**	**Process**	**Description and units**
PARCEmax	9.9	Biomass accumulation	Maximum (no stress) radiation conversion efficiency expressed as assimilate produced before respiration, per unit photosynthetically-active radiation (g MJ^{-1}).
APFMX	0.88	Biomass accumulation	Maximum fraction of dry mass increments that can be allocated to aerial dry mass (t t^{-1}).
STKPFMAX	0.65	Biomass partitioning	Fraction of daily aerial dry mass increments partitioned to stalk at high air temperatures in a mature crop (t t^{-1} on a dry mass basis).
SUCA	0.58	Sucrose accumulation	Sucrose partitioning parameter: maximum sucrose content in the base of stalk (t t^{-1})
TBFT	25	Sucrose accumulation	Sucrose partitioning: air temperature at which partitioning of unstressed stalk mass increments to sucrose is 50% of the maximum value (°C)
LFMAX	12	Canopy – leaves	Maximum number of green leaves a healthy, adequately-watered plant will have after it is old enough to lose some leaves.
MXLFAREA	360	Canopy – leaves	Maximum leaf area assigned to all leaves above leaf number MXLFARNO (cm^2).
MXLFARNO	14	Canopy – leaves	Leaf number above which leaf area is limited to MXLFAREA.
PI1	69	Leaf phenology	Thermal time interval between consecutive leaf tip appearance (phyllocron interval), for leaf numbers below PSWITCH, °C d (base TTBASELFEX).
PI2	169	Leaf phenology	Phyllocron interval 2 (for leaf numbers above Pswitch, °C d (base TTBASELFEX)).
PSWITCH	18	Leaf phenology	Leaf number at which the phyllocron changes.
MAX_POP	30	Tiller phenology	Maximum tiller population (shoots m^{-2}).
POPTT16	13.3	Tiller phenology	Shoot population after 1600 °C d (shoots m^{-2}).
TTPLNTEM	428	Phenology	Thermal time to emergence for a plant crop (°C d, base TTBASEEM).
TTRATNEM	203	Phenology	Thermal time to emergence for a ratoon crop (°C d, base TTBASEEM).
CHUPIBASE	1050	Phenology	Thermal time (°C d , base TTBASEEM)

			from emergence to start of stalk elongation.
TT_POPGROWTH	600	Phenology	Thermal time from emergence to peak shoot population (°C d, base TTBASEPOP).
LG_AMBASE	220	Lodging	Above-ground fresh mass at which lodging starts (t ha^{-1}).

6.4 Heading Definitions for Output Module

The DSSAT file ‘DATA.CDE’, typically stored in C:\DSSAT45\, contains definitions for heading names used in model input (e.g. weather and soil data) and output (e.g. simulated plant growth data) data files.

Table 6.6 Data variable/heading names and definitions introduced for the DSSAT CSM Canegro model's output modules.

Variable name in source code	**Heading name in output file**	**Description and Units**
AERLDM	AELH	Aerial dry biomass (t ha^{-1}) at harvest
AERLDM	BADMD	Aerial dry biomass (t ha^{-1})
DWDT	BRDMD	Biomass increase per day (kg ha^{-1} day^{-1})
LAI	LAIGD	Green leaf area index (m^2 m^{-2})
TRASDM	LDDMD	Trash (residue) dry mass (t ha^{-1}), excluding green tops and meristem.
TOPDM	LGDMD	Green leaf canopy + meristem dry mass (t ha^{-1}) ['green tops']
GROSSP	PGRD	Gross photosynthesis rate (t ha^{-1} day^{-1})
ROOTDM	RDMD	Root dry mass (t ha^{-1})
RTLD	RTLD	Total root length (cm cm^{-2})
SHGT	SHTD	Stalk height (m)
STKDM	SMDMD	Stalk (millable) dry mass (t ha^{-1})
STKWM	SMFMD	Stalk (millable) fresh mass (t ha^{-1})
STKDM	STKH	Stalk (millable) dry mass (t ha^{-1}) at harvest
SUDMD	SU%DMD	Stalk sucrose content (%) (dry mass basis)
SUFMD	SU%FMD	Stalk sucrose content (%) (fresh mass basis)
SUCMAS	SUCH	Sucrose mass (t ha^{-1}) at harvest
SUCMAS	SUCMD	Sucrose mass (t ha^{-1})
TRASDM	TRSH	Trash (residue) dry mass (t ha^{-1}) at harvest
CHU_EM	TTEBC	Thermal time (°C d) for emergence base temperature
CHUPI	TTLEC	Thermal time (°C d) for shoot population base temperature
CHUPOP	TTSPC	Thermal time (°C d) for leaf extension base temperature

A number of new variables were added for the DSSAT CSM Canegro plant module. Sugarcane produces much higher biomass than most other crops, so the DSSAT convention of expressing mass values in kg ha^{-1} was impractical. A number of new codes for similar mass variables were defined in t ha^{-1}. The list of new variables introduced for the Canegro plant module is presented in Table 6.6. Output is provided by the SC_OPGROW (daily growth values) and SC_OPHARV (summarised growth values at harvest).

6.5 Canegro algorithm changes

The tiller senescence algorithm in SASRI Canegro was simplified, such that it required a only a single genotype parameter for calibration. Both the new DSSAT CSM Canegro approach (Eqn (4)) and the original SASRI Canegro approach (Eqn (5)) are described in Section 2.2.2.

6.6 Source code error corrections

Four major errors were identified and corrected (detailed in Sections 6.6.1-6.6.4). It is acknowledged that these have the potential to change the outcomes of simulations, and model re-calibration may be necessary in future. Additionally, two further inconsistencies were noted (described Sections 6.6.5 and 6.6.6), but these did not result in any changes to the source code of DSSAT CSM Canegro.

6.6.1 Rooting front depth

The daily change in the depth of the rooting front (*RTDEP*, m) was calculated as function of *DTT*, which represented (in SASRI Canegro) cumulative thermal time (°C), not the daily increment in thermal time. *RTDEP* was then updated with the daily change. This resulted in extremely rapid simulated rooting front depth increase. In the 10-year verification runs, rooting front depth typically reached the full depth of the soil profile (0.95 m) within one or two days of emergence, whereas the intention of the algorithm was a more gradual and realistic increase over a longer period. *DTT* was replaced with a daily, rather than cumulative, thermal time variable in the root growth and development module (ROOTGROWTH, SC_ROOTG.for) in DSSAT CSM Canegro.

6.6.2 Tillering water stress

The $SWDF_{30}$ variable (Section 2.2.7) is a stress index that affects tillering rate. $SWDF_{30}$ is intended to be a function of the soil water content of the top 30 cm of the soil profile. SASRI Canegro allowed arbitrary soil layer thicknesses, however, and did not make provision for soil layers that did not fit wholly within 30 cm. As a result, any soil layer that crossed the 30 cm threshold was ignored, yet the model still assumed that the full 30 cm depth was taken into account. If the model was run with the same soil profile, but with the layer that crossed the 30 cm boundary divided into two layers (with identical physical properties, and thus scientifically no different from the first case) separated on the 30 cm boundary, the model produced different simulation results. This arbitrary sensitivity to how inputs with identical meanings, but described differently, produced different model outputs, was clearly incorrect. The code was corrected by including the appropriate part of a soil layer that crossed the 30 cm threshold.

6.6.3 Root dry mass

In SASRI Canegro, root dry mass each day was assigned the product, rather than sum, of root mass and the daily increment in root mass. This error applied to the integration of root mass rather than the calculation of the daily root mass increment, on which the calculation of new root length was based. The only consequence of this error was incorrectly-reported root mass in the model output, and had no further implications on other variables.

6.6.4 Root length density

The SASRI Canegro source code revealed two approaches to accumulating root mass. For determining the values of state variables for above-ground dry mass and root mass, the *ROOTDWDT* (t ha^{-1} d^{-1}) variable was taken into account, based on a biomass partitioning fraction (*ROOTPF*, t t^{-1}), which was calculated as function of total dry mass (as described in Section 2.2.6). However, it was discovered that for the purposes of calculating root length density (root length per volume of soil, *RLV*, cm cm^{-3}), a different partitioning fraction *ROOTF* (t t^{-1}) was used to calculate a different daily root dry mass increment *GRORT* (t ha^{-1} d^{-1}). *ROOTF* had a constant value of 0.5 for the 10-year verification run, resulting in an average daily total root length of 32.3 cm cm^{-2}, compared with 10.52 as simulated by the corrected DSSAT CSM Canegro model. The code correction involved assigning the value of *ROOTDWDT* to *GRORT*.

6.6.5 Soil surface evaporation from top soil layer

SASRI Canegro calculated soil surface evaporation in a two stage process that had a special case for the topmost (shallowest) soil layer, the soil surface. It did not, however, give consideration to the thickness of that soil layer. The soil evaporation algorithm allowed this top layer to dry out more rapidly than deeper layers. Noticeably different simulation results could be produced by varying the thickness of the top layer, without otherwise altering the soil description in any way (i.e. while ensuring that the soil remained physically identical). No specific action was taken to address this in DSSAT CSM Canegro, as the DSSAT CSM calculates soil surface evaporation outside of the plant module, and the CSM (generally, unless specifically instructed) maps the input soil profile to fixed internal profile (limited by the input soil depth) in which the top layer is always 5 cm thick.

6.6.6 Radiation extinction coefficient for soil evaporation

The SASRI Canegro model calculated an exponential *PAR* interception coefficient, ranging from 0.58 to 0.84 (for cultivar NCo376) depending upon green leaf number, to account for changes in canopy architecture as the plant develops (Section 2.2.3). The calculation of potential soil surface evaporation, however, was based on a full-spectrum radiation interception coefficient of 0.4 (applied to total leaf area index, including senesced leaves). The DSSAT CSM makes provision for plant modules to supply a different radiation extinction coefficient for soil evaporation via the *KSEVAP* variable. This was not done, however, because *KSEVAP* is set by the CSM to default to 0.4. A further inconsistency is that it appears that potential soil surface evaporation in DSSAT CSM Canegro is calculated from EO *after* EO has been adjusted by the crop coefficient. Given that the crop coefficient of 1.15 for sugarcane (discussed in the Results chapters) increases EO by 15%, it follows that potential soil surface evaporation is estimated to be 15% higher on a soil growing sugarcane than a soil growing a short grass. This seems to be an incorrect assumption.

6.7 Documentation

The documentation for the model is separated into two parts. The first, the 'Scientific Documentation' (Singels *et al.*, 2008), describes the scientific concepts underlying the model and its algorithms. The second part, 'User documentation' (Jones *et al.*, 2008) represents a step-by-step user guide for people wishing to operate the model.

Documentation is available at the following web addresses:

- Scientific documentation – http://sasri.sasa.org.za/misc/DSSAT%20Canegro%20SCIENTIFIC%20documentation_20081215.pdf

- User documentation – http://sasri.sasa.org.za/misc/DSSAT_Canegro_USER_doc_disclaimercopyright.pdf

6.8 Discussion and Conclusion

The plant growth and development aspects of the SASRI Canegro model were re-structured from an integrated and procedural approach to the modular approach used in the DSSAT CSM. The SASRI Canegro model uses its own water balance and simulation management code, along with input files in formats specific to this model. The new DSSAT Canegro CSM uses the common CSM modules for these functions and processes, and uses DSSAT-compatible input and output file formats.

This effectively allows use of the Canegro model from within DSSAT. It means that the Canegro model in the CSM can make use of ongoing refinements to non-plant processes, and allows Canegro users and developers access to the myriad advantages offered by modularity, as discussed in Chapter 3 of this thesis.

The intention (and experience since completing the software implementation phase of this project) is to use the CSM implementation of Canegro as the ongoing development platform for the Canegro model at SASRI. While this has numerous advantages, it is not without risk. Canegro developers only have authority over the Canegro plant module in the CSM – in other words, just the plant growth and development aspects. Development of algorithms that require changes to common modules run the risk of being rejected by the DSSAT development team, leading either to the abandonment of those algorithms or the adoption of a non-standard version of the CSM at SASRI, which would lead to a divergence of model versions once again.

The implementation of the DSSAT CSM version of Canegro was the primary objective of this project. It was however necessary to verify that the DSSAT CSM Canegro simulates as closely as possible to the SASRI Canegro model, and the results of this verification exercise are presented in the next chapter. Error corrections made to the DSSAT CSM Canegro code were also made in the version of SASRI Canegro used for verification. It was also necessary to validate the DSSAT CSM Canegro model against experimental data, to measure its simulation performance (results shown in Chapter 8).

7. VERIFICATION RESULTS

7.1 Introduction

The purpose of verification is to establish how similarly the DSSAT CSM Canegro and the standalone SASRI Canegro models simulate sugarcane crops.

This verification process had three distinct aims, and phases:

1. **To show that the Canegro CSM plant module code works properly by running with water balances disabled**
   This first phase was performed to establish that the Canegro code in the DSSAT CSM operates identically as it does in the SASRI Canegro model. It entailed running both models with their water balances disabled. Any coding errors introduced during the incorporation process would have been revealed during this verification step.
2. **To assess the effects of enabling the water balances in both models**
   Given that there are differences in the non-plant aspects, i.e. the simulated environment in which the plant grows, some differences between the models could be expected when the water balance was enabled. This phase aimed to describe and quantify the simulation similarity between DSSAT CSM Canegro and SASRI Canegro for well-irrigated and water-stressed scenarios. These runs also facilitated the calibration of an FAO-56 crop reference evaporation coefficient.
3. **To identify and understand causes of differences between DSSAT CSM Canegro and SASRI Canegro simulation results**
   These runs aimed to identify and explain, using sensitivity analysis, the differences in simulation output between SASRI Canegro and DSSAT CSM Canegro arising from differences in the simulations of the water balances in the two models.

Ongoing verifications were performed during model development in order to ensure that each sub-process, as it was moved from Canegro to the CSM, operated identically in the CSM as in Canegro. Results from these verifications are not presented. Rather, verifications of the complete model are presented.

A validation – a comparison of DSSAT CSM Canegro results and actual observations from experiments – is presented in the next chapter.

7.2 Results of Verification Runs

Graphs and statistics for Runs 1 – 7 are presented in the following sections. Table 7.1 provides a key to the meanings and units of variables examined for verification. Statistical measures of similarity used are root mean squared error (RMSE) and average prediction error (APE). These statistics are expressed in the same units as those of the variables listed in Table 7.2. In the other tables in this chapter where RMSE and APE statistics are presented, these measures are expressed as fractions of the average value simulated by the SASRI Canegro model for each variable. These error terms are named 'RMSE_f' and 'APE_f'. Fractions were chosen to provide a unit-less indication of the severity of the error between the models' outputs across different variables. The average SASRI Canegro-simulated value was chosen as the reference value because the aim of the DSSAT CSM Canegro model was to simulate as closely as possible SASRI Canegro-simulated values, so expressing the error factors in these terms was most appropriate.

Table 7.1 List of output variable names and their units.

Description	**Variable**	**Units**
Shoot population	STKPOP	shoots ha^{-1}
Stalk height	STKHGT	cm
Green (healthy) leaf area index (LAI)	GLAI	$\text{m}^2\ \text{m}^{-2}$
Total (living and dead) leaf area index (LAI)	TLAI	$\text{m}^2\ \text{m}^{-2}$
Fractional interception of photosynthetically-active radiation	FI	
Gross photosynthesis	PGROSS	g m^{-2}
Root dry mass	ROOTDM	t ha^{-1}
Leaf and meristem (canopy) dry mass	TOPDM	t ha^{-1}
Millable stalk dry mass	STKDM	t ha^{-1}
Sucrose mass	SUCDM	t ha^{-1}
Trash dry mass	TRASHDM	t ha^{-1}
Soil water deficit factor 2 (water stress index affecting growth)	SWDF2	
Soil water deficit factor 1 (water stress index for photosynthesis)	SWDF1	
Reference evaporation	EO	mm d^{-1}
Potential plant evaporation (transpiration)	EOP	mm d^{-1}
Potential root water uptake	TRWUP	mm d^{-1}
Soil water deficit factor 30 (water stress index affecting tillering)	SWDF30	
Stalk fresh mass	STKWM	t ha^{-1}
Potential soil surface evaporation	EOS	mm d^{-1}
Actual soil surface evaporation	ES	mm d^{-1}
Soil water content of whole soil profile	TSW	cm
Total root length	TLROOT	cm cm^{-2}
Actual transpiration (root water uptake)	EP	mm d^{-1}

7.2.1 Run 1: Disabled water balance

This run verified the similarity between Canegro plant growth and development code sets in the DSSAT CSM and SASRI Canegro.

The water balance was disabled in both SASRI Canegro and DSSAT CSM Canegro. This ensured that both models effectively simulated the plant growth and development processes at climatic potential rates, without the confounding effects of differences in the water balance.

Table 7.2 presents results of verification Run 1.

Run 1 demonstrates that the Canegro plant growth and development algorithms in the DSSAT CSM operate very similarly to the SASRI Canegro model. All variables show negligible root mean squared errors (RMSEs) and average prediction errors (APEs).

Simulation of shoot population showed an RMSE of 200 shoots ha^{-1}, which is insignificant (given that final shoot population is typically in the order of 130 000 shoots ha^{-1}), particularly as the average prediction error (*APE*) is only -2.6 shoots ha^{-1}. This difference is ascribed to rounding differences in the input weather data files, as shoot population is determined entirely by air temperature in a non-water stressed scenario such as this. The difference in root dry mass is attributed to a negative minimum initial *ROOTDM* (root dry mass, t ha^{-1}) value in Canegro, which originates from a software bug.

Table 7.2 Root mean squared error (RMSE) and average prediction error (APE) statistics, minimum (MIN) and maximum (MAX) values for Run 1 (water balance disabled). Negative APE values indicate a bias towards SASRI Canegro simulations, while positive values indicate a bias towards DSSAT CSM Canegro simulations.

				SASRI Canegro		DSSAT CSM Canegro	
Variable	**Units**	**RMSE**	**APE**	**MIN**	**MAX**	**MIN**	**MAX**
STKPOP	shoots ha^{-1}	200.14	-2.60	0	300000	0.00	300000
STKHGT	cm	0.03	0.00	0.00	267.36	0.00	267.40
GLAI	m^2 m^{-2}	0.00	0.00	0.00	9.18	0.00	9.18
TLAI	m^2 m^{-2}	0.00	0.00	0.00	14.53	0.00	14.53
FI		0.00	0.00	0.00	0.99	0.00	0.99
PGROSS	g m^{-2}	0.00	0.00	0.00	1.19	0.00	1.18
ROOTDM	t ha^{-1}	0.02	0.02	-0.04	10.37	0.00	10.39
TOPDM	t ha^{-1}	0.02	-0.01	0.00	18.50	0.00	18.48
STKDM	t ha^{-1}	0.01	0.00	0.00	35.48	0.00	35.49
SUCDM	t ha^{-1}	0.00	0.00	0.00	15.61	0.00	15.62
TRASHDM	t ha^{-1}	0.01	0.00	0.00	9.87	0.00	9.87

All apparent discrepancies are insignificantly small and for all practical purposes the plant growth and development algorithms can be considered identical in the two models.

7.2.2 Run 2: Well-irrigated scenario

The water balance was enabled in both models and generous irrigation applied. The intention was to reduce water stress to negligible levels. This run is important for two reasons:

1. By enabling the water balances, this run demonstrated the similarity of the simulation outputs of the complete models. Any shortcomings in the DSSAT CSM Canegro's communication with CSM water balance would be apparent in the results of this run.
2. The run also allowed the calibration of a FAO-56 (Allen *et al.*, 1998) short grass reference evaporation crop coefficient for sugarcane. A value of 1.15 was determined.

Table 7.3 Root mean squared errors (RMSE$_f$), average prediction error (APE$_f$), R^2, slope and intercept of linear regression statistics for Run 2 (well-irrigated). Root mean squared error (RMSE) and average prediction error (APE) are presented as fractions of the average SASRI Canegro-simulated value for each variable.

Variable	Units	SASRI Canegro average	RMSE$_f$	APE$_f$	R^2	Slope	Intercept
GLAI	$m^2 m^{-2}$	3.46	0.03	0.00	1.00	1.00	0.01
ROOTDM	$t ha^{-1}$	4.61	0.00	0.00	1.00	1.00	0.00
STKHGT	cm	104.01	0.00	0.00	1.00	1.00	0.01
STKDM	$t ha^{-1}$	11.09	0.00	0.00	1.00	1.00	0.00
SUCDM	$t ha^{-1}$	4.24	0.01	0.00	1.00	1.00	0.00
SWDF1		1.00	0.00	0.00			
SWDF2		1.00	0.01	0.00	0.14	1.68	-0.68
TOPDM	$t ha^{-1}$	8.63	0.01	0.01	1.00	1.00	0.02
STKPOP	shoots ha^{-1}	198058	0.01	0.00	1.00	1.00	729.22
TRASHDM	$t ha^{-1}$	3.24	0.01	0.00	1.00	1.00	0.00
EO	$mm d^{-1}$	3.49	0.12	0.09	0.97	1.00	0.33
EOP	$mm d^{-1}$	2.71	0.12	0.08	0.99	1.05	0.09
TRWUP	$mm d^{-1}$	1.09	0.04	0.00	1.00	1.00	0.00
FI		0.71	0.01	0.00	1.00	1.00	0.00
SWDF30		0.98	0.02	0.00	0.81	0.92	0.08
PGROSS	$g m^{-2}$	0.44	0.01	0.00	1.00	1.00	0.00
STKWM	$t ha^{-1}$	42.27	0.01	0.00	1.00	1.00	0.02
EOS	$mm d^{-1}$	0.82	0.84	-0.10	0.98	1.09	0.01
ES	$mm d^{-1}$	0.56	0.58	0.04	0.82	0.99	0.03
TSW	cm	24.56	0.04	0.02	0.92	1.11	-2.25
TLROOT	$cm cm^{-2}$	16.02	0.01	0.01	1.00	1.00	0.13
EP	$mm d^{-1}$	2.71	0.12	0.08	0.99	1.05	0.09

Table 7.4 R^2, slope and intercept of linear regression statistics, and values predicted by DSSAT CSM Canegro expressed as percentages of SASRI Canegro-predicted values ('DSSAT% SASRI Canegro'), for variables at harvest (Run 2 – well irrigated).

Variable	**Units**	**R^2**	**Slope**	**Intercept**	**DSSAT% SASRI Canegro**
GLAI	$m^2\ m^{-2}$	0.99	0.98	0.09	100.14
ROOTDM	$t\ ha^{-1}$	1.00	0.99	0.09	100.11
STKHGT	cm	1.00	1.00	0.35	99.94
STKDM	$t\ ha^{-1}$	1.00	1.00	0.08	100.04
SUCMAS	$t\ ha^{-1}$	1.00	1.01	-0.17	100.10
TOPDM	$t\ ha^{-1}$	0.97	0.99	0.19	100.36
STKPOP	shoots ha^{-1}	1.00	1.00	0.00	100.00
TRASHDM	$t\ ha^{-1}$	0.99	1.00	0.00	100.29
FI		1.00	1.00	0.00	100.00
STKWM	$t\ ha^{-1}$	1.00	0.99	1.34	99.98

Tables 7.3 and 7.4, and Figures 7.1-7.3 present results of verification Run 2.

The results from the well-irrigated run indicate a very high level of agreement between the two models' output. *RMSE* and *APE* are low values, linear regression slopes are at or near one, and linear regression intercepts are generally at or close to zero (Table 7.4). The similarity in water-related variables – $SWDF_{30}$ (water stress index for the top 30 cm of soil), *EP* (actual transpiration, mm d^{-1}), *EOS* (potential soil surface evaporation, mm d^{-1}) and *ES* (actual soil surface evaporation, mm d^{-1}) – indicate that the linkages between the Canegro plant module in the DSSAT CSM operate as they should.

These values are not, however, identical. Differences in the values were expected because the water balance is simulated differently in DSSAT CSM Canegro and SASRI Canegro. These differences needed to be investigated and explained, however, which was done in verification Runs 4-7.

Time series graphs of leaf area index, stalk dry mass and sucrose mass simulated by both models provide a visual indication of the similarity of model performance (Figure 7.3).

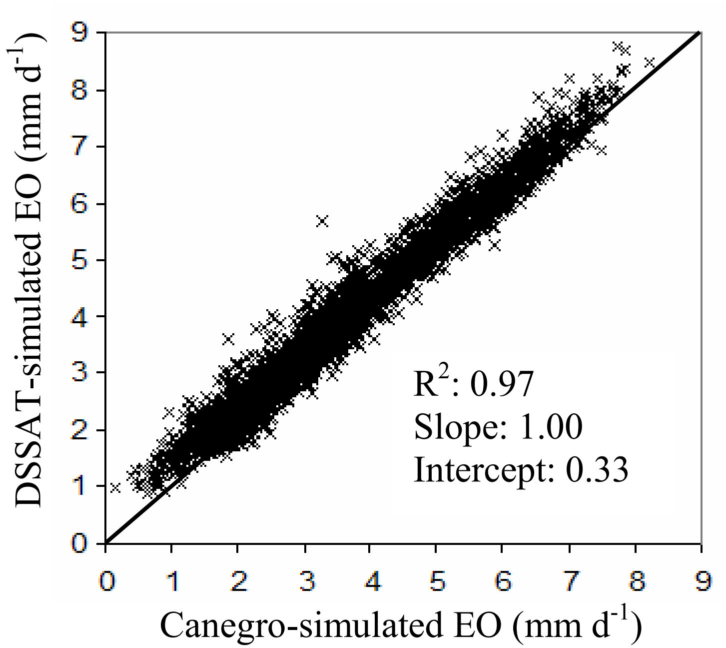


Figure 7.1 Reference evaporation (mm d^{-1}) simulated by the SASRI Canegro model (x-axis) and DSSAT CSM Canegro (y-axis), for Run 2 (well-irrigated). R^2, slope and intercept of linear regression values are shown. The solid line represents the 1:1 relationship.

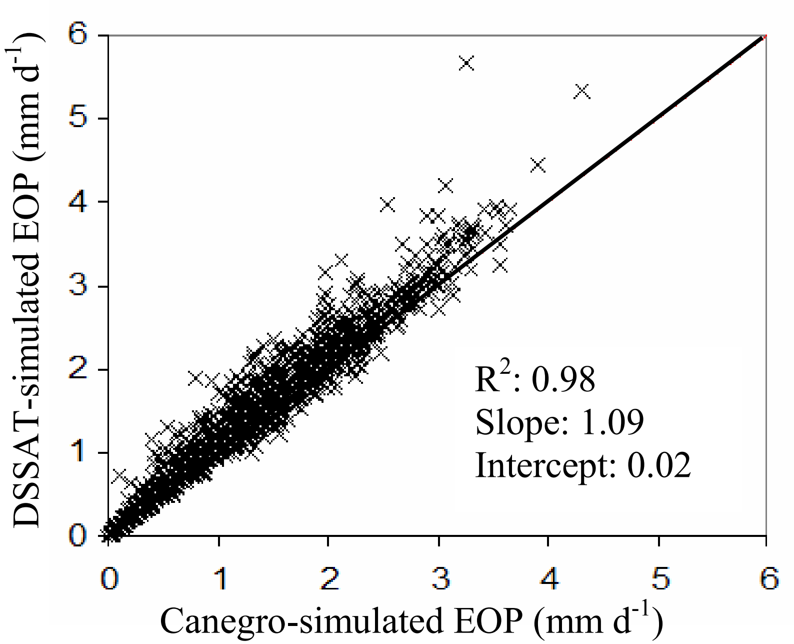


Figure 7.2 Potential transpiration (mm d^{-1}) simulated by the SASRI Canegro model (x-axis) and DSSAT CSM Canegro (y-axis), for Run 2 (well-irrigated). R^2, slope and intercept of linear regression values are shown. The solid line represents the 1:1 relationship.

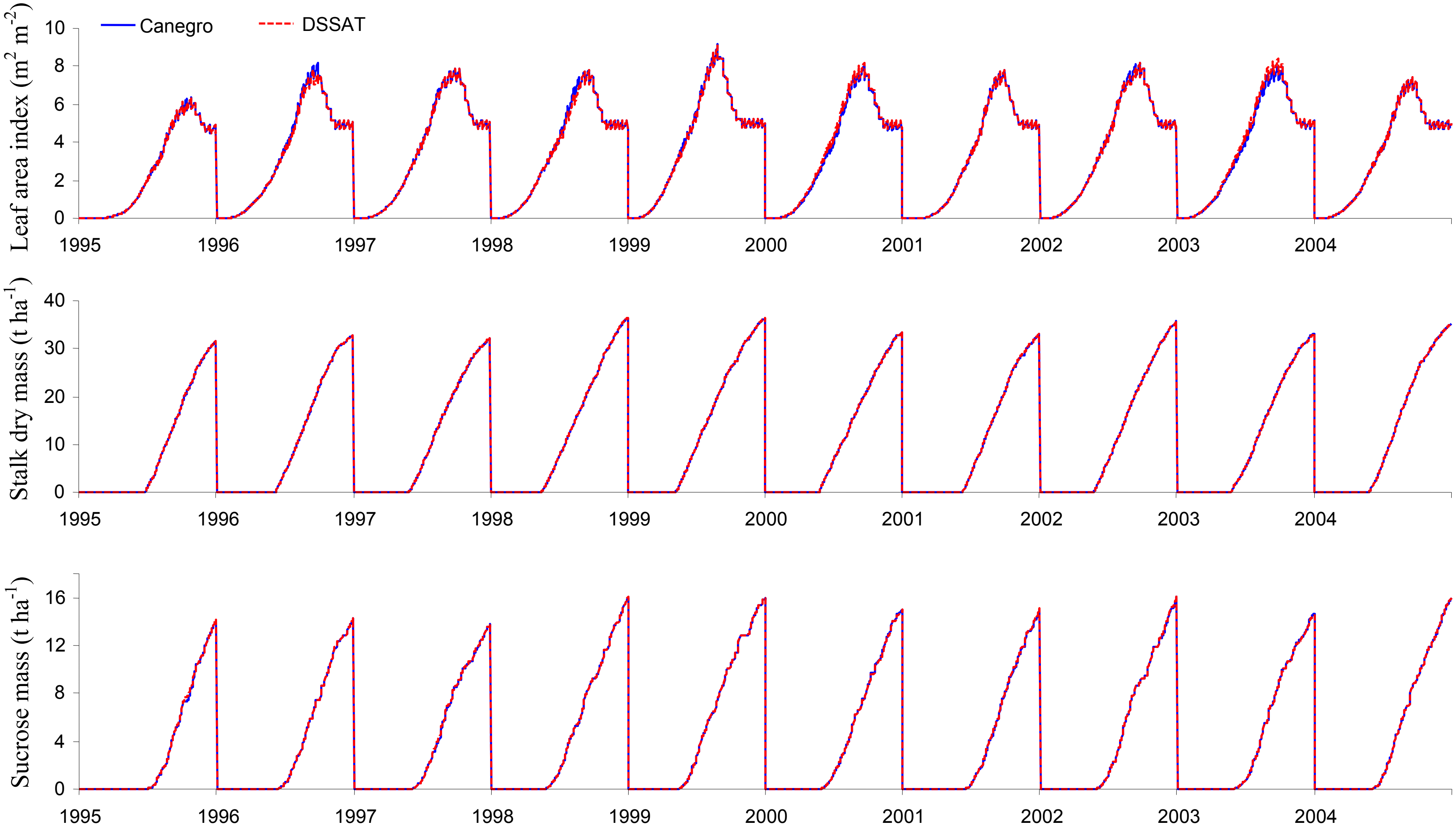


Figure 7.3 Line graphs showing leaf area index (m^2 m^{-2}), stalk dry mass (t ha^{-1}) and sucrose mass (t ha^{-1}) for Run 2 (well-irrigated scenario). Simulation results from the SASRI Canegro model are blue lines marked 'Canegro'. The red lines marked 'DSSAT' display DSSAT CSM Canegro simulation results. Simulations are for a 10-year period, a plant crop followed by nine ratoons starting in October 1995 at Mount Edgecombe.

A (maximum) FAO-56 evapotranspiration crop coefficient value of 1.25 is published for sugarcane by Allen *et al.* (1998). Inman-Bamber and McGlinchey (2003) and Olivier and Singels (2012) confirmed the value of 1.25. The results in this study suggest, however, a that value of 1.15 is appropriate. This run was repeated with several possible values and 1.15 yielded the most favourable correlation for simulated reference evaporation (*EO*), as indicated by a slope of one and low intercept (0.33 mm d^{-1}) (Figure 7.1) for the regression between Canegro- and DSSAT CSM-simulated values. The discrepancy between the outcome of this calibration experiment and the published values may be because this value was calibrated against values simulated using a more the more detailed Penman-Monteith – based sugarcane reference evaporation model in SASRI Canegro, rather than actual evaporation values measured in the field. Despite the discrepancy with published values, 1.15 was used for subsequent model runs. The R^2 value of 0.97 suggests there is some scatter in this relationship, which may translate into more noticeable impacts on plant growth and development. This scatter is likely to be explained by the use of daily maximum relative humidity in the FAO-56 calculation in the DSSAT CSM, rather than maximum and minimum values as used by the SASRI Canegro model. The differences in the calculation of reference evaporation were investigated in Run 4.

The DSSAT CSM Canegro model appears to simulate slightly higher potential transpiration (*EOP*, mm d^{-1}) values than SASRI Canegro (Figure 7.2): results show a slope of linear regression of 1.09 and an intercept of 0.02 mm d^{-1}. This could be expected to slightly increase simulated water stress levels in the DSSAT CSM Canegro model relative to SASRI Canegro.

7.2.3 Run 3: Water-stressed scenario

This run demonstrates the differences that are apparent between the models under water-limited (stressed) conditions. Section 4.7.7 describes this run in detail.

Table 7.5 Root mean squared error (RMSE$_f$), average prediction error (APE$_f$), R^2, slope and intercept of linear regression values for DSSAT CSM Canegro and SASRI Canegro results, for verification Run 3 (water-stressed scenario, Mount Edgecombe, 1995-2005 (daily values)). Root mean squared error (RMSE) and average prediction error (APE) are presented as fractions of the average SASRI Canegro-simulated value for each variable.

Variable	**Units**	**SASRI Canegro average**	**RMSE$_f$**	**APE$_f$**	**R^2**	**Slope**	**Intercept**
GLAI	m^2 m^{-2}	1.85	0.12	-0.05	0.98	0.94	0.03
ROOTDM	t ha^{-1}	3.27	0.04	-0.03	1.00	0.98	-0.01
STKHGT	cm	72.72	0.06	-0.04	1.00	0.95	0.62
STKDM	t ha^{-1}	6.48	0.07	-0.03	1.00	0.98	-0.09
SUCDM	t ha^{-1}	2.31	0.10	-0.04	1.00	0.97	-0.04
SWDF1		0.78	0.16	-0.03	0.88	0.97	0.00
SWDF2		0.69	0.17	-0.04	0.92	0.98	-0.01
TOPDM	t ha^{-1}	5.30	0.07	-0.04	1.00	0.97	-0.07
STKPOP	shoots ha^{-1}	185724	0.03	0.00	0.99	1.00	1135
TRASHDM	t ha^{-1}	1.88	0.11	-0.06	1.00	0.96	-0.05
EO	mm d^{-1}	3.46	0.11	0.06	0.96	0.93	0.45
EOP	mm d^{-1}	2.41	0.10	0.02	0.98	0.98	0.08
TRWUP	mm d^{-1}	0.41	0.23	-0.06	0.96	0.95	0.00
FI		0.62	0.04	-0.02	1.00	0.99	0.00
SWDF30		0.86	0.06	0.01	0.91	0.91	0.08
GROSSP	g m^{-2}	0.39	0.04	-0.01	1.00	0.99	0.00
STKWM	t ha^{-1}	25.04	0.06	-0.03	1.00	0.98	-0.34
EOS	mm d^{-1}	1.16	0.60	-0.02	0.47	0.69	0.34
ES	mm d^{-1}	0.61	0.62	0.05	0.71	0.95	0.06
TSW	cm	17.38	0.03	0.00	0.99	1.02	-0.38
TLROOT	cm cm^{-2}	11.46	0.03	-0.01	1.00	0.98	0.07

Table 7.6 R^2, slope and intercept of linear regression values indicating similarity between DSSAT CSM Canegro and SASRI Canegro results, and DSSAT CSM Canegro-simulated values expressed as percentages of SASRI-Canegro values, for verification Run 3 (water-stressed scenario), Mount Edgecombe, 1995-2005 (values at harvest).

Variable	**Units**	**R^2**	**Slope**	**Intercept**	**DSSAT/SASRI Canegro %**
GLAI	m^2 m^{-2}	0.98	1.07	-0.39	94.64
ROOTDM	t ha^{-1}	0.99	1.01	-0.20	97.97
STKHGT	Cm	0.99	0.94	2.88	95.80
STKDM	t ha^{-1}	0.99	1.01	-0.72	97.85
SUCDM	t ha^{-1}	0.99	1.05	-0.76	96.79
TOPDM	t ha^{-1}	0.99	0.99	-0.30	96.77
STKPOP	shoots ha^{-1}	1.00	1.00	0.00	100.00
TRASHDM	t ha^{-1}	0.99	1.00	-0.21	96.77
FI		0.98	1.23	-0.23	98.63
STKWM	t ha^{-1}	1.00	1.02	-3.71	97.73

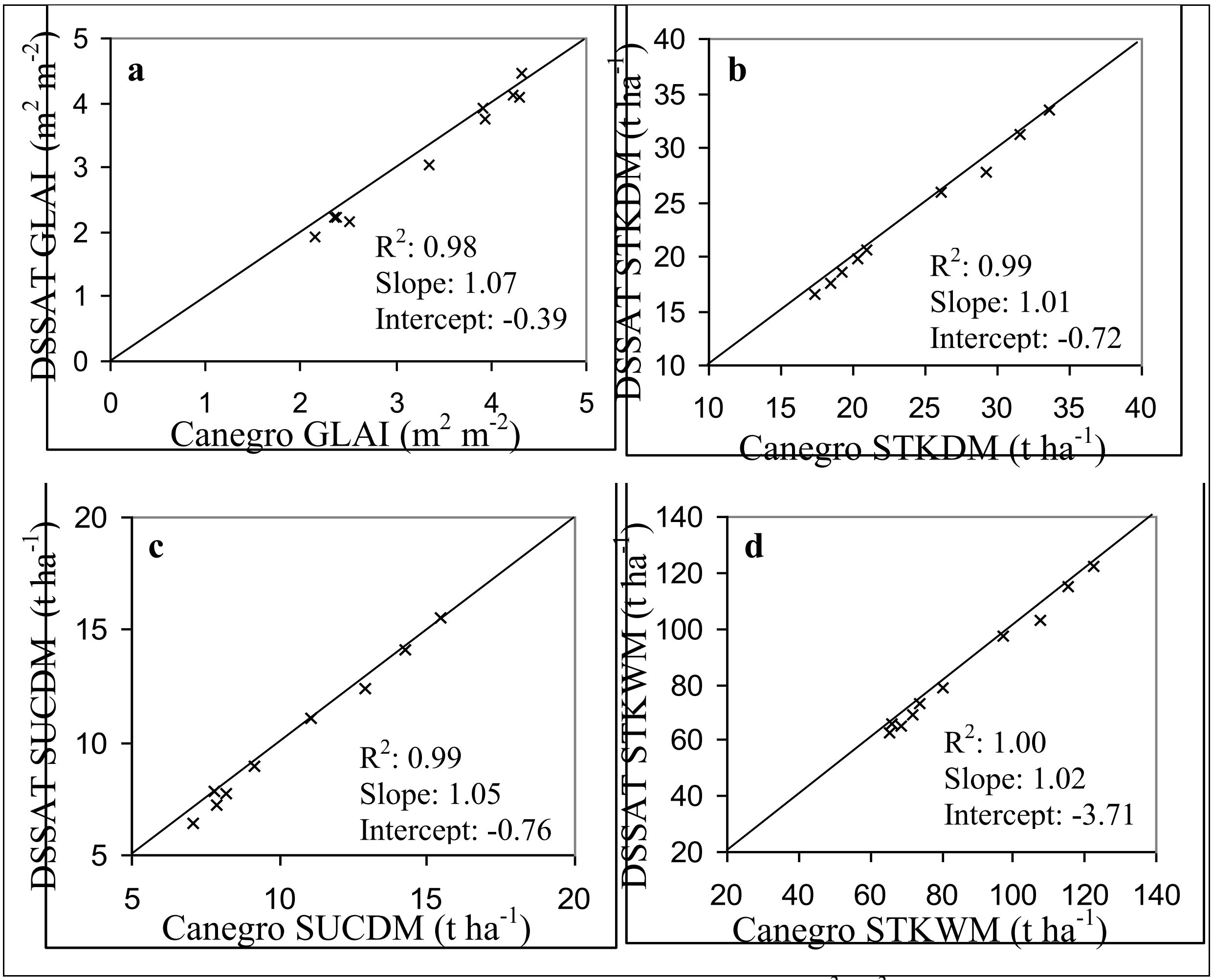


*Figure 7.4 Scatter plots of green leaf area index (GLAI, $m^2.m^{-2}$, **a**), stalk dry mass (STKDM, t ha^{-1}, **b**), sucrose mass (SUCDM, t ha^{-1}, **c**) and stalk fresh mass (STKWM, t ha^{-1}, **d**) at harvest for verification Run 3 (water-stressed scenario), Mount Edgecombe, 1995-2005. SASRI Canegro values plotted on the x-axes, DSSAT CSM Canegro values on the y-axes, and solid lines represent 1:1 relationships. R^2, slope and intercept of linear regression values are shown.*

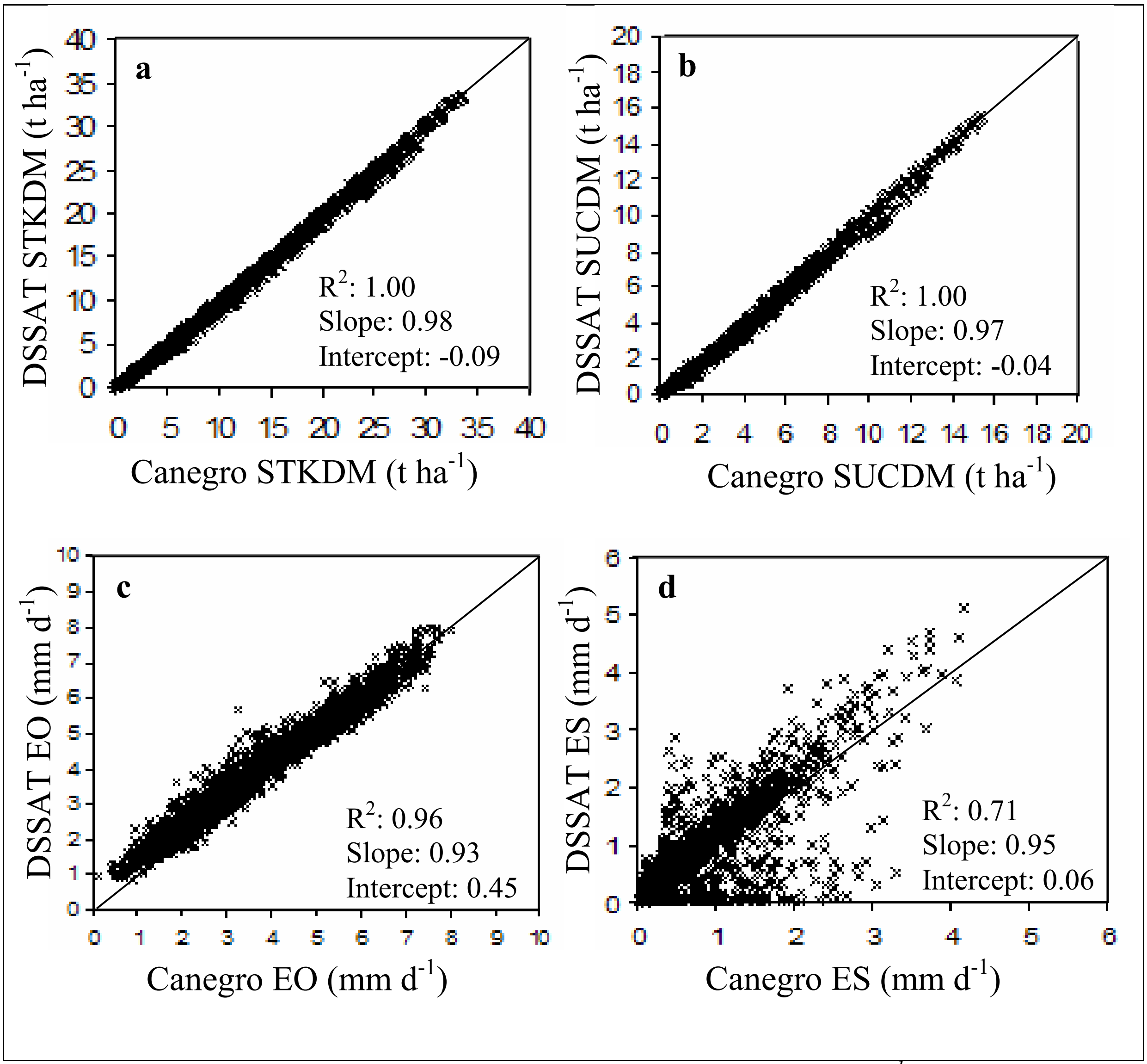


*Figure 7.5 Scatter plots of daily values of stalk dry mass (STKDM, t ha^{-1}, **a**), sucrose mass (SUCDM, t ha^{-1}, **b**), reference evaporation (EO, mm d^{-1}, **c**) and soil surface evaporation (ES, mm d^{-1}, **d**) for entire duration of verification Run 3 (water-stressed scenario), Mount Edgecombe, 1995-2005. SASRI Canegro values plotted on the x-axes, DSSAT CSM Canegro values on the y-axes, and solid lines represent 1:1 relationships. R^2, slope and intercept of linear regression values are shown.*

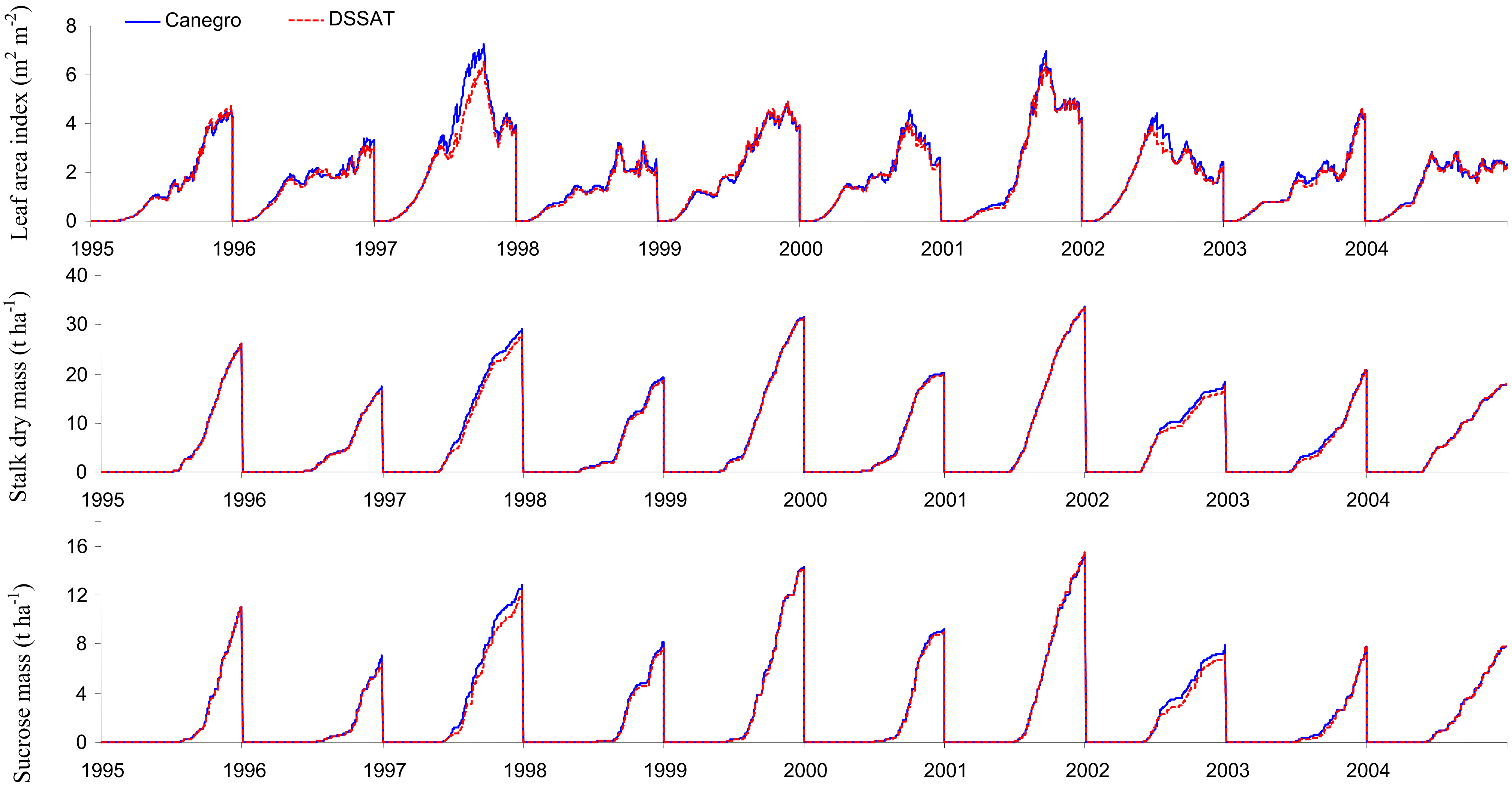


Figure 7.6 Line graphs showing leaf area index (m^2 m^{-2}), stalk dry mass (t ha^{-1}) and sucrose mass (t ha^{-1}) for Run 3 (water-stressed scenario). Simulation results from the SASRI Canegro model are blue lines marked 'Canegro'. The red lines marked 'DSSAT' display DSSAT CSM Canegro simulation results. Simulations are for a 10-year period, a plant crop followed by nine ratoons starting in October 1995 at Mount Edgecombe.

Run 3 results describe how closely DSSAT CSM Canegro simulates sugarcane under water-limited conditions compared to Canegro. It was important to verify the performance of the model under these conditions because most sugarcane in South Africa is grown under non-irrigated conditions where water stress is a frequent occurrence. The frequent and moderate-to-severe water stress conditions test the differences in the simulation of the water balance, as this drives the calculation of stress variables that reduce growth and development.

Overall, the DSSAT CSM Canegro module appears to simulate very much like the SASRI Canegro model. Statistical measures shown in Tables 7.5 and 7.6 indicate close correlations between daily values simulated by the two models. Figure 7.5 visually indicates close similarity. *RMSE* values are within 5% of the average simulated SASRI Canegro value, with the exception of *GLAI*. This appears to have had relatively little effect on yields (*STKDM*, *SUCDM*, *TOPDM*), as revealed in Table 7.6 and shown in Figure 7.6. Values at harvest indicate close agreement between DSSAT CSM Canegro and SASRI Canegro simulations (Figure 7.4). DSSAT-simulated values are generally slightly lower than those of Canegro (Tables 7.5 and 7.6).

Reference evaporation (*EO*), actual soil surface evaporation (*ES*) and particularly potential soil surface evaporation (*EOS*) compared relatively poorly (Table 7.5, Figure 7.5c and Figure 7.5d). Run 2 revealed that slight differences between EO simulated by the two models exists even under non water-limited conditions. The differences in the water balances in the two models resulted in differing simulated values for water stress indices $SWDF_1$ and $SWDF_2$ (Table 7.5), which affected canopy development, leaf area index (*LAI*) and fractional PAR interception (*FI*), which resulted in additional discrepancies in simulated *EO*. The magnitude of APE% for EOS is small (indicating that the models simulated roughly similar demand for soil surface evaporation), and its negative sign indicates that SASRI Canegro simulated a slightly greater demand for soil surface evaporation compared with DSSAT CSM Canegro. The large $\mathrm{RMSE_f}$ statistic and small APE% statistic for *EOS*, and its poor correlation statistics (slope is 0.69, and $R^2 = 0.47$), suggest the SASRI Canegro's simulation of EOS is routinely higher than that of DSSAT CSM Canegro, but that overall the values are similar. A possible explanation is that SASRI Canegro calculates *EOS* after integration of other variables, while DSSAT CSM Canegro does beforehand, meaning that *EOS* values are delayed by a day. Despite this, *ES* $\mathrm{APE_f}$ indicates that actual soil surface evaporation was higher with the DSSAT CSM Canegro simulation. These factors combined to increase water stress ($SWDF_1$ and $SWDF_2$) in the DSSAT CSM Canegro model relative to SASRI Canegro.

The better correlation statistics (i.e. slope closer to 1.0 and intercept closer to 0.0) shown by *ES* compared to *EOS* might be explained by *ES* being buffered with respect to short-term changes in *EOS*: after a soil-wetting event, the top soil layer (in both models) dries out rapidly (usually within 1-2 days) as result of soil surface evaporation (driven by *EOS*), with minimal soil surface evaporation thereafter. Whether this happens on the same day or the next day (i.e. depending on when *EOS* is integrated) is not terribly important, unless soil-wetting events routinely occur every 2-3 days.

In Run 4, the source of the soil surface evaporation discrepancies is investigated.

7.2.4 Run 4: Forcing reference evaporation to 4 mm d^{-1}

The DSSAT CSM and SASRI Canegro models calculate reference evaporation (*EO*, mm d^{-1}) differently. This run attempted to demonstrate, and describe, by comparison with Run 3, the contribution this difference made to inter-model simulation discrepancies.

Table 7.7, Figure 7.7, and Figure 7.8 present results of verification Run 4.

Table 7.7 Root mean squared error ($RMSE_f$), average prediction error (APE_f), R^2, slope and intercept of linear regression values for DSSAT CSM Canegro and SASRI Canegro results, for verification Run 4 (reference evaporation forced to 4 mm d^{-1}, Mount Edgecombe, 1995-2005 (daily values)). Root mean squared error (RMSE) and average prediction error (APE) are presented as fractions of the average SASRI Canegro-simulated value for each variable.

Variable	**Units**	**Canegro average**	**$RMSE_f$**	**APE_f**	**R^2**	**Slope**	**Intercept**
GLAI	m^2 m^{-2}	1.63	0.09	0.04	0.99	1.01	0.05
ROOTDM	t ha^{-1}	2.85	0.05	0.03	1.00	1.01	0.07
STKHGT	cm	61.63	0.04	0.03	1.00	1.01	1.23
STKDM	t ha^{-1}	6.02	0.05	0.02	1.00	1.01	0.10
SUCDM	t ha^{-1}	2.02	0.06	0.02	1.00	1.00	0.04
SWDF1		0.71	0.19	0.02	0.88	0.93	0.06
SWDF2		0.62	0.19	0.02	0.92	0.95	0.04
TOPDM	t ha^{-1}	3.81	0.10	0.07	0.99	1.03	0.16
STKPOP	shoots ha^{-1}	178284	0.05	0.03	0.99	1.03	-6.60
TRASHDM	t ha^{-1}	1.29	0.18	0.08	0.99	1.03	0.07
EO	mm d^{-1}	4.00					
EOP	mm d^{-1}	2.15	0.09	0.03	0.98	0.99	0.09
TRWUP	mm d^{-1}	0.37	0.18	0.04	0.98	1.00	0.02
FI		0.55	0.05	0.03	1.00	1.00	0.02
SWDF30		0.82	0.08	0.03	0.90	0.87	0.13
GROSSP	g m^{-2}	0.35	0.04	0.02	1.00	1.00	0.01
STKWM	t ha^{-1}	23.54	0.05	0.02	1.00	1.01	0.37
EOS	mm d^{-1}	2.08	0.46	-0.14	0.62	0.78	0.17

ES	mm d^{-1}	0.76	0.80	-0.07	0.59	0.73	0.15
TSW	cm	16.98	0.03	0.02	0.99	1.05	-0.51
TLROOT	cm cm^{-2}	10.48	0.07	0.04	0.99	0.99	0.49

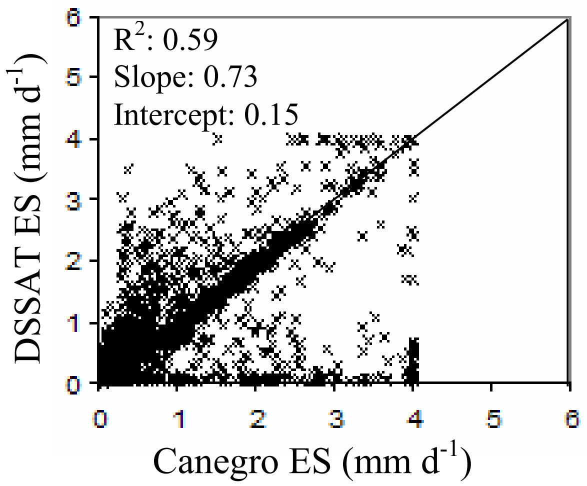


Figure 7.7 Actual (ES) soil surface evaporation (mm d^{-1}) for verification Run 4 (water-stressed, reference evaporation forced to 4 mm d^{-1}). Simulations are for a 10-year period, a plant crop followed by nine ratoons starting in October 1995 at Mount Edgecombe. SASRI Canegro values plotted on the x-axis, DSSAT CSM Canegro values on the y-axis, and the solid line represents the 1:1 relationship. R^2, slope and intercept of linear regression values are shown.

Forcing both models to use the same reference evaporation values improved simulations of *LAI*, *STKDM*, *SUCDM* and *TRWUP*, as indicated by reduced $RMSE_f$ and APE_f values, and regression slopes closer to one (Table 7.7). Simulated potential soil surface evaporation (*EOS*) values were more closely correlated between the models, with R^2 of linear regression increasing from 0.47 in Run 3 (Table 7.5) to 0.69 (Table 7.7), and the slope of linear regression also showing a substantial improvement (increasing to 0.78 from 0.69 in Run 3). Despite this, simulation discrepancies of actual soil surface evaporation (*ES*) worsened slightly, its R^2 decreasing from 0.71 to 0.59 and the slope decreasing from 0.95 to 0.73. The consequent differences in simulated soil water content of the top of the soil profile between the models affected the simulation of $SWDF_{30}$ (water stress affecting shoot population) and therefore *STKPOP* (shoot population), the statistics of which worsened slightly (Table 7.7). It appears then that discrepancies in the simulation of *EO* and *ES* compensated for each other to some extent.

The FAO-56 potential evapotranspiration method, as implemented in the DSSAT CSM, includes a leaf area index (*LAI*) term. The *LAI* term regulates the extent to which EO differs from short grass reference evaporation (i.e. up to 15% additional evaporative demand with the *EORATIO* value of 1.15 for sugarcane). *LAI* is sensitive to shoot population, which in turn is responsive to the soil water content of the top 30 cm of the soil profile via $SWDF_{30}$. *ES* has a significant impact on $SWDF_{30}$, so a possible mechanism by which *EO* and *ES* compensated for each other was that an overestimation of *ES* reduced *LAI*, which in turn reduced the crop-dependent additional evapotranspiration and hence calculated *EO*.

This compensatory mechanism explains why other variables are relatively unaffected. From the perspective of DSSAT CSM Canegro model operation, this is a favourable finding.

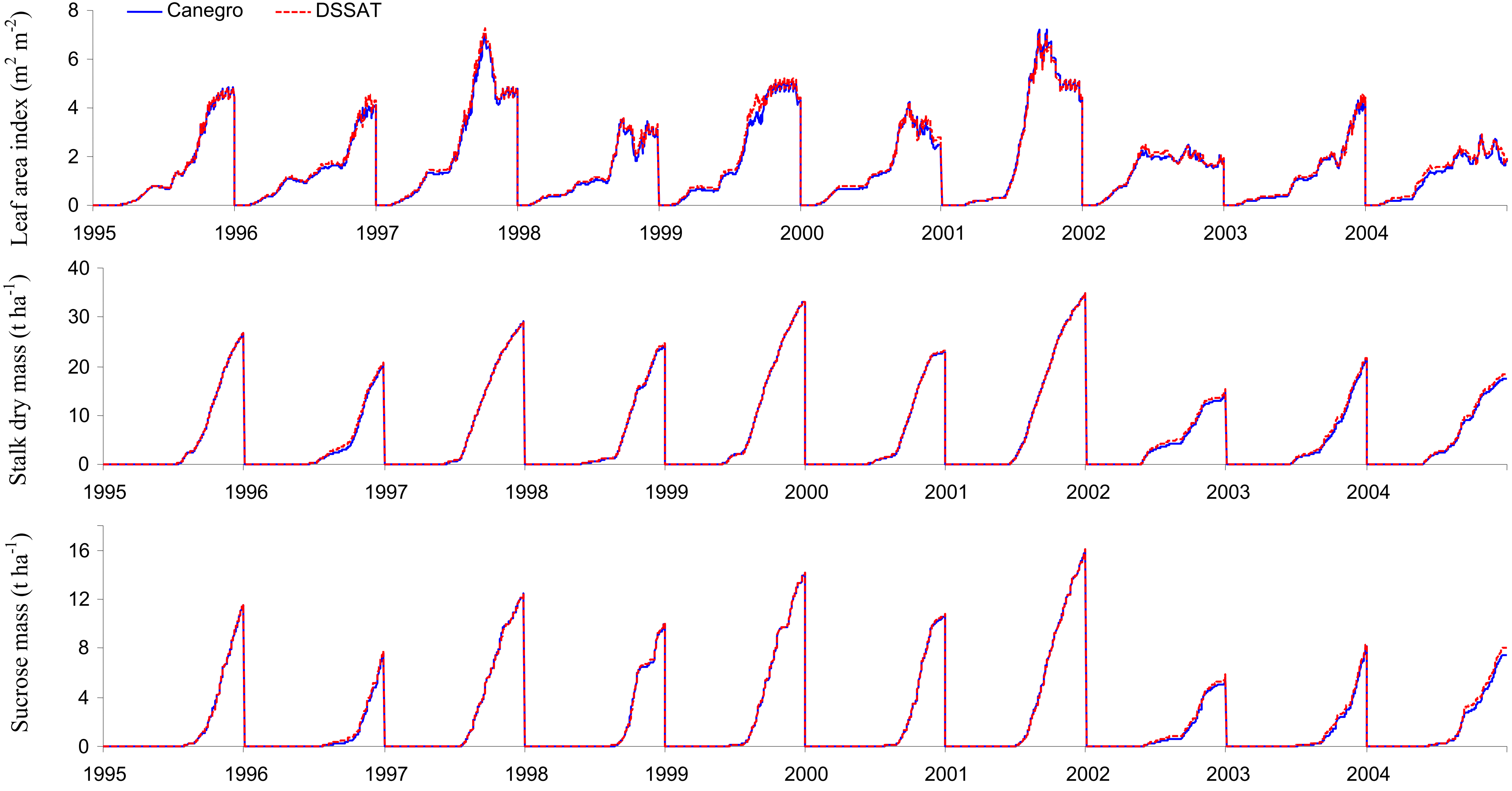


Figure 7.8 Line graphs showing leaf area index (m^2 m^{-2}), stalk dry mass (t ha^{-1}) and sucrose mass (t ha^{-1}) for Run 4 (water-stressed scenario, reference evaporation forced to 4 mm d^{-1}). Simulation results from the SASRI Canegro model are blue lines marked 'Canegro'. The red lines marked 'DSSAT' display DSSAT CSM Canegro simulation results. Simulations are for a 10-year period, a plant crop followed by nine ratoons starting in October 1995 at Mount Edgecombe.

Although differences in *EO* are responsible for some of the discrepancy between the models' outputs, it is clearly not the only cause. Results from Run 4 suggested that differences in the simulation of *ES* might have a significant contribution towards inter-model discrepancies. Run 5 assessed the impacts of forcing *ES* to be identical in the two models.

7.2.5 Run 5: Forcing soil surface evaporation to 0 mm d^{-1} in both models

Results from Run 4 indicated that soil surface evaporation (*ES*) was simulated noticeably differently between the two models. Run 5 assessed the contribution made towards overall inter-model discrepancies by just the differences in *ES* between the two models. This scenario is otherwise the same as Run 3 (the water-stressed scenario).

Table 7.8, Table 7.9 and Figure 7.9 present results of verification Run 5.

Table 7.8 Root mean squared error (RMSE$_f$), average prediction error (APE$_f$), R^2, slope and intercept of linear regression values for DSSAT CSM Canegro and SASRI Canegro results, for verification Run 5 (water-stressed scenario, soil surface evaporation set to 0 mm d^{-1}, Mount Edgecombe, 1995-2005 (daily values)). Root mean squared error (RMSE) and average prediction error (APE) are presented as fractions of the average SASRI Canegro-simulated value for each variable.

Variable	**Units**	**Canegro average**	**RMSE$_f$**	**APE$_f$**	**R^2**	**Slope**	**Intercept**
GLAI	m^2 m^{-2}	2.33	0.10	-0.04	0.99	0.95	0.02
ROOTDM	t ha^{-1}	3.90	0.03	-0.02	1.00	0.98	-0.01
STKHGT	cm	84.07	0.04	-0.03	1.00	0.96	0.58
STKDM	t ha^{-1}	7.80	0.06	-0.04	1.00	0.98	-0.13
SUCDM	t ha^{-1}	2.89	0.09	-0.05	1.00	0.97	-0.07
SWDF1		0.87	0.14	-0.02	0.84	0.95	0.03
SWDF2		0.78	0.14	-0.03	0.90	0.98	-0.01
TOPDM	t ha^{-1}	7.44	0.04	-0.03	1.00	0.97	0.00
STKPOP	shoots ha^{-1}	203147	0.00	0.00	1.00	1.00	-6.39
TRASHDM	t ha^{-1}	2.79	0.06	-0.04	1.00	0.98	-0.03
EO	mm d^{-1}	3.47	0.11	0.07	0.97	0.94	0.43
EOP	mm d^{-1}	2.78	0.09	0.03	0.99	0.99	0.12
TRWUP	mm d^{-1}	0.59	0.18	-0.05	0.96	0.96	0.00
FI		0.68	0.52	-0.31	1.00	0.99	0.00
SWDF30		0.95	0.02	-0.01	0.94	0.93	0.07
PGROSS	g m^{-2}	0.42	0.08	0.02	1.00	0.99	0.00
STKWM	t ha^{-1}	29.91	0.00	0.00	1.00	0.98	-0.49
EOS	mm d^{-1}	0.76	0.90	-0.06	0.97	1.08	0.05
ES	mm d^{-1}	0.00					
TSW	cm	19.15	0.02	-0.01	0.99	1.01	-0.16
TLROOT	cm cm^{-2}	13.40	0.00	0.00	1.00	0.98	0.12

Table 7.9 R^2, slope and intercept of linear regression values indicating similarity between DSSAT CSM Canegro and SASRI Canegro results, and relative performance of DSSAT CSM-simulated results expressed as percentages of SASRI-Canegro results, for verification Run 5 (water-stressed scenario, soil surface evaporation set to 0 mm d^{-1}), Mount Edgecombe, 1995-2005 (values at harvest).

Variable	Units	R^2	Slope	Intercept	DSSAT/CANEGRO %
GLAI	$m^2.m^{-2}$	0.99	0.97	0.08	99.39
ROOTDM	t ha^{-1}	1.00	1.04	-0.41	98.69
STKHGT	cm	1.00	1.00	-4.71	96.86
STKDM	t ha^{-1}	0.99	1.02	-0.84	98.51
SUCDMS	t ha^{-1}	0.99	1.06	-0.87	98.06
TOPDM	t ha^{-1}	0.99	0.99	-0.18	98.08
STKPOP	shoots ha^{-1}	0.00	0.00	0.00	100.00
TRASHDM	t ha^{-1}	0.99	1.00	-0.14	98.01
FI		1.00	1.08	-0.08	99.75
STKWM	t ha^{-1}	0.99	1.02	-3.07	98.69

With soil surface evaporation forced to be equal in both models, correlations between corresponding output variables generally improved compared with results from Run 3 (water-stressed scenario, Table 7.5) and Run 4 (reference evaporation forced to 4 mm d^{-1}, Table 7.7). This suggested that the differences in the simulation of *ES* were a strong determinant of the inter-model discrepancies.

Shoot population (*STKPOP*) in particular was nearly identical in both models, and slopes of linear regression values were close to one for most variables. *LAI* (leaf area index) also showed some improvement (Figure 7.9). Water stress indices $SWDF_1$ and $SWDF_2$ remained sufficiently different to have affected gross photosynthesis rates (*PGROSS*) and biomass partitioning, such that dry mass values for stalks (Figure 7.9), green tops and sucrose (Figure 7.9) were still simulated lower by DSSAT CSM Canegro than SASRI Canegro.

Runs 4 and 5 revealed that differences in *EO* and *ES* appeared to account for the most of the differences between the model outputs for this water-stressed simulation run. This would be confirmed if *ES* and *EO* were controlled (ES = 0 mm d^{-1}, EO = 4 mm d^{-1}) and model output correlations improved significantly – this was investigated in Run 6. Run 6 was similar to Run 3 (water-stressed scenario), different only in that *EO* and *ES* were simultaneously forced to predetermined constant values of 4 mm d^{-1} and 0 mm d^{-1} respectively.

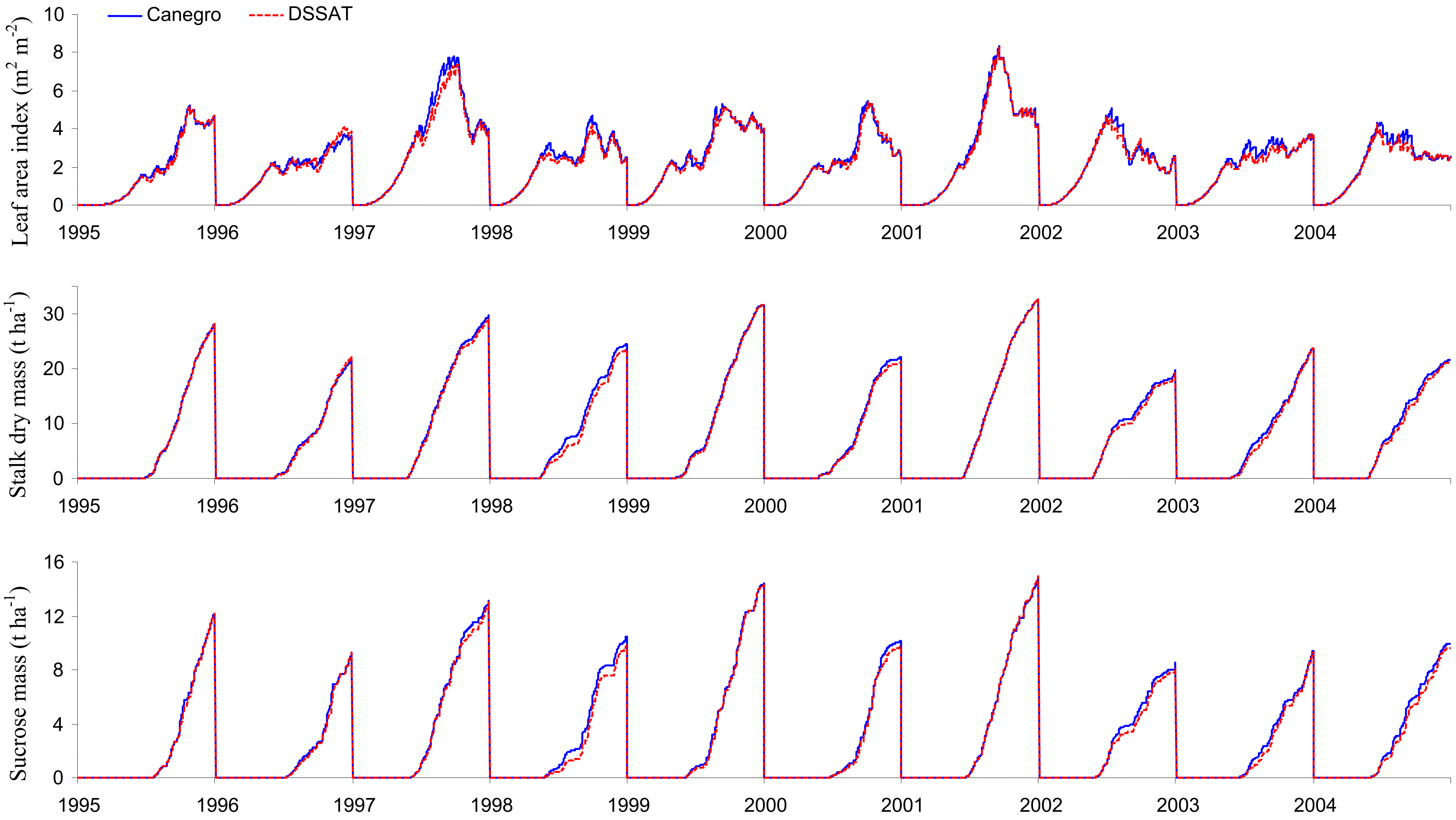


Figure 7.9 Line graphs showing leaf area index (m^2 m^{-2}), stalk dry mass (t ha^{-1}) and sucrose mass (t ha^{-1}) for Run 5 (water-stressed scenario, soil surface evaporation forced to 0 mm d^{-1}). Simulation results from the SASRI Canegro model are blue lines marked 'Canegro'. The red lines marked 'DSSAT' display DSSAT CSM Canegro simulation results. Simulations are for a 10-year period, a plant crop followed by nine ratoons starting in October 1995 at Mount Edgecombe.

$SWDF_{30}$ is an index of soil water content in the top 30 cm of the soil profile. Soil surface evaporation is a major determinant of the soil water content of the top of the soil profile, and hence $SWDF_{30}$. Following forcing ES to 0 mm d^{-1}, $SWDF_{30}$ statistics still indicated inter-model differences. This suggested that differences in water infiltration, arising from differences in the simulation of runoff (stormflow), might explain further inter-model differences. Furthermore, differences in the simulations of cumulative runoff and drainage were noted in the results from Runs 3 and 5 (results not shown). Differences in runoff and drainage tended to be inversely-related – i.e. where SASRI Canegro-simulated runoff was higher, DSSAT CSM Canegro-simulated drainage was slightly lower. These observations were investigated in Run 7.

7.2.6 Run 6: Forcing reference evaporation to 4 mm d^{-1} and soil evaporation to 0 mm d^{-1} in both models

By comparison with Run 3 (water-stressed scenario), the results of Run 6 demonstrate the contribution made towards overall simulation differences between the models by the combined effects of differences in reference evaporation and soil surface evaporation calculations.

Table 7.10 and Figures 7.10-7.13 present results of verification Run 6.

Table 7.10 Root mean squared error ($RMSE_f$), average prediction error (APE_f), R^2, slope and intercept of linear regression values for DSSAT CSM Canegro and SASRI Canegro results, for verification Run 6 (water-stressed scenario, soil surface evaporation set to 0 mm d^{-1} and EO set to 4 mm d^{-1}, Mount Edgecombe, 1995-2005 (daily values)). Root mean squared error (RMSE) and average prediction error (APE) are presented as fractions of the average SASRI Canegro-simulated value for each variable.

Variable	**Units**	**Canegro average**	**$RMSE_f$**	**APE_f**	**R^2**	**Slope**	**Intercept**
GLAI	m^2 m^{-2}	2.18	0.04	0.01	1.00	1.00	0.01
ROOTDM	t ha^{-1}	3.58	0.02	0.01	1.00	1.01	0.00
STKHGT	cm	75.90	0.01	0.01	1.00	1.01	0.08
STKDM	t ha^{-1}	7.62	0.03	0.01	1.00	1.00	0.02
SUCDM	t ha^{-1}	2.70	0.04	0.00	1.00	1.00	0.01
SWDF1		0.83	0.00	0.00	0.99	0.99	0.01
SWDF2		0.73	0.00	-0.01	0.99	0.99	0.01
TOPDM	t ha^{-1}	5.94	0.03	0.01	1.00	1.01	0.02
STKPOP	shoots ha^{-1}	202847	0.00	0.00	1.00	1.00	-89.15
TRASHDM	t ha^{-1}	2.14	0.04	0.01	1.00	1.01	0.02
EO	mm d^{-1}	4.00					
EOP	mm d^{-1}	2.72	0.02	0.00	1.00	0.99	0.01

TRWUP	mm d^{-1}	0.54	0.00	-0.01	0.99	1.01	0.00
FI		0.65	0.02	0.00	1.00	1.00	0.00
SWDF30		0.94	0.04	0.01	0.95	0.93	0.07
PGROSS	g m^{-2}	0.40	0.02	0.00	1.00	1.00	0.00
STKWM	t ha^{-1}	29.50	0.03	0.01	1.00	1.00	0.07
EOS	mm d^{-1}	1.68	0.57	-0.14	1.00	1.00	-0.01
ES	mm d^{-1}	0.00					
TSW	cm	18.65	0.03	0.02	1.00	1.04	-0.25
TLROOT	cm cm^{-2}	12.76	0.02	0.01	1.00	1.01	0.07

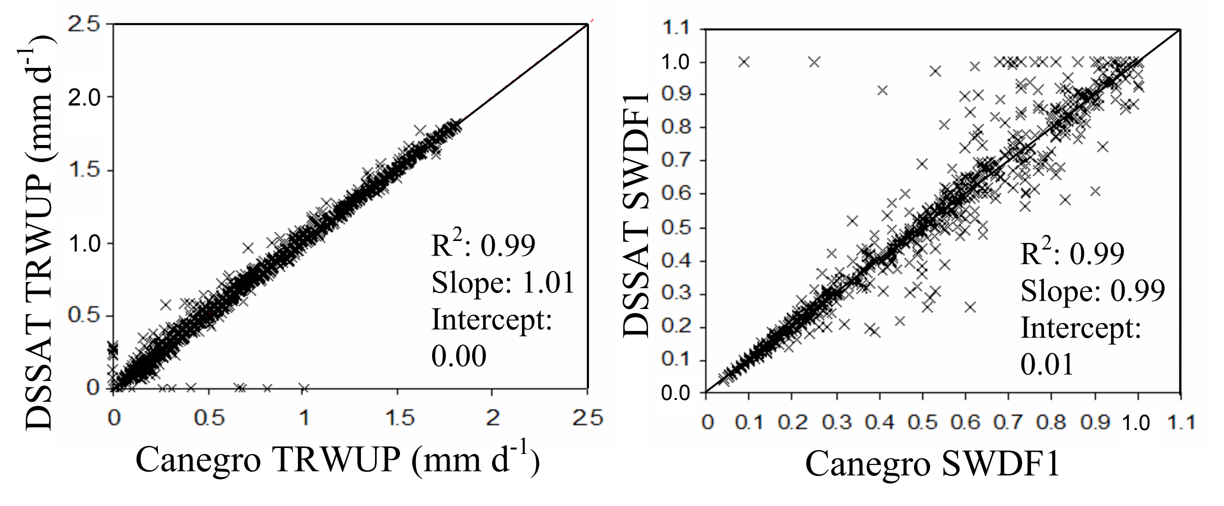


Figure 7.10 Total potential root water uptake (TRWUP, mm d^{-1}) and soil water deficit factor 1 (SWDF$_1$) for verification Run 6 (water-stressed scenario, soil surface evaporation forced to 0 mm d^{-1} and reference evaporation forced to 4 mm d^{-1}), Mount Edgecombe 1995-2005. Values are offset by one day. SASRI Canegro values plotted on the x-axes, DSSAT CSM Canegro values on the y-axes, and solid lines represent 1:1 relationships. R^2, slope and intercept of linear regression values are shown for SWDF$_1$.

Correlations between corresponding output variables generally improved, once again relative to Run 3 (water-stressed scenario), Run 4 (*EO* = 4 mm d^{-1} in both models) and Run 5 (*ES* = 0 mm d^{-1} in both models). It is concluded then that differences in output between the models can be attributed primarily to variations in the calculation of actual soil evaporation (*ES*) and reference evaporation (*EO*). The differences appear to compensate for each other slightly as well, which is supported by the fact that *EO* is an input to the calculation of *ES*.

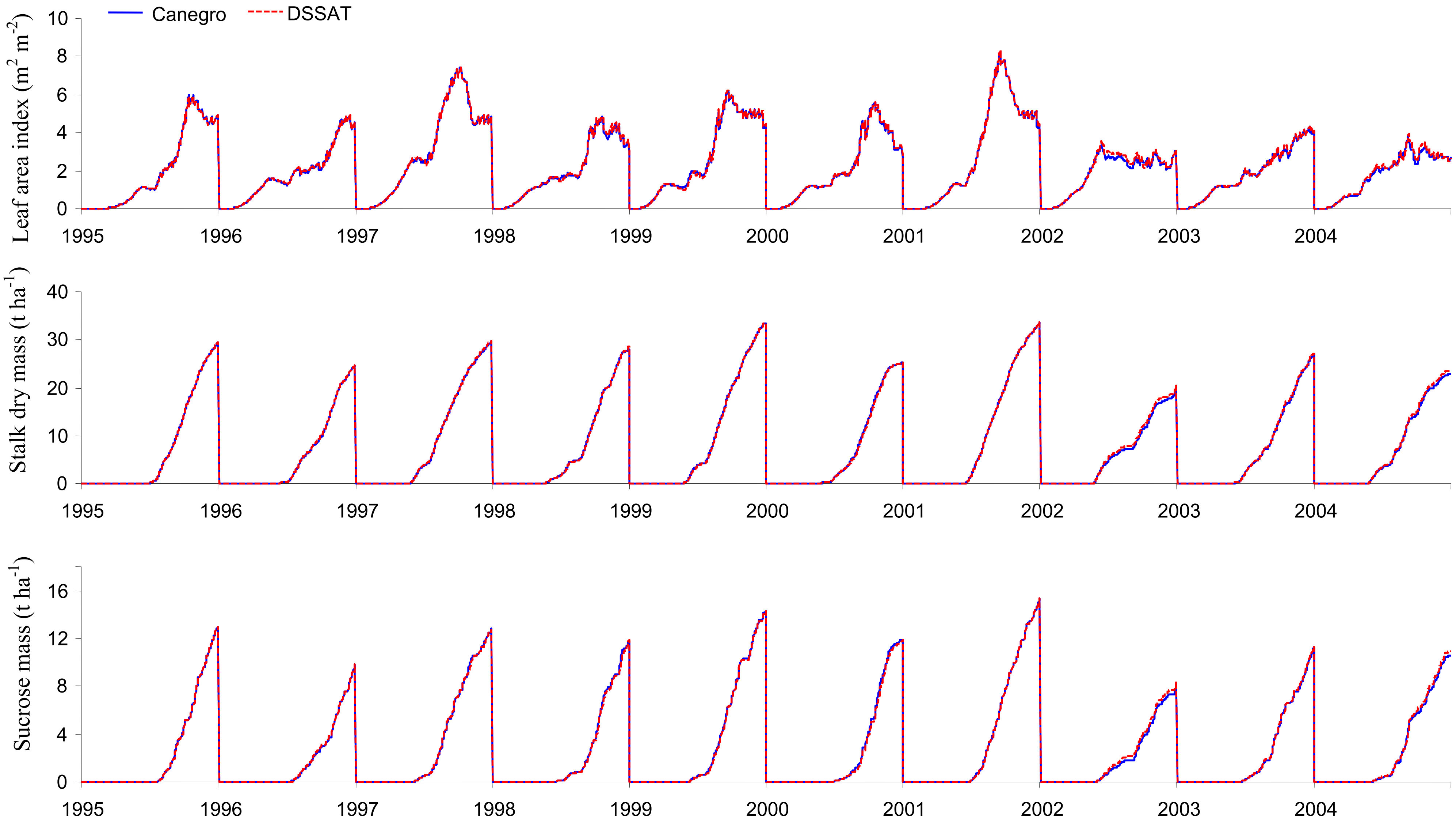


Figure 7.11 Line graphs showing leaf area index (m^2 m^{-2}), stalk dry mass (t ha^{-1}) and sucrose mass (t ha^{-1}) for Run 6 (water-stressed scenario, soil surface evaporation forced to 0 mm d^{-1} and reference evaporation forced to 4 mm d^{-1}). Simulation results from the SASRI Canegro model are blue lines marked 'Canegro'. The red lines marked 'DSSAT' display DSSAT CSM Canegro simulation results. Simulations are for a 10-year period, a plant crop followed by nine ratoons starting in October 1995 at Mount Edgecombe.

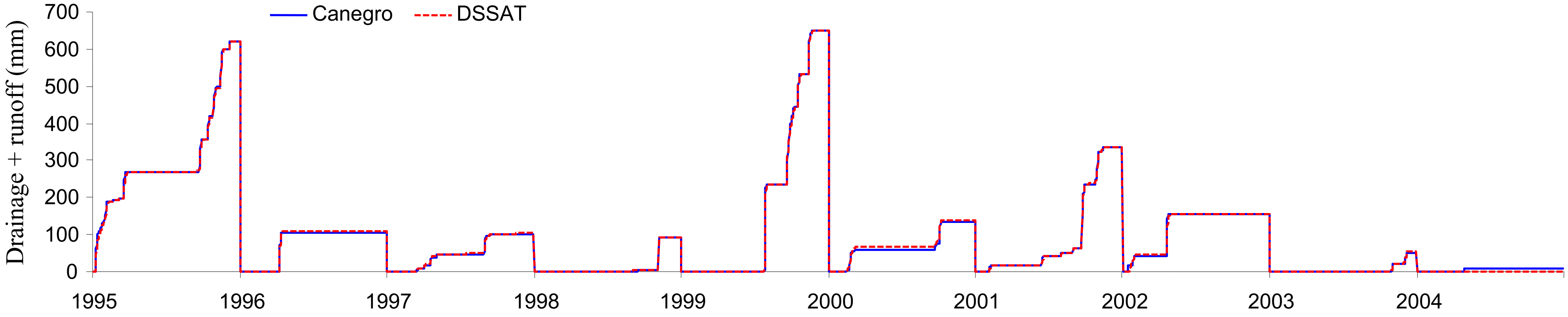


Figure 7.12 The cumulative (per season) sum of drainage and runoff (mm) simulated by SASRI Canegro (red line) and DSSAT CSM Canegro (blue line) for Run 6 (water-stressed scenario, soil surface evaporation forced to 0 mm d^{-1}, reference evaporation forced to 4 mm d^{-1}).

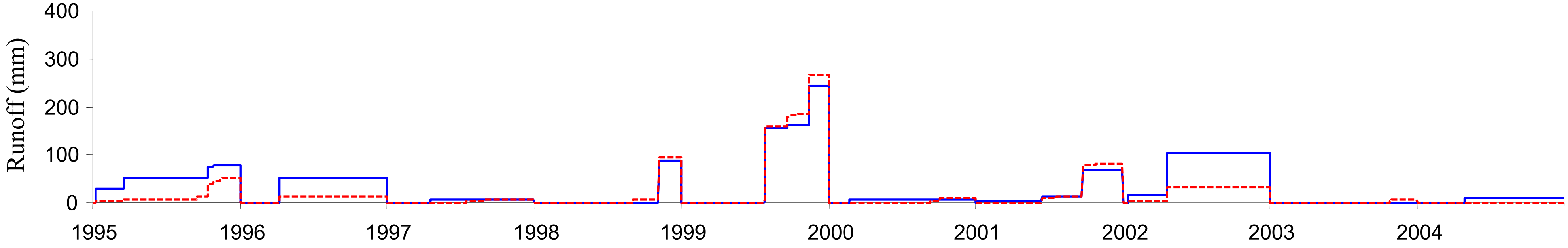


Figure 7.13 Cumulative (per season) runoff (mm) simulated by SASRI Canegro (red line) and DSSAT CSM Canegro (blue line) for Run 6 (water-stressed scenario, soil surface evaporation forced to 0 mm d^{-1}, reference evaporation forced to 4 mm d^{-1}).

The remaining discrepancies between the models can be explained by the differences in runoff (Figure 7.13). It was found, however, that the differences in runoff were largely offset by the differences in drainage. This is demonstrated in Figure 7.12, which shows the sum of runoff and drainage to be nearly identical in both models.

Run 7 forced runoff to be equal in both models, in addition to forcing *EO* to 4 mm d^{-1} and *ES* to 0 mm d^{-1}.

7.2.7 Run 7: Reference evaporation set to 4 mm d^{-1}, soil evaporation set to 0 mm d^{-1}, and runoff set to 0 mm d^{-1}

This run served to verify the observation that while simulated runoff and drainage are different between the two models, the calculations compensate for each other.

By setting runoff to 0 mm d^{-1}, all water is infiltrated and equivalent drainage would confirm the hypothesis that the processes compensate for each other. This hypothesis was borne out in the results. Very slight improvements to all variables' correlations were noted. Results are not presented.

Forcing runoff, in addition to *EO* and *ES*, yielded nearly identical results for the two models. The differences in the simulation of runoff are explained (1) mainly by the discrepancy in soil surface evaporation (*ES*), as this is an input to the SCS (Soil Conservation Service, 1972) runoff algorithm used in DSSAT and (2) partly by the slight extra sophistication of the SASRI Canegro approach, which varies the runoff curve number of the soil with crop cover.

7.3 Conclusion

Run 1 (water balances disabled) confirms that Canegro plant growth and development algorithms inside the DSSAT CSM behave identically to the algorithms inside the SASRI Canegro model. Run 2 (well-irrigated scenario) indicates that the linkages with the CSM water balance appear to function correctly, and facilitated the calibration of a crop coefficient (1.15), for scaling FAO-56 short grass reference evaporation to sugarcane reference evaporation. Results of Run 3 (water-stressed scenario) present a broad indication of the performance of the DSSAT CSM Canegro in comparison with the SASRI Canegro model. Despite the differences in the simulation of the water balance, the DSSAT CSM Canegro and SASRI Canegro models appear to simulate similarly with respect to all variables. Green leaf area index and SWDF30 (soil water stress index affecting shoot population) appear to be the state variables most sensitive to the differences in the water balance. These, however, translated into relatively small differences in stalk and sucrose mass, which are the most important model outputs for most simulation applications. The statistical measures indicate a general high level of agreement between SASRI Canegro and DSSAT CSM Canegro results.

The sensitivity analysis revealed that the apparent discrepancies between the simulation results from the two models are explained by differences in the simulation of, in order of severity: soil surface evaporation, reference evaporation and soil surface runoff.

8. VALIDATION RESULTS

8.1 Introduction

The validation of the DSSAT CSM Canegro model gives an indication of how accurate it is at simulating real crops. The model was used to simulate 16 crops in two experiments: A/Growth/HR, conducted in Pongola, 1968-1971 (Rostron, 1972a), chosen because it represents crops grown under high-potential conditions (fully irrigated, high air temperatures and solar radiation); and A/Growth/07, run at La Mercy, 1989-1990 (Inman-Bamber, 1994), water-stressed rainfed crops grown under more temperate conditions which represent more common practice in the South African sugar industry. The simulation results from the model were compared with observations made in the field. The graphs and statistics describe the accuracy and performance of the DSSAT CSM Canegro model. SASRI Canegro simulation performance figures are also shown for comparison. This serves as an additional verification that the models perform similarly. The previous chapter described the verification of the DSSAT CSM Canegro model compared with an intermediate version of the SASRI Canegro that included a number of small corrections to software errors, as well as changes to the tiller senescence calculations. The validation described in this chapter compares the DSSAT CSM Canegro model results with simulations generated by an unaltered and original version of the SASRI Canegro model. The DSSAT CSM Canegro model was intentionally not recalibrated in any way before performing these validation runs.

RMSE and APE values are expressed as percentages of the average observed value, and are named RMSE% and APE% respectively.

8.2 A/Growth/HR – Pongola 1968-71

Each of the eight treatments of the A/Growth/HR trials represents an 18-month ratoon crop started on a different date, ranging from December 1968 to January 1970. Time-series plots of simulated and observed stalk dry mass values for treatments 1, 3, 5 and 7 (corresponding with December, April, July and November crop starts), are shown in Figure 8.1. The simulated values in Figure 8.1 were generated by the SASRI Canegro model and the DSSAT CSM Canegro model. It is clear from the figure that the two models simulated the experiment very similarly.

The RMSE values for SASRI Canegro and DSSAT CSM Canegro simulations compared with observations (Table 8.1) were of very similar magnitudes. Stalk mass RMSE was 6.31 t ha^{-1} for SASRI Canegro and 6.50 t ha^{-1} for DSSAT CSM Canegro. RMSE for sucrose mass was 3.60 t ha^{-1} for both models. This serves as additional (to the verification runs) confirmation that the models simulated this experiment very similarly, although small differences are evident.

Scatter plots of SASRI Canegro vs DSSAT CSM Canegro values for stalk dry mass and sucrose mass (Figure 8.2a and b respectively) provide additional verification that the models simulated almost identically for this experiment.

The verification exercises revealed that the plant growth and development aspects of the two models almost were identical in the absence of water stress. The fully-irrigated A/Growth/HR runs were intended to be set up such that water stress was avoided, and so the SASRI Canegro and DSSAT CSM Canegro models should have produced near-identical output for these validation runs as well. A detailed investigation of the small differences between the models (evident in the figures and Table 8.1) revealed that neither models' irrigation strategy was able to conclusively eliminate water stress. DSSAT CSM Canegro simulated water stress events (of low severity) affecting expansive growth and sucrose accumulation ($SWDF_2$, described in Section 2.2.7) on several dates during the simulations. SASRI Canegro initially indicated stress affecting tiller development ($SWDF_{30}$, Section 2.2.7), but it was possible to largely eliminate this by instructing the model to irrigate smaller amounts more frequently. As the two models provide different mechanisms for triggering irrigation events (SASRI Canegro allows the user to specify green leaf number or an $SWDF_2$ threshold as a trigger, while DSSAT allows a soil water content threshold for a specified management depth), it was not practically possible for both models to have identical irrigation application amounts and timing. The irrigation system parameters chosen ensured that both models simulated practically feasible irrigation schedules. Further investigation revealed that both models simulated near-identically when the water stress indices were forced to zero-stress values. Such an approach was not considered appropriate for this validation exercise, however, so the small stress events simulated by the models were accepted.

Remaining model differences were of effectively negligible magnitude, and were satisfactorily explained by the source code modifications and corrections described in sections 6.5 and 6.6. In the model verification presented and discussed in Chapter 7, the version of SASRI Canegro used for verification had had the same code corrections made to it as the DSSAT CSM Canegro plant module, to ensure that the verification was meaningful. These validation runs were performed with an unchanged SASRI Canegro model. For this reason, small simulation output differences between the SASRI Canegro and the DSSAT CSM Canegro simulations are not necessarily inconsistent with the verification results.

The simulation of the A/Growth/HR trial showed good performance for the time-series analyses. Figure 8.1 provides a qualitative indication of similarity between DSSAT CSM Canegro simulations and observations for stalk dry mass. The RMSE values for DSSAT CSM Canegro simulations (Table 8.1) are of acceptable magnitude, by virtue of being of similar to the values (stalk dry mass 5.48 t ha^{-1}, sucrose mass 3.18 t ha^{-1}) reported by Singels and Bezuidenhout (2002), particularly as the simulations in the validation resulting in the latter figures had had their dates of emergence adjusted to fit the observations.

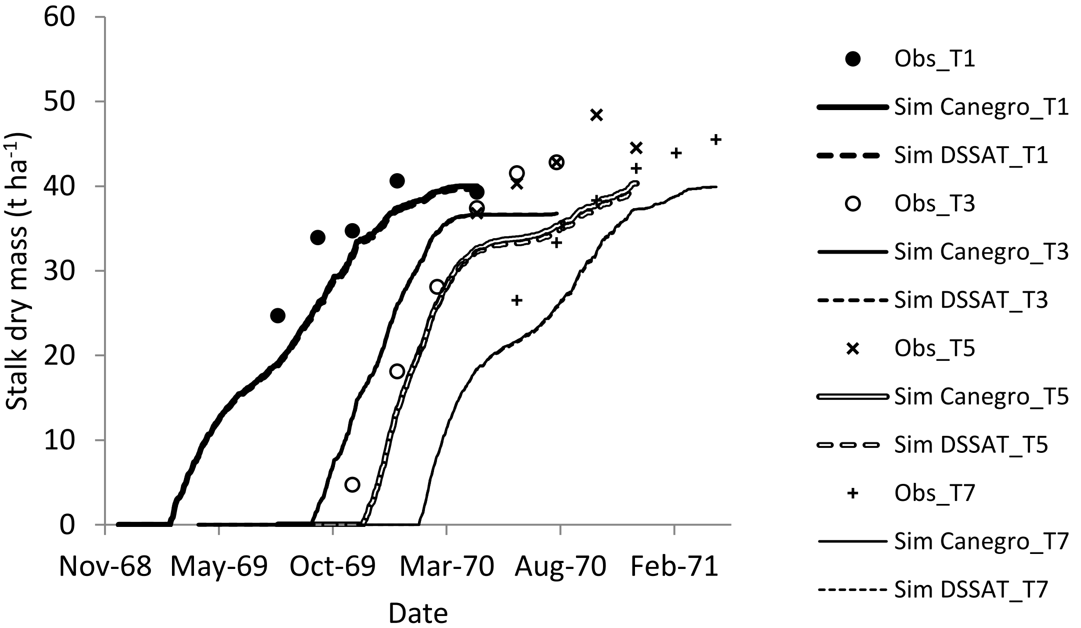


Figure 8.1 Time series plot of observed ('Obs_') and simulated ('Sim_') stalk dry mass (t ha^{-1}), for four start date treatments (December 1968 ('T1'), April 1969 ('T3'), July 1969 ('T5') and November 1969 ('T7')) for the A/Growth/HR trial. Simulated values are predicted by the SASRI Canegro model ('Canegro'), and the DSSAT CSM Canegro model ('DSSAT').

Table 8.1 Average prediction error (APE, t ha^{-1}), root mean squared error (RMSE, t ha^{-1}), maximum and minimum values of observed, DSSAT CSM Canegro-simulated and SASRI Canegro-simulated values for stalk dry mass, sucrose mass, and stalk fresh mass for the A/Growth/HR trial.

Variable name	**Model version**	**APE**	**RMSE**	**APE% avg. obs.** †	**RMSE% avg. obs.**†	**Maximum simulated***	**Maximum observed**	**Average simulated***	**Average observed**	**Minimum simulated***	**Minimum observed**
Stalk dry mass (t ha^{-1})	SASRI Canegro	3.69	6.31	10.21	17.46	40.34	50.60	32.43	36.12	12.67	4.70
	DSSAT CSM Canegro	3.96	6.50	10.96	18.00	39.95	50.60	32.16	36.12	12.61	4.70
Sucrose mass (t ha^{-1})	SASRI Canegro	1.49	3.60	9.11	22.02	19.43	26.30	14.86	16.35	3.51	0.70
	DSSAT CSM Canegro	1.57	3.60	9.58	21.99	19.41	26.30	14.79	16.35	3.48	0.70
Stalk fresh mass (t ha^{-1})	SASRI Canegro	24.37	29.62	17.01	20.68	146.02	194.00	118.88	143.25	51.07	32.00
	DSSAT CSM Canegro	25.49	30.53	17.79	21.31	144.32	194.00	117.76	143.25	50.86	32.00

*Simulated values on dates corresponding with observations

†'APE% avg. obs.' and 'RMSE% avg. obs.' are APE and RMSE values expressed as percentages of the average observed value.

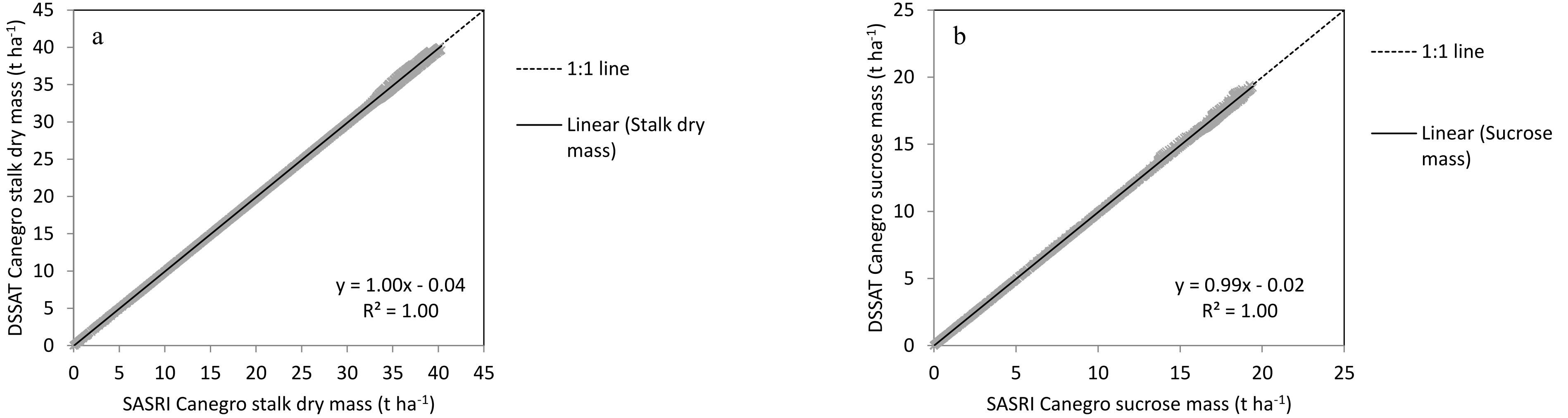


Figure 8.2 Scatter plots of (a) stalk dry mass (t ha^{-1}) and (b) sucrose mass (t ha^{-1}) values simulated by the DSSAT CSM Canegro against corresponding values simulated by SASRI Canegro, for the A/Growth/HR trial. SASRI Canegro values are plotted on the x-axis and the DSSAT CSM Canegro-simulated values on the y-axis.

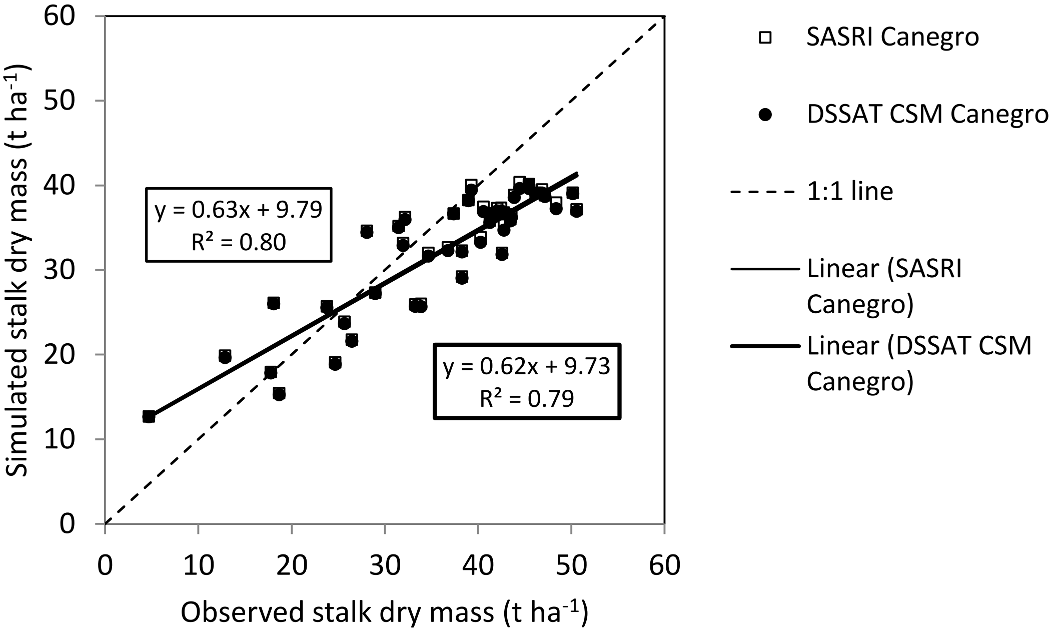


Figure 8.3 Scatter plot of stalk dry mass (t ha^{-1}) values, observed and simulated by DSSAT CSM Canegro and SASRI Canegro, for the A/Growth/HR trial.

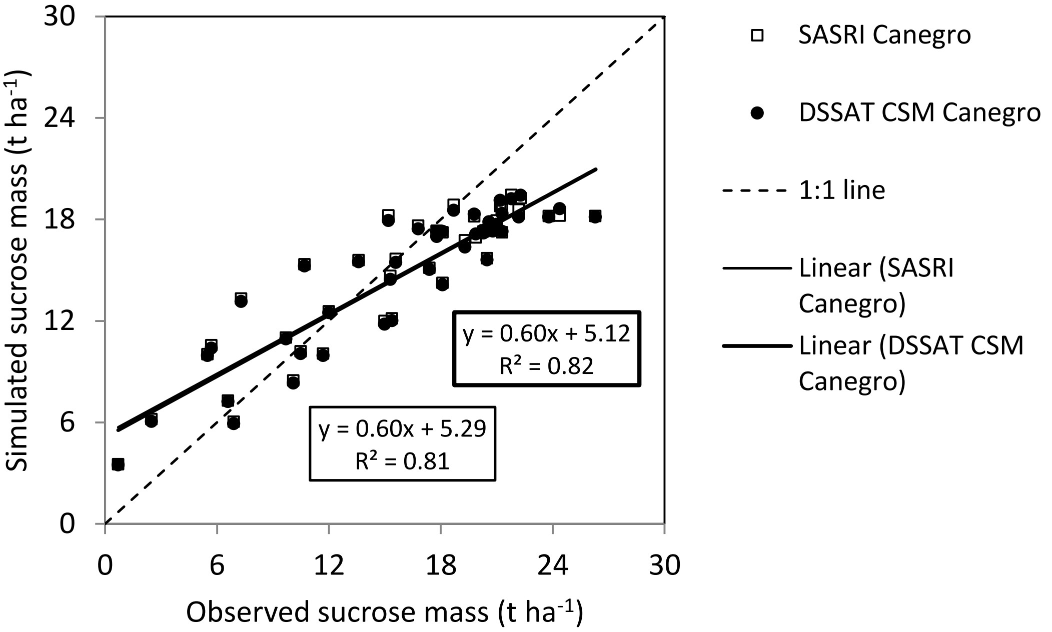


Figure 8.4 Scatter plot of sucrose mass (t ha^{-1}) values, observed and simulated by DSSAT CSM Canegro and SASRI Canegro, for the A/Growth/HR trial.

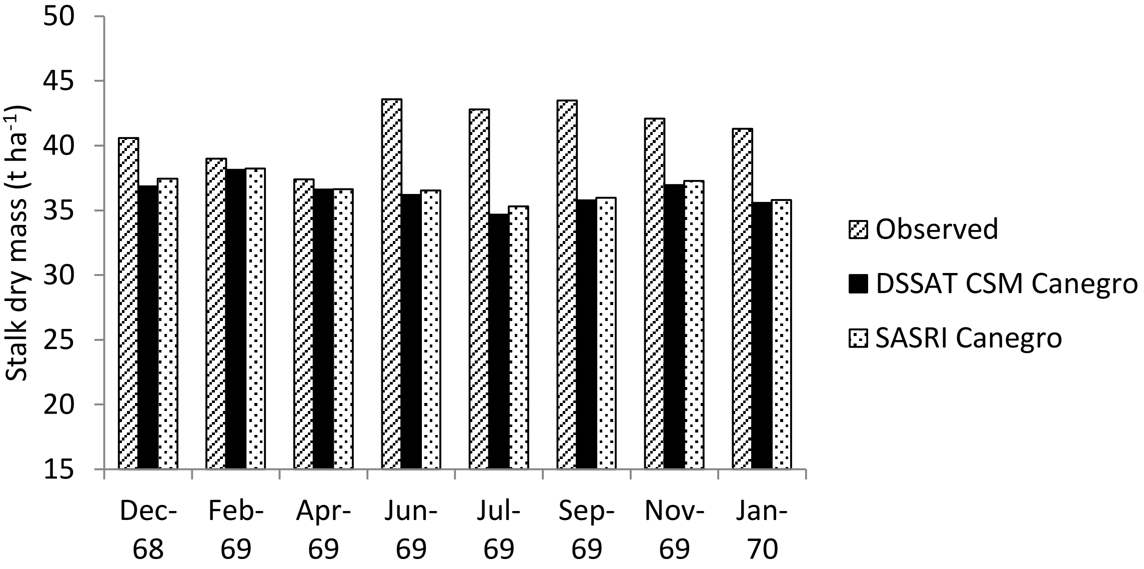


Figure 8.5 Stalk dry mass (t ha^{-1}) at 13 months' age, observed and simulated by SASRI Canegro and DSSAT CSM Canegro, per treatment (identified by start month and year), for the A/Growth/HR trial.

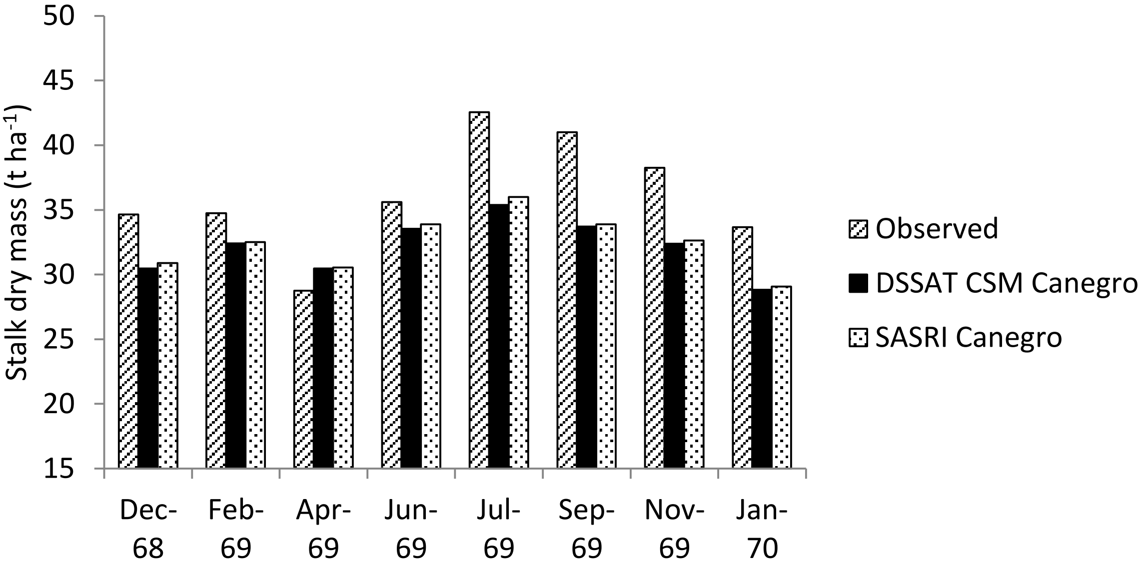


Figure 8.6 Average stalk dry mass (t ha^{-1}), observed and simulated by SASRI Canegro and DSSAT CSM Canegro, per treatment, for the A/Growth/HR trial.

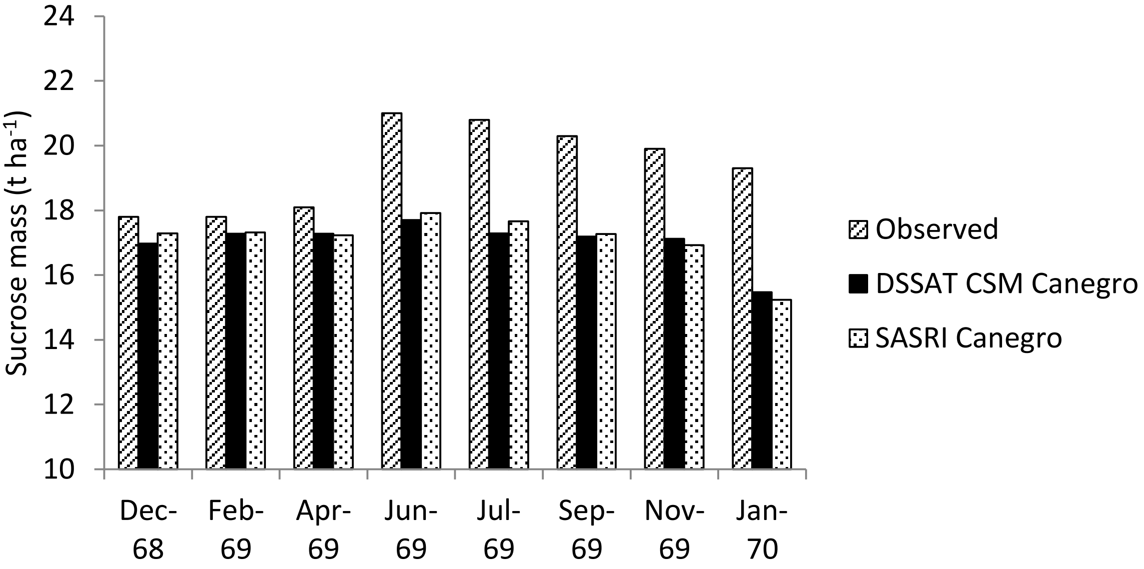


Figure 8.7 Sucrose mass (t ha^{-1}) at 13 months' age, observed and simulated by SASRI Canegro and DSSAT CSM Canegro, per treatment, for the A/Growth/HR trial.

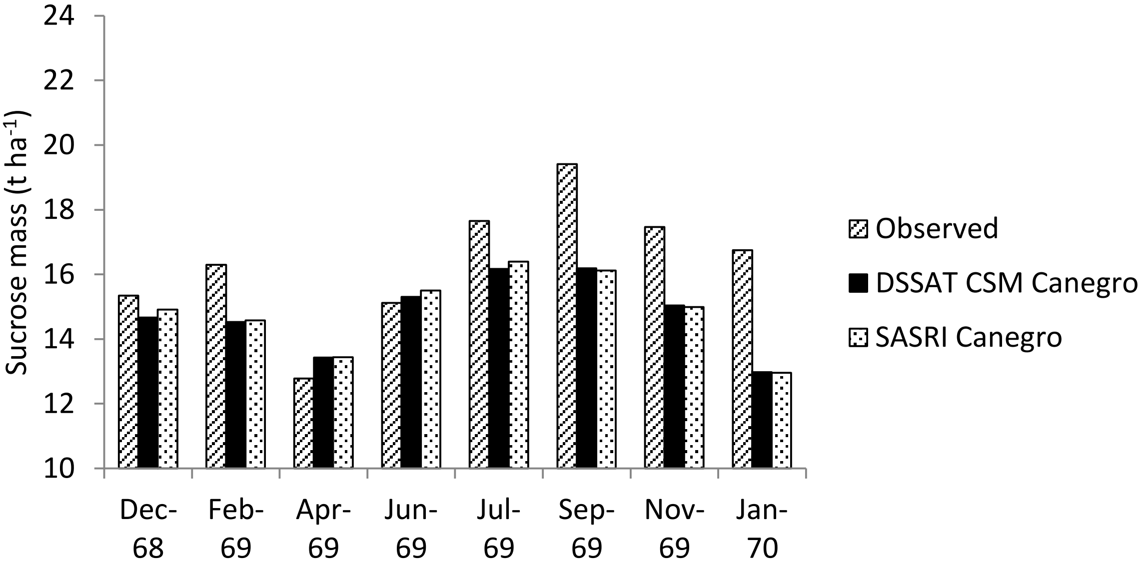


Figure 8.8 Average sucrose mass (t ha^{-1}), observed and simulated by SASRI Canegro and DSSAT CSM Canegro, per treatment, for the A/Growth/HR trial.

The scatter plots of simulated vs observed stalk dry mass (Figure 8.3) and sucrose mass (Figure 8.4) show good agreements. The slope of linear regression statistic for stalk dry mass for DSSAT CSM Canegro is 0.62 (0.63 for SASRI Canegro). While these values would ideally be 1, what matters more is that a clearly positive relationship exists between the simulated and observed values, and that the R^2 values are high – 0.80 and 0.79 respectively for DSSAT CSM Canegro and SASRI Canegro – suggesting that the models capture the subtleties of observations well. The scatter plots also show that the DSSAT CSM Canegro and SASRI Canegro models simulate very similarly. Scatter plots for sucrose mass (Figure

8.4) show similarly acceptable prediction accuracy – slopes of 0.60 for both models and R^2 values of 0.81 and 0.82 for SASRI Canegro and DSSAT CSM Canegro respectively.

Figures 8.5-8.8 illustrate the DSSAT CSM Canegro model's ability to simulate seasonal differences. Two indicators of model performance in this regard are presented: average simulated values were compared with average observed values over the full duration of each crop (stalk dry mass is shown in Figure 8.6 and sucrose mass in Figure 8.8); and values corresponding with a 13-month harvest are shown in Figures 8.5 (stalk dry mass) and 8.7 (sucrose mass). Values are reported at 13 months' age rather than at harvest (18 months) because this is a more realistic harvest age at Pongola, and so represents a more practically useful and relevant measure of model performance.

The simulation performance of the model at 13 months' age is poor (Figures 8.5 and 8.7). Seasonal differences for this experiment are explained primarily by differences in air temperature and intensity of solar radiation, as water stress was negligible. Recent research by Smit (2010) suggests that the choice of base temperatures for calculating thermal time for emergence and the start of stalk elongation may need revision. Implementing such changes in the DSSAT CSM Canegro model might facilitate better seasonal differentiation on the basis of air temperature, although other model parameters, such as the thermal time to emergence (*TTPLNTEM* and *TTRATNEM*, °C d) and the thermal time to the stalk of stalk elongation (*CHUPIBASE*, °C d), will likely require corresponding adjustments.

The comparison of average simulated and observed values, per treatment, gives a better account of average model performance (Figures 8.6 and 8.8). The rationale behind presenting model performance data in this averaged form (in addition to values at harvest) is to provide a description of model performance for the season as a whole, rather than just at harvest when values might, for example, be affected by lodging events not sufficiently captured by the lodging algorithms in the models. The additional measurement events considered for the average values are mostly taken at younger crop stages. The improved performance of the models on average compared with values at harvest only suggests that the model is more accurate for younger crops.

The better performance shown by Figure 8.6 shows that the relative change in simulated stalk dry mass between treatments is captured far better (than values at 13 months) by the model, but the seasonal (treatment) differences are less accentuated than in observations. The differences between average values simulated by SASRI Canegro and DSSAT CSM Canegro simulations are small, but SASRI Canegro nevertheless provides a slightly more accurate simulation of seasonal differences. This is attributed, as discussed above, to water stress being simulated by DSSAT CSM Canegro. Both models tended to under-estimate observed values.

8.3 A/Growth/07 – La Mercy 1989-90

The A/Growth/07 experiment comprised eight treatments representing 18-month ratoon crops starting on different dates, ranging from June 1989 to August 1990, grown under rainfed conditions at La Mercy on the north coast of KwaZulu-Natal in South Africa.

The time series plot of stalk dry mass (Figure 8.10) shows that some quite significant differences between the simulated values for the two models are apparent. Both this figure and the verification analysis of this experiment (Figure 8.9) indicate that in general, the DSSAT CSM Canegro model underestimated values this experiment relative to SASRI Canegro. This is confirmed by the scatter plots of simulated vs observed values (Figures 8.11-8.14), which generally show reduced slopes of linear regression (with similar intercept values) for DSSAT CSM Canegro compared with SASRI Canegro.

The scatter plot of simulated vs observed above-ground dry biomass (Figure 8.11) reveals a general trend that both the DSSAT CSM Canegro and SASRI Canegro models tend to over-simulate low values and under-simulate high values. The DSSAT CSM Canegro model had a smaller slope of linear regression statistic (0.66) than that of the SASRI Canegro model (0.78) for roughly equivalent intercept values (14.2 and 12.6 t ha^{-1} respectively). However, the R^2 value of the DSSAT CSM Canegro (0.75) was slightly greater than that of SASRI Canegro (0.72). The APE and RMSE statistics (Table 8.2) favoured DSSAT CSM Canegro (-0.35 and 6.02 t ha^{-1}, compared with -3.35 and 7.18 t ha^{-1} respectively for SASRI Canegro). The graph showing average above-ground dry mass (*ADM,* t ha^{-1}) per treatment (Figure 8.15) clearly shows that SASRI Canegro consistently over-estimates *ADM*, whereas DSSAT CSM Canegro simulated ADM more realistically.

Table 8.2 Average prediction error (APE, t ha^{-1}), root mean squared error (RMSE, t ha^{-1}), maximum and minimum values of observed, DSSAT CSM Canegro-simulated and SASRI Canegro-simulated values for stalk dry mass (t ha^{-1}), sucrose mass (t ha^{-1}), and stalk fresh mass (t ha^{-1}) for the eight treatments of the A/Growth/07 trial.

Variable name	**Model version**	**APE**	**RMSE**	**APE% avg. obs.†**	**RMSE % avg. obs.†**	**Max. simu-lated***	**Max. observed**	**Average simu-lated***	**Average observed**	**Min. simu-lated***	**Minimum observed**
Stalk dry mass (t ha^{-1})	SASRI Canegro	0.42	5.12	1.7	20.5	38.3	42.5	24.6	25.0	11.4	2.8
	DSSAT CSM Canegro	4.14	6.75	16.5	27.0	31.9	42.5	20.9	25.0	6.9	2.8
Sucrose mass (t ha^{-1})	SASRI Canegro	0.21	2.70	1.9	23.8	17.9	19.9	11.1	11.3	3.4	0.5
	DSSAT CSM Canegro	1.92	3.58	16.9	31.6	15.3	19.9	9.4	11.3	1.6	0.5
Above-ground dry biomass (t ha^{-1})	SASRI Canegro	-3.53	7.18	-8.6	17.6	63.0	62.8	44.4	40.8	20.4	10.6
	DSSAT CSM Canegro	-0.35	6.02	-0.9	14.7	60.4	62.8	41.2	40.8	20.4	10.6
Green leaf area index (m^2 m^{-2})	SASRI Canegro	-0.72	1.08	-23.5	35.2	5.0	5.1	3.8	3.1	2.7	1.3
	DSSAT CSM Canegro	0.17	0.70	5.5	22.8	4.0	5.1	2.9	3.1	2.1	1.3

*Simulated values on dates corresponding with observations

†'APE% avg. obs.' and 'RMSE% avg. obs.' are APE and RMSE values expressed as percentages of the average observed value.

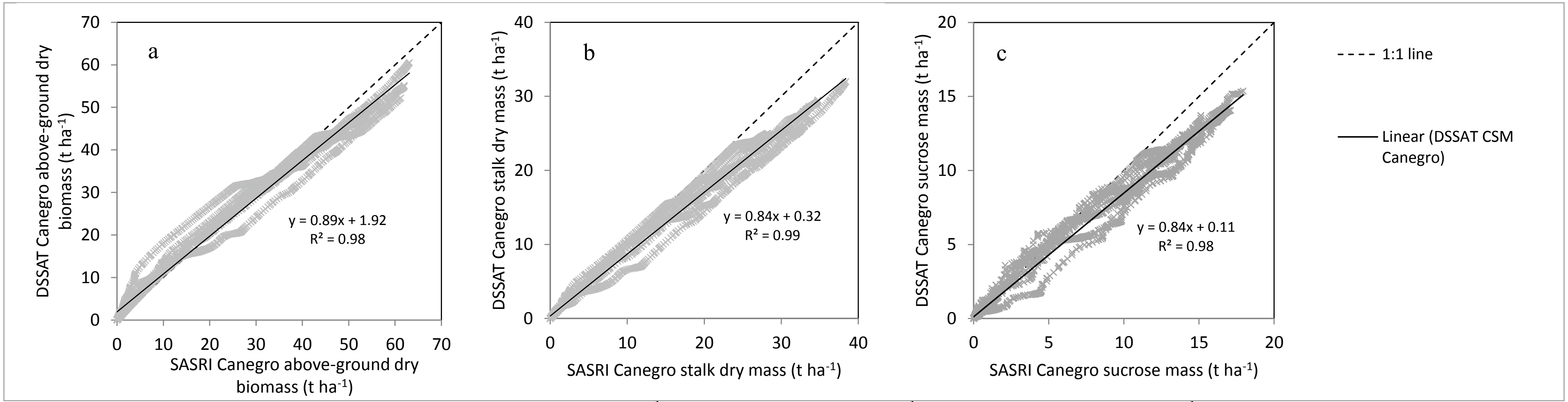


Figure 8.9 Scatter plots of (a) above-ground dry biomass (t ha^{-1}), (b) stalk dry mass (t ha^{-1}) and (c) sucrose mass (t ha^{-1}) values simulated by DSSAT CSM Canegro against corresponding values simulated by SASRI Canegro, for the A/Growth/07 trial.

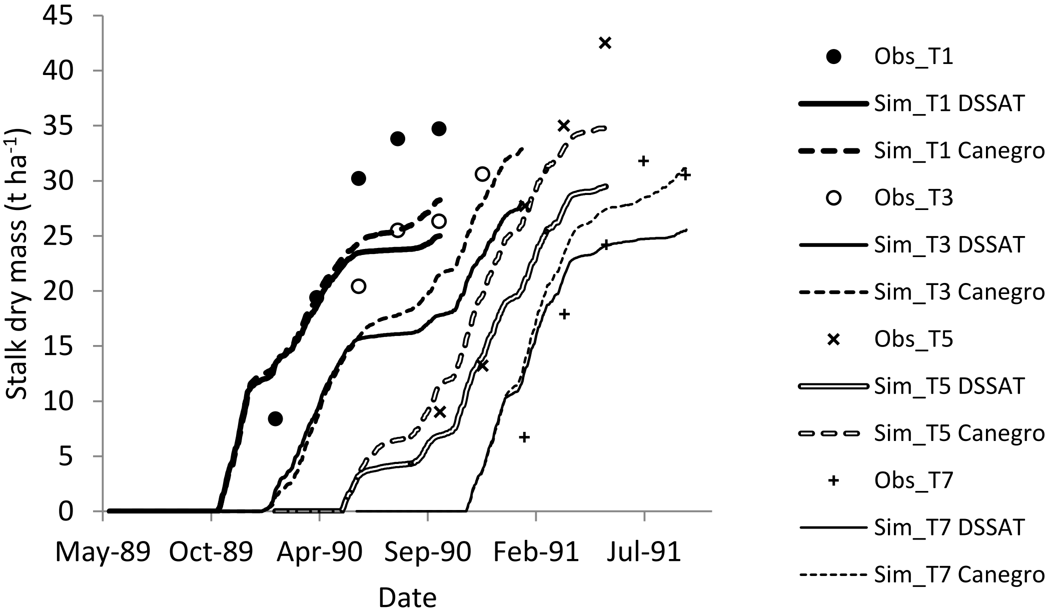


Figure 8.10 Time series plot of observed (series names marked 'Obs_') and simulated ('Sim_') stalk dry mass (t ha^{-1}), for four start date treatments (June 1989 ('T1'), October 1989 ('T3'), February 1990 ('T5') and June 1990 ('T7')) , for the A/Growth/07 trial.

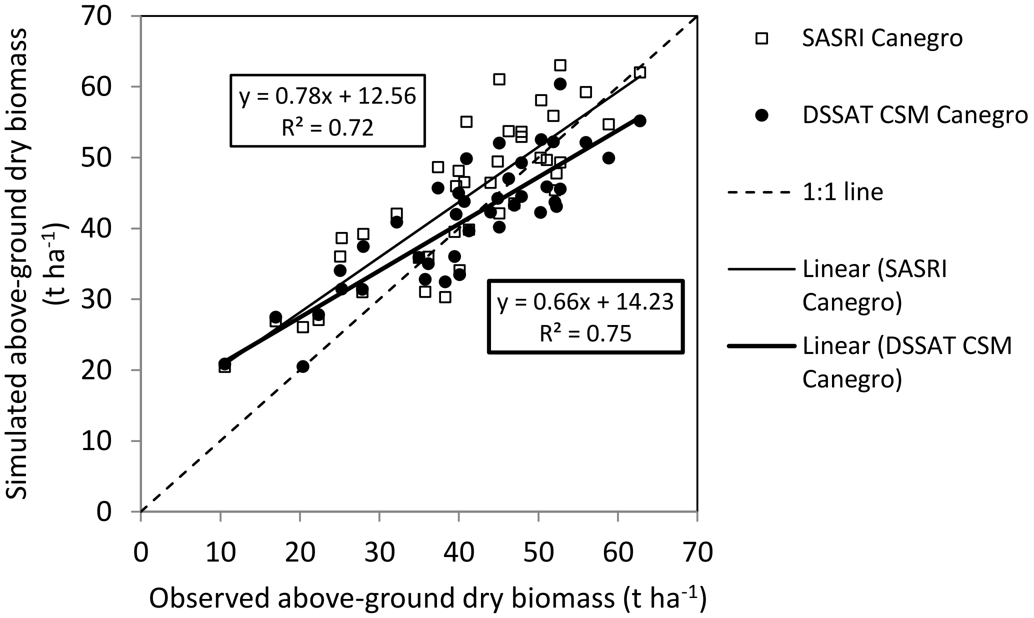


Figure 8.11 Scatter plot of above-ground dry biomass (t ha^{-1}) values, observed and simulated by DSSAT CSM Canegro and SASRI Canegro, for the A/Growth/07 trial.

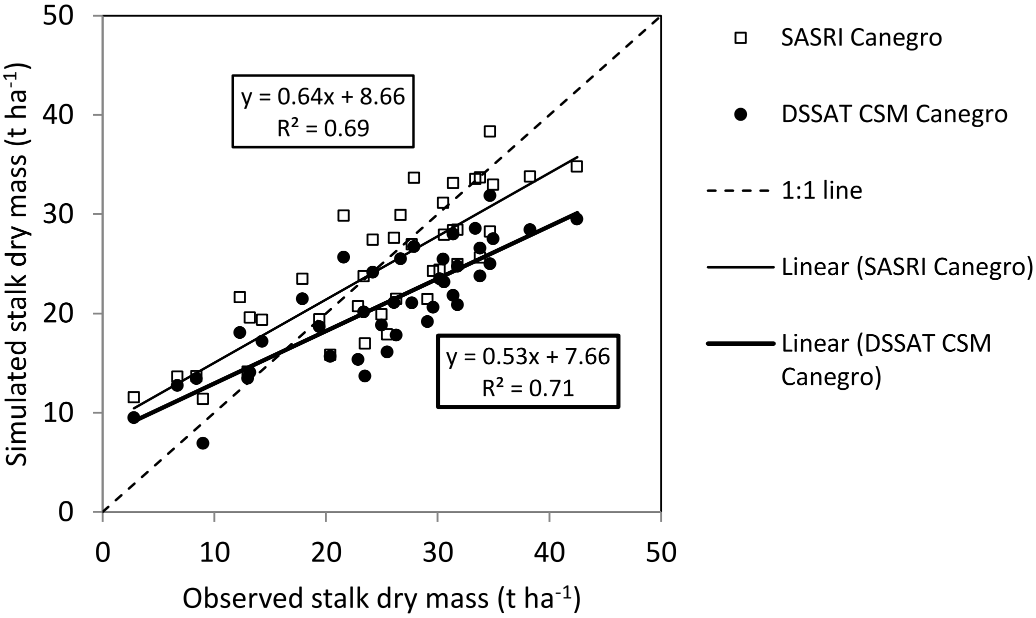


Figure 8.12 Scatter plot of stalk dry mass (t ha^{-1}) values, observed and simulated by DSSAT CSM Canegro and SASRI Canegro, for the A/Growth/07 trial.

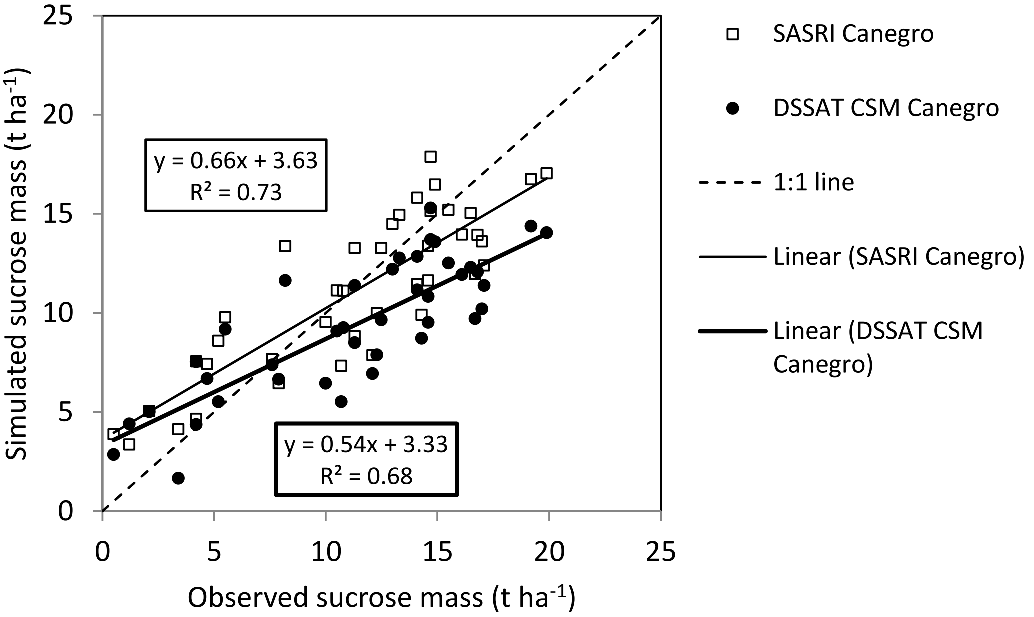


Figure 8.13 Scatter plot of sucrose mass (t ha^{-1}*) values, observed and simulated by DSSAT CSM Canegro and SASRI Canegro, for the A/Growth/07 trial.*

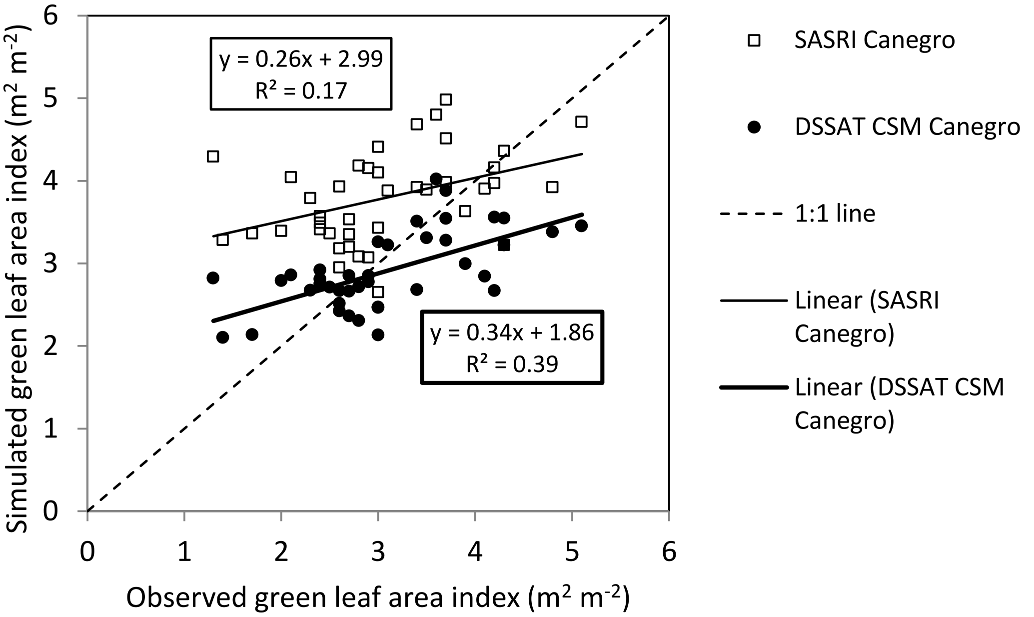


Figure 8.14 Scatter plot of green leaf area index (m^{2} *m*$^{-2}$*) values, observed and simulated by DSSAT CSM Canegro and SASRI Canegro, for the A/Growth/07 trial.*

ADM is largely determined by photosynthetically-active radiation interception (i.e. canopy size) and water stress, and it is interesting to note that the SASRI Canegro model consistently over-estimated green leaf area index (*GLAI*, Figure 8.14 and Figure 8.20), whereas the DSSAT CSM Canegro model gave better correspondence.

The APE value for *GLAI* of SASRI Canegro (-0.72 $m^2.m^{-2}$) confirms this; it is also nearly five times the magnitude of the APE of DSSAT CSM Canegro (0.17 $m^2.m^{-2}$) (Table 8.2). The RMSE values for *GLAI* also favour DSSAT CSM Canegro (0.7 t ha^{-1} vs 1.08 t ha^{-1}). The DSSAT CSM Canegro model showed a higher R^2 value for *GLAI* than SASRI Canegro (Figure 8.14). This suggests that an over-estimation of *GLAI* led to an over-estimation of radiation interception and photosynthesis rates (further enhanced by lower water stress) by SASRI Canegro, resulting in higher *ADM* values. It must be acknowledged that neither model gave a particularly good account of *GLAI*, with relatively poor R^2 values of 0.17 (SASRI Canegro) and 0.39 (DSSAT CSM Canegro). The DSSAT CSM Canegro model's smaller simulated values of *GLAI* are attributed to greater rates of water stress, arising from differences in reference evaporation

DSSAT CSM Canegro values are lower than, but otherwise consistent with, SASRI Canegro values (Figures 8.16-8.19). The values are also mostly lower than observed values. The SASRI Canegro model showed better correspondence with stalk dry mass statistics (APE = 0.42 and RMSE = 5.12 t ha^{-1}) than the DSSAT CSM Canegro model (APE = 4.14 and RMSE = 6.75 t ha^{-1}). There is a much larger difference in APE relative to RMSE between the models. SASRI Canegro model's *STKDM* values were more realistic partly because *ADM* was over-estimated: the carbohydrate partitioning ratio of *ADM* to *STKDM* is the same for both models, and is possibly too low – which in SASRI Canegro was compensated by the over-estimation of *ADM*.

The verification exercises indicated that the calculation of water balance variables (*EO* (reference evaporation, mm d^{-1}), *ES* (soil surface evaporation, mm d^{-1}), runoff and drainage) explained the discrepancies between the models' outputs. In order to confirm this, and to better understand the causes of the discrepancies in these validation runs, a single treatment was selected for further scrutiny. Treatment 5 was chosen as it that showed the greatest discrepancy in stalk dry mass (Figure 8.10). This treatment is characterised by a number of periods in which the rate of stalk dry mass accumulation (for both models) slows noticeably. Analysis of daily output of the models revealed that these were caused by water stress events (rather than low temperatures).

In the DSSAT CSM Canegro treatment 5 run, stalk dry mass accumulated at a noticeably lower rate than SASRI Canegro from approximately 120 days after planting. This is the first water stress period. After a period of about 20 days, the DSSAT CSM Canegro's stalk mass accumulation appeared to recover to a rate near to that of SASRI Canegro. Analysis of daily simulated output revealed that green leaf number per shoot was similar between the models throughout the simulation, and tiller population was identical from 50 to 250 days after planting. The output discrepancies were not, therefore, explained by differences in simulated phenology. Analysis of daily simulated water stress ($SWDF_1$) revealed a strong similarity in timing of water stress events between the models. However, the DSSAT CSM Canegro model consistently estimated slightly earlier onset of stress (about 15 days earlier in the case of the first water stress event), and more rapid development of water stress. These stress periods typically ended on the same days in both models. The result of this is that the DSSAT CSM Canegro model simulated longer and more severe water stress events than Canegro.

The impacts of this additional water stress, for DSSAT CSM Canegro in comparison with SASRI Canegro, were reduced *GLAI* (2.6 vs 3.5 $m^2\ m^{-2}$ on average, respectively) and (consequently) reduced interception of PAR (particularly after peak shoot population), which, combined with increased water stress, resulted in decreased photosynthesis rates which reduced dry mass accumulation rates. PAR interception was an average 5% lower for DSSAT CSM Canegro than SASRI Canegro.

The cause of the more severe water stress in the DSSAT CSM Canegro model appeared to be driven by differences in *EO* and *ES*.

EO was simulated by DSSAT CSM Canegro to be, on average, 12.5% higher than SASRI Canegro. The greater simulated atmospheric demand for water, combined with the proportionally greater increase in demand water lost to evaporation from the soil surface, resulted in reduced soil water availability and stronger $SWDF_1$ and $SWDF_2$ factors (in Eqns (44) and (45), the *EOP* term is larger via Eqn (38) ($EREF \approx EO$), reducing the $SWDF_1$ and $SWDF_2$ values).

Simulated *ES* was, on average, 17% higher for DSSAT CSM Canegro compared with SASRI Canegro. The calculation of *EOS* (potential soil surface evaporation, mm d^{-1}) is determined by *EO* and PAR interception; *ES* is limited by *EOS* and simulated water availability in soil layers near to the soil surface. Given the dynamic nature of these interactions, it is difficult to confirm whether or not differences in the *ES*-related algorithms (such as the calculation of soil water content near the soil surface, which is affected by runoff as well as the *ES* calculation itself) also partly explain the differences in water stress.

This analysis confirms the differences in the models' outputs for the rainfed validation experiment runs are explained by differences in water balance-related variables. These are calculated by common CSM modules outside of the plant module, and so the differences are not due to an error or omission in the CSM Canegro plant module code. It is recommended that future work include a more detailed investigation of the differences between the water balance algorithms, in order to develop appropriate adjustments to the DSSAT CSM Canegro plant module code or parameters.

The models' accounts of seasonal differences for these water-stressed situations appear better than those of the fully irrigated experiment. In every plant process in both versions of Canegro, water stress indices are used to limit rates from maximum, potential rates (fully-irrigated conditions) to attainable rates given water availability. Given that the seasonal differentiation of fully-irrigated crops (representing growth at potential rates) was demonstrated in A/Growth/HR validation exercise to be poor, it appears that the seasonal variation in rainfall availability, rather than air temperature and/or radiation, is the overwhelming determinant of seasonal variation in these rainfed simulations.

The simulation of seasonal *ADM* differences, on average (Figure 8.15), was reasonably good by both models. Both models gave a better account of seasonal (treatment) differences at 14 months' age (a typical age at harvest in the Pongola region, Figure 8.16), compared with the average of all corresponding simulated and observed values per treatment (Figure 8.17). The simulation of seasonal differences in sucrose mass was poor (relative to stalk dry mass) by both models at 14 months' age (Figure 8.18), and slightly better on average (Figure 8.19). Both models simulated conservatively, in that inter-treatment variability of simulated values was generally lower than with the observations (Figures 8.15-8.20).

The RMSE values of 6.75 and 3.58 t ha^{-1} respectively for stalk dry mass and sucrose mass reported for the DSSAT CSM Canegro simulations (Table 8.2) are similar to the values reported by Singels and Bezuidenhout (2002) (stalk dry mass 5.48 t ha^{-1}, sucrose mass 3.18 t ha^{-1}).

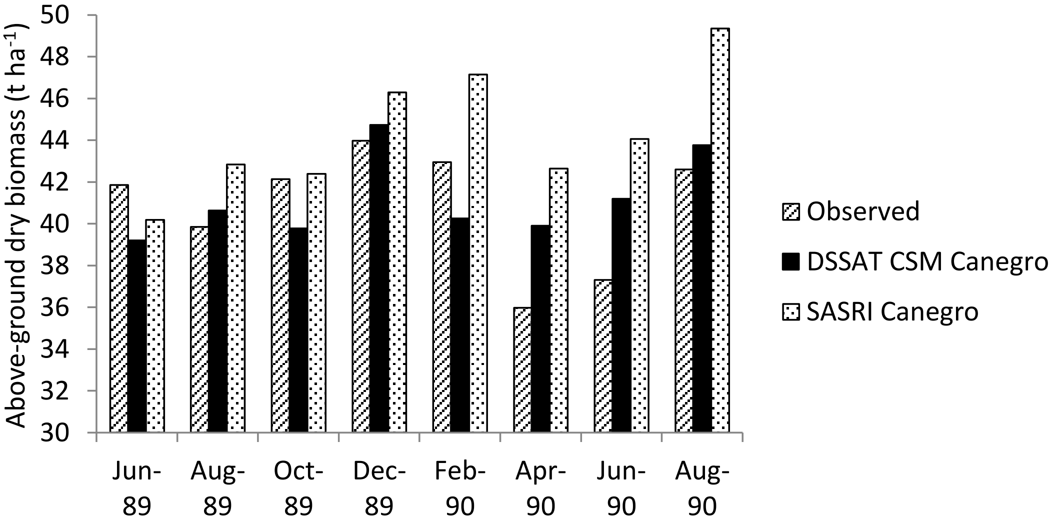


Figure 8.15 Average (of corresponding observed and simulated time-series values, per-treatment) above-ground dry biomass (t ha^{-1}), observed and simulated by SASRI Canegro and DSSAT CSM Canegro, per treatment (identified by start month and year) , for the A/Growth/07 trial.

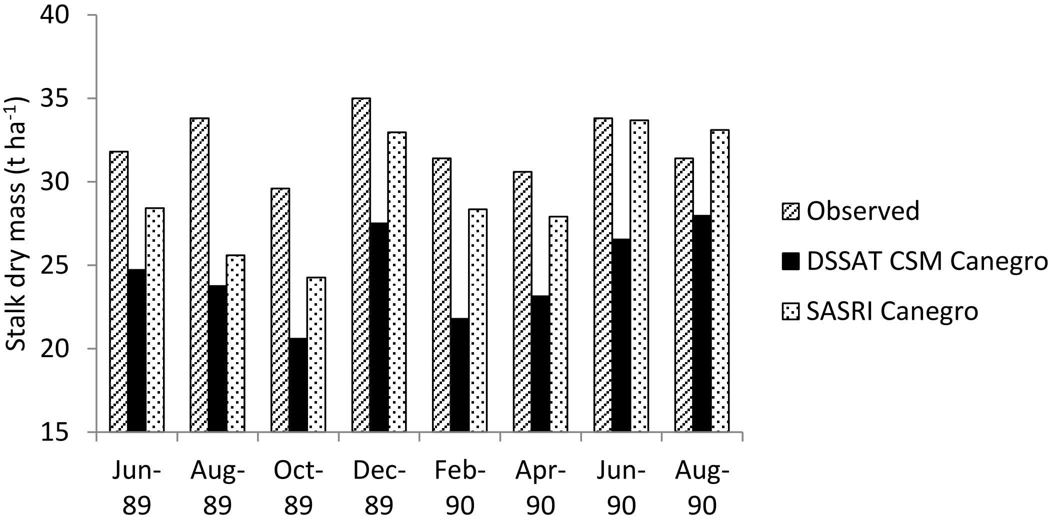


Figure 8.16 Stalk dry mass (t ha^{-1}) at 14 months, observed and simulated by SASRI Canegro and DSSAT CSM Canegro, for the A/Growth/07 trial.

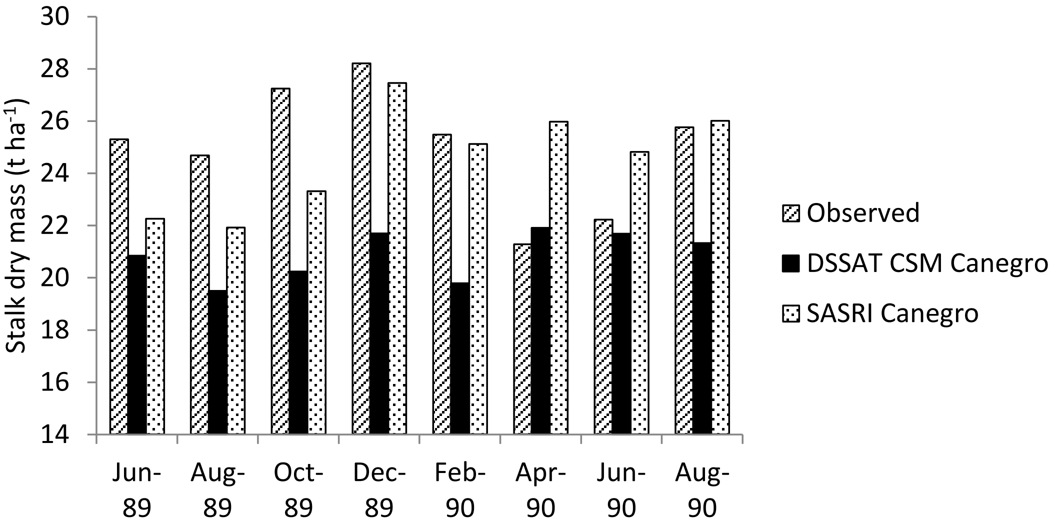


Figure 8.17 Average (of corresponding observed and simulated time-series values, per-treatment) stalk dry mass (t ha^{-1}), observed and simulated by SASRI Canegro and DSSAT CSM Canegro, for the A/Growth/07 trial.

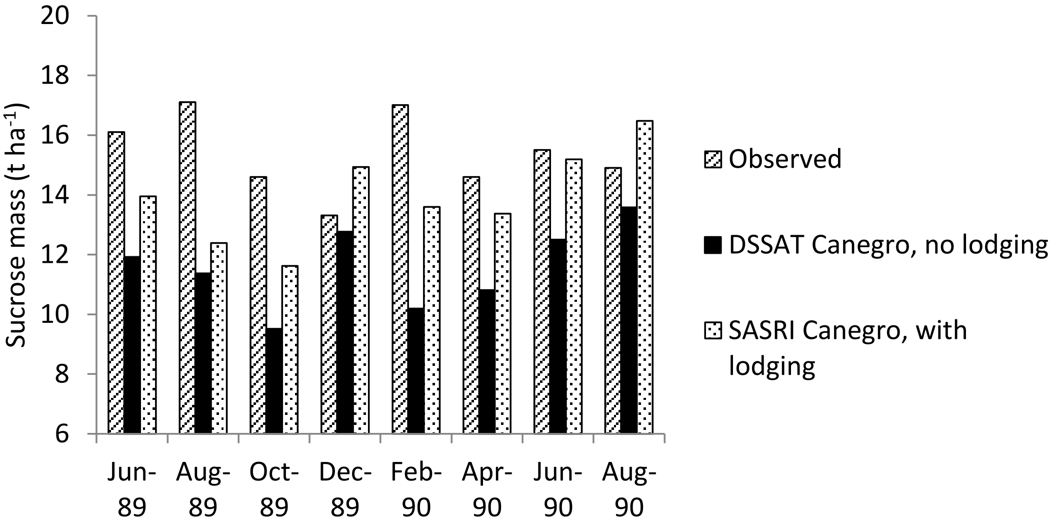


Figure 8.18 Sucrose mass (t ha^{-1}) at 14 months, observed and simulated by SASRI Canegro and DSSAT CSM Canegro, for the A/Growth/07 trial.

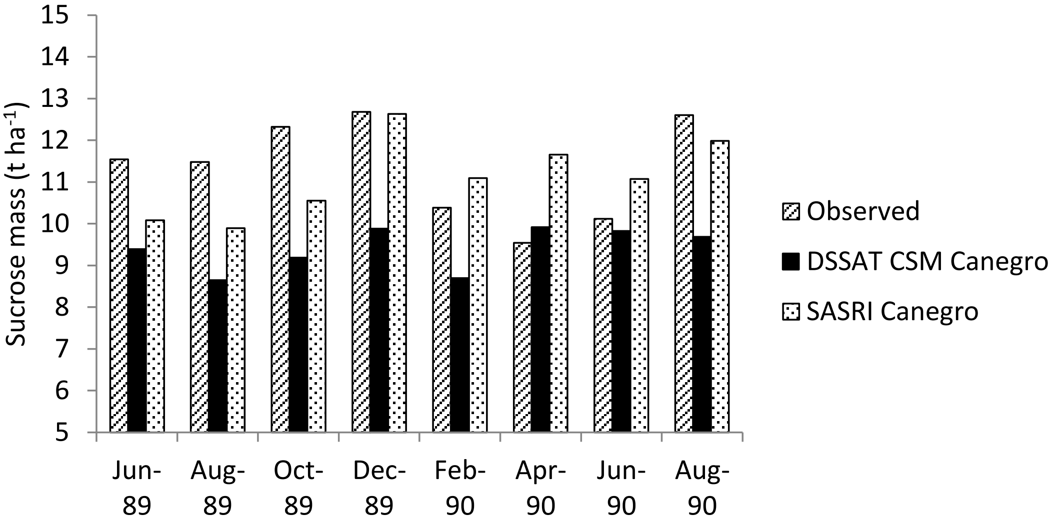


Figure 8.19 Average (of corresponding observed and simulated time-series values, per-treatment) sucrose mass (t ha^{-1}), observed and simulated by SASRI Canegro and DSSAT CSM Canegro, for the A/Growth/07 trial.

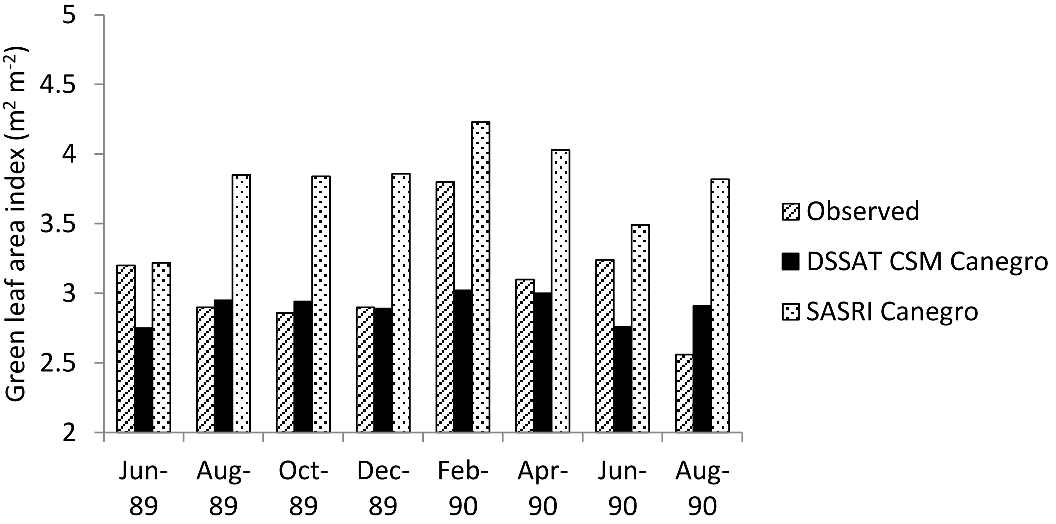


Figure 8.20 Average (of corresponding observed and simulated time-series values, per-treatment) green leaf area index (m^{2} m^{-2}), observed and simulated by SASRI Canegro and DSSAT CSM Canegro, for the A/Growth/07 trial.

8.4 Combined Experiments

The A/Growth/HR and A/Growth/07 trial datasets were combined and then analysed together to provide a general indication of model performance.

Table 8.3 Average prediction error (APE, t ha^{-1}), root mean squared error (RMSE, t ha^{-1}), and average values of observed, DSSAT CSM Canegro-simulated and SASRI Canegro-simulated values for stalk dry mass (t ha^{-1}) and sucrose mass (t ha^{-1}), for the A/Growth/HR and A/Growth/07 trials combined.

Variable name	**Model version**	**APE**	**RMSE‡**	**APE% avg. obs. †**	**RMSE% avg. obs. †**	**Average simulated***	**Average observed**
Stalk dry mass (t ha^{-1})	SASRI Canegro	2.13	5.77 (*5.48*)	6.91	18.72	28.7	30.83
	DSSAT CSM Canegro	4.05	6.62 (*5.48*)	13.12	21.48	26.79	30.83
Sucrose mass (t ha^{-1})	SASRI Canegro	1.00	3.22 (*3.18*)	7.11	22.86	13.16	14.07
	DSSAT CSM Canegro	1.82	3.59 (*3.18*)	12.93	25.48	12.34	14.07

*Simulated values on dates corresponding with those of observations

†'APE% avg. obs.' and 'RMSE% avg. obs.' are APE and RMSE values expressed as percentages of the average observed value.

‡Values in brackets are RMSE values published for SASRI Canegro in Singels and Bezuidenhout (2002).

The APE value of stalk dry mass for the DSSAT CSM Canegro model was 4.05 t ha^{-1}, compared with 2.13 t ha^{-1} for the SASRI Canegro model, while for sucrose mass the APE values were 1.82 and 1.00 t ha^{-1} respectively (Table 8.3). These indicate that the DSSAT CSM Canegro model under-estimates values compared with SASRI Canegro simulations, because a higher APE indicates a larger average error between observed and simulated values, and the positive signs of the APE values indicate that simulated values were lower than observed values. On average, the SASRI Canegro model produced estimates closer to the observed values in the experiments (as evidenced by smaller APE values).

The difference in stalk dry mass APE values between the models for the combined experiments was 1.93 t ha^{-1}. This APE discrepancy was attributed to water stress differences arising from differences in the reference evaporation and water balance modules outside of the DSSAT CSM Canegro module in the non-irrigated A/Growth/07 experiment, as the APE differences for stalk mass in the A/Growth/HR experiment were negligible (3.96 vs. 3.69, Table 8.1).

The discrepancy in stalk dry mass *RMSE* values (6.62 vs. 5.77 t ha^{-1}, Table 8.3) is relatively much smaller than that of *APE*. This indicates that the models simulate very similarly, but for a tendency by DSSAT CSM Canegro to under-estimate observed values by 4 t ha^{-1} on average (the APE value). The RMSE discrepancy is attributed mainly to differences in the water balance, as the RMSE value discrepancy for the non-stressed trial was negligible compared with the water-stressed trial. The verification exercises revealed a number of water balance-related differences (reference evaporation, soil evaporation, runoff and drainage) between the models. These same differences have been shown to have explained both the APE and RMSE differences between the models for the water-stressed experiment, while for the irrigated experiment the differences were explained by minimal water stress events arising from differences in the irrigation regimes in the two models. The DSSAT CSM Canegro's higher RMSE value reflects somewhat more variability in simulations relative those of SASRI Canegro. This is explained by the fact that the calculation of reference evaporation in the DSSAT CSM is based on a single (maximum) daily relative humidity value, rather than a pair (maximum and minimum) of values as in SASRI Canegro (discussed in Section 4.8).

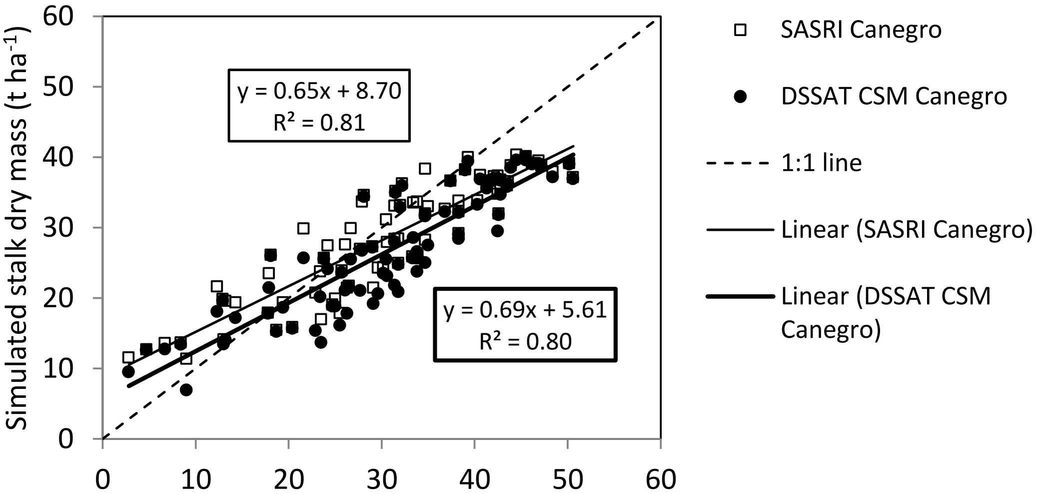


Figure 8.21 Scatter plot of stalk dry mass (t ha^{-1}) values, observed and simulated by DSSAT CSM Canegro and SASRI Canegro, for the A/Growth/HR and A/Growth/07 trials combined.

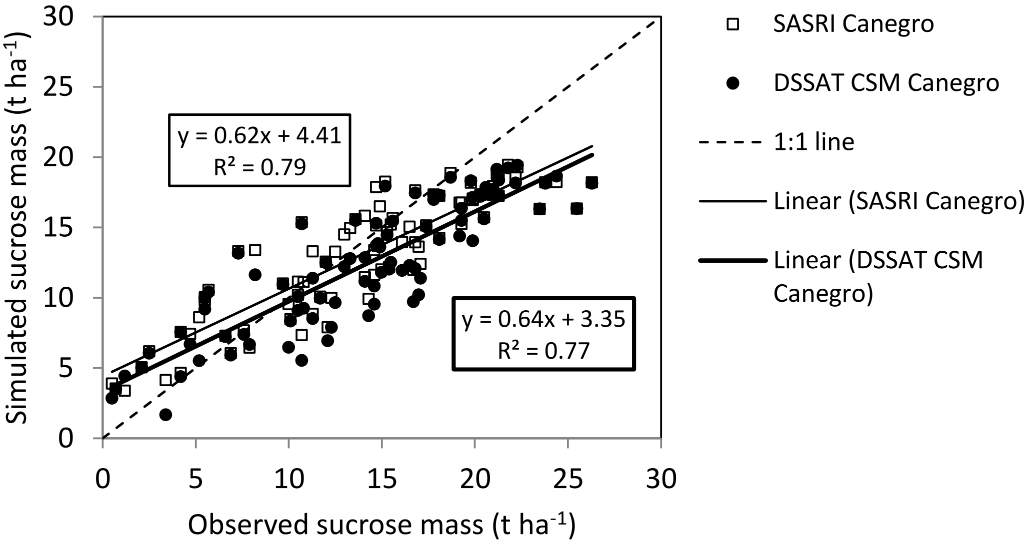


Figure 8.22 Scatter plot of sucrose mass (t ha^{-1}) values, observed and simulated by DSSAT CSM Canegro and SASRI Canegro, for the A/Growth/HR and A/Growth/07 trials combined.

The scatter plot of stalk dry mass (Figure 8.21) shows, subjectively, a close match between SASRI Canegro and DSSAT CSM Canegro performance for the combined data of the two experiments. The DSSAT CSM Canegro model has a slope of linear regression statistic that is closer to 1 (0.69 vs 0.65), and an intercept value closer to 0 (5.61 vs 8.70 t ha^{-1}). Both models under-estimated low values and over-estimated high values. The R^2 value for SASRI Canegro is 0.81, and 0.80 for DSSAT CSM Canegro.

The scatter plot of sucrose mass (Figure 8.22) shows more scatter (more variation) compared with that of stalk dry mass, as evidenced by the lower R^2 values of 0.77 and 0.79 respectively for DSSAT CSM Canegro and SASRI Canegro. The slopes of linear regression are very similar, at 0.64 for DSSAT CSM Canegro and 0.62 for SASRI Canegro. The DSSAT CSM model also shows an intercept of linear regression value closer to 0 t ha^{-1} (3.35 t ha^{-1}, compared with 4.41 t ha^{-1} for SASRI Canegro).

Overall, the APE and RMSE values for the DSSAT CSM Canegro model are considered acceptable, by virtue of being similar to the values published by Singels and Bezuidenhout (2002) (RMSE values in brackets in Table 8.3).

The poor simulation of the effects of seasonal (temperature) differences in the fully irrigated experiment warrants further investigation. Modifications to base temperatures for thermal time accumulation for emergence and the start of stalk elongation (Smit, 2010) may go some way to improving the DSSAT CSM Canegro's seasonal responses. This is expected to improve the model's performance for non-irrigated conditions as well. Concepts identified in recent research into possible causes of low growth rates in sugarcane crops started in spring (compared with those started in autumn), the combination of factors including high-sucrose feedback inhibition impacts on photosynthesis, temperature-dependent maintenance respiration, specific leaf nitrogen content, flowering and lodging (van Heerden *et al.*, 2010), may be further investigated using the DSSAT CSM Canegro model, and may in turn lead to future improvements to the model.

The discrepancies in reference evaporation (*EO*) should be investigated further for rainfed conditions, because *EO* was estimated to be 12.5% higher by DSSAT CSM Canegro than SASRI Canegro for the A/Growth/07 experiment. Simulating a lysimeter experiment (such as described by Thompson (1988)), for example, may be possible avenue for investigating this. Of slight concern was SASRI Canegro's over-estimation of green leaf area index and aerial dry mass under water-stressed conditions, necessary to achieve the realistic stalk dry mass yields estimated. Improvements to the simulated timing of plant phenology and ensuring that the DSSAT CSM Canegro model simulates *EO* accurately under a wide range of circumstances will be necessary to ensure that model calibration is physiologically realistic and meaningful. Further calibration would ideally include a fully-irrigated trial dataset that includes aerial dry biomass and leaf area index for different crop starts, to ensure that timing of phenology, photosynthetically-active radiation interception and biomass accumulation are as estimated as accurately as possible.

8.5 Summary

This validation exercise provided measures of DSSAT CSM Canegro model performance by analysing comparisons of simulated values with observed values for two experiments. It also provided an additional verification that the DSSAT CSM Canegro and the SASRI Canegro models simulate similarly.

In general, the DSSAT CSM Canegro model was found to estimate lower dry mass values than SASRI Canegro. Average prediction errors (APE) were lower for SASRI Canegro than DSSAT CSM Canegro. The similarity in slopes and R^2 values of linear regressions of simulated vs observed stalk dry mass values, and RMSE values, suggests that the models simulate similarly.

RMSE values for the modelled vs measured regressions were marginally greater for DSSAT CSM Canegro than SASRI Canegro, but sufficiently similar to each other and published values for SASRI Canegro, that the performance of the DSSAT CSM Canegro model is considered very satisfactory in this respect. DSSAT CSM Canegro RMSE values for the combined experimental data were 6.62 t ha^{-1} for stalk dry mass and 3.59 t ha^{-1} for sucrose mass. Both models tended to over-simulate low values of dry mass variables and under-simulate high values.

The slopes and intercepts of linear regressions of DSSAT CSM Canegro-simulated and observed values were good. Simulations of stalk dry mass correlated with observations with a slope of 0.69 and an intercept of 5.61 t ha^{-1}, with an R^2 value of 0.80. Sucrose mass correlations showed a slope of 0.64 and an intercept of 3.35 t ha^{-1}, with an R^2 of 0.77. The DSSAT CSM Canegro model showed similar simulation accuracy compared with the SASRI Canegro model, by these measures.

The RMSE and APE differences between the models are attributed mainly to differences in the calculations of reference evaporation and the soil water balance between the SASRI Canegro and DSSAT CSM systems, and partly to (small) corrections to program logic and changes to the tiller senescence algorithm. The validation study confirmed that the plant growth and development components of the SASRI Canegro model operate identically to their counterparts in the DSSAT CSM Canegro model, but that differences in non plant-related processes, with which the plant processes interact, result in differences in simulation outputs under water-stressed conditions.

The model's ability to capture seasonal differences needs further investigation. Recent research into the influences of base temperature, and a possible high sucrose photosynthesis inhibition feedback mechanism, may lead to model improvements that address these weaknesses in seasonal simulations. Further investigation of and exploration into the calculation of reference evaporation in DSSAT CSM Canegro is also recommended for water-stressed conditions.

Overall, this validation points to two satisfactory conclusions. Firstly, that the model adequately simulates the growth and development of sugarcane. Secondly, that the DSSAT Canegro CSM gives very similar simulation performance compared with the SASRI Canegro model, confirming that the models behave nearly identically and achieving the main objective of this study.

9. SUMMARY, FUTURE WORK, AND CONCLUSIONS

9.1 Summary

The objective of this study was to incorporate the SASRI Canegro model into the DSSAT Cropping System Model (CSM) framework. The DSSAT CSM Canegro model was to be fully modular in order to enjoy the benefits the CSM had to offer. It also needed to simulate near-identically to the SASRI Canegro model.

Plant growth and development aspects of the SASRI Canegro model source code were modularised and incorporated as a plant module into the DSSAT CSM. This plant module was verified to simulate identically to the SASRI Canegro model, when non-plant components of the models were disabled. Differences in the water balances in the two models introduced some small simulation discrepancies between the models, with these discrepancies most pronounced under water-stressed conditions. A sensitivity analysis revealed that these discrepancies could be attributed mainly to differences in the calculation of reference evaporation and soil surface evaporation, and to a smaller extent, precipitation runoff and drainage. Both the DSSAT CSM Canegro and SASRI Canegro models were validated against data from 16 crops in two experiments. This validation exercise confirmed that the two models simulate very similarly. It also showed that simulation performance of the DSSAT CSM Canegro model is good. Its root mean squared errors (RMSEs) of 6.62 t ha^{-1} for stalk dry mass and 3.59 t ha^{-1} for sucrose mass, are similar too, but slightly larger than, published values for SASRI Canegro.

An accurate, accessible, up-to-date version of the Canegro model available in the DSSAT system is beneficial to several audiences. As a research platform at SASRI, Canegro can now be run independently by crop scientists without the assistance of programmers. The DSSAT system's large research and development group, user base, consistent compatibility with utility software, and extensive documentation, mean that training, support and assistance are readily available. Comprehensive user and scientific documentation specific to the DSSAT CSM Canegro model has been produced and is freely available on the internet.

The modularisation of the SASRI Canegro code provided the opportunity to improve the structure, clarity and maintainability of the model's source code. Ongoing model development and maintenance will be (and has been) greatly facilitated by this. The modular structure means that the Canegro model has access to a richer set of functionality and features afforded by the CSM, such as a nitrogen-balance, surface mulch model, extensive irrigation options, weather modification and generation functions, crop rotations and sequences, and support for multiple-treatment runs. These are beneficial both for allowing a wide range of model applications with this version of the DSSAT CSM Canegro model, and for supporting future model development and functionality.

Two hypotheses were postulated. The first, that the shared pedigree of the Canegro and some DSSAT CSM crop models (CERES-Maize, for example) would make it possible to incorporate the Canegro model into the DSSAT CSM, was demonstrated to be true. The second hypothesis was that the Canegro model in the DSSAT CSM would simulate near-identically to the standalone SASRI Canegro model, and that any discrepancies between the simulation output of the models could be explained. The verification exercises revealed that the aspects of the two models that are common – algorithms representing the plant growth and development processes – do indeed operate identically. Verification also revealed some discrepancies under water-stressed conditions, and these were primarily attributed to differences in the calculation of reference evaporation and soil surface evaporation. Both the verification and validation exercises showed that the models still produce very similar simulation results, even under stressed conditions. It is therefore considered that this second hypothesis also holds.

9.2 Future work

The DSSAT CSM Canegro provides an ideal platform for future model development. A nitrogen module for the model has already been developed by colleagues at SASRI, which builds on the DSSAT CSM N-balance foundations (van der Laan *et al.*, 2010). The DSSAT CSM Canegro model has been validated for sugarcane crops internationally (Singels *et al.*, 2010a), and is being used in the Agricultural Model Inter-comparison Project (Rosenzweig *et al.*, 2012), which aims to assess the impacts of future climate scenarios on food production for the period 2080-2100 and whose outcomes will be included in the 5th Assessment Report of the International Panel on Climate Change. At SASRI, model development work is underway to make the model suitable for predicting complex genotype by environment interactions from simple genetic traits, with the eventual ambition of being used for model-assisted breeding (Jones *et al.*, 2011).

9.3 Conclusions

The SASRI Canegro model has been successfully incorporated into the DSSAT Cropping System model. The model simulates near-identically to the SASRI Canegro model. The new DSSAT CSM Canegro is an easy-to-use, accessible, accurate and well-documented sugarcane model for a wide range of applications. It also provides an ideal platform for continued model development and improvement. The hypotheses initially postulated have been shown to hold, and the objectives of the project have been achieved.

10. REFERENCES

Allen, RG, Pereira, LS, Raes, D and Smith, M. 1998. Crop evapotranspiration: guidelines for computing crop water requirements. In: *FAO Irrigation and Drainage Paper 56.* Food and Agriculture Organisation of the United Nations, Rome, Italy.

Beebe, NHF. 2001. Using C and C++ with Fortran. Internet. Department of Mathematics - University of Utah, USA. Available from: http://www.math.utah.edu/software/c-with-fortran.html. [Accessed: 15 June 2010].

Boote, KJ, Jones, JW, Hoogenboom, G, Wilkerson, GG and Jagtap, SS. 1987. PNUTGRO V1.0, Peanut crop growth simulation model, user's guide. *Florida Agricultural Experiment Station Journal* 8420. University of Florida, Gainesville, Florida, USA.

Bouman, BAM, van Keulen, H, van Laar, HH and Rabbinge, R. 1993. The 'School of de Wit' crop growth simulation models: A pedigree and historical overview. *Agricultural Systems* 52(2/3):171-198.

Campbell, GS and Norman, JM. 1998. *Introduction to Environmental Biophysics*. Springer Science+Business Media, Inc., New York, USA.

Cheeroo-Nayamuth, BF and Nayamuth, ARH. 2001. Preliminary testing of the Canegro-DSSAT v3.5 model in Mauritius. In: eds. O'Leary, GJ and Kiker, GA, *Proceedings of the First International Workshop on the Canegro Sugarcane Model.* South African Sugar Association Experiment Station, Mount Edgecombe, South Africa.

Cheeroo-Nayamuth, FC, Robertson, MJ, Wegener, MK and Nayamuth, ARH. 2000. Using a simulation model to assess potential and attainable sugar cane yield in Mauritius. *Field Crops Res*earch 66:225-243.

Clark, DJ. 2011. Personal communication, University of KwaZulu-Natal, Pietermaritzburg, South Africa. December 2011.

Hewlett-Packard. 2003. Visual Fortran product description. Internet. Hewlett-Packard Corporation, Palo Alto, CA, USA. Available from: http://h21007.www2.hp.com/portal/site/dspp/menuitem.863c3e4cbcdc3f3515b49c108973a801/?ciid=eb08d7c682f02110d7c682f02110275d6e10RCRD. [Accessed 15 June 2010].

Hoogenboom, G, White, JW, Jones, JW, Boote, KJ. 1994. BEANGRO, a process-oriented dry bean model with a versatile user interface. *Agronomy Journal* 86:182-190.

Hoogenboom, G, Wilkens, PW, Thornton, PK, Jones, JW, Hunt, LA and Imamura, DT. 1999. Decision support system for agrotechnology transfer v3.5. In: eds. Hoogenboom, G, Wilkens, PW and Tsuji, GY, *DSSAT version 3*, vol. 4, 1-36. University of Hawaii, Honolulu, HI, USA.

Hoogenboom, G, Jones, JW, Wilkens, PW, Porter, CH, Batchelor, WD, Hunt, LA, Boote, KJ, Singh, U, Uryasev, O, Bowen, WT, Gijsman, AJ, du Toit, A, White, JW and Tsuji, GY. 2004. *Decision Support System for Agrotechnology Transfer Version 4.0 [CD-ROM]*. University of Hawaii, Honolulu, HI, USA.

Hunt, LA and Boote, KJ. 1998. Data for model operation, calibration and evaluation. In: eds. Tsuji, GY, Hoogenboom, G and Thornton, PK, *Understanding Options for Agricultural Production*, 9-39. Kluwer Academic Publishers, Dordrecht, The Netherlands.

IBM. 1975. *Continuous System Modeling Program III (CSMP III), Program Reference Manual*. IBM Technical Publishing Department, White Plains, USA.

IBSNAT 1989. *Decision Support System for Agrotechnology Transfer, version 2.1. User Guide*. Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI, USA.

IBSNAT, 1993. *The IBSNAT Decade*. Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI, USA.

ICSM. 2006a. International Consortium for Sugarcane Modelling. Internet. International Consortium for Sugarcane Modelling, South African Sugarcane

Research Institute, Mount Edgecombe, South Africa. Available from: http://sasri.sasa.org.za/misc/icsm.html. [Accessed: 20 November 2009].

ICSM. 2006b. Minutes of the inaugural meeting. Internet. International Consortium for Sugarcane Modelling, South African Sugarcane Research Institute, Mount Edgecombe, South Africa. Available from: http://sasri.sasa.org.za/misc/documents/ICSM_meeting_22May2006_minutes.doc. [Accessed: 20 November 2009].

Inman-Bamber, NG. 1991. A growth model for sugar-cane based on a simple carbon balance and the CERES-Maize water balance. *South African Journal of Plant and Soil* 8(2):93-99.

Inman-Bamber, NG. 1994. Temperature and seasonal effects on canopy development and light interception of sugarcane. *Field Crops Research* 36:41-51.

Inman-Bamber, NG and Thompson, GD. 1989. Models of dry matter accumulation by sugarcane. *Proceedings of the South African Sugar Technologists Association* 63:212-216.

Inman-Bamber, NG and Kiker, GA. 1997. CANEGRO 3.10. *DSSAT version 3.1 1998 distribution software*. IBSNAT, University of Hawaii, Honolulu, HI, USA.

Inman-Bamber, NG, Singels, A and Muchow, RC. 1998. A systems approach to benchmarking for sugarcane production in Australia and South Africa. *Proceedings of the South African Sugar Technologists Association* 72:3-9.

Inman-Bamber, NG. 2001. History of the Canegro Model. In: eds. O'Leary, GJ and Kiker, GA, *Proceedings of the First International Workshop on the Canegro Sugarcane Model*. South African Sugar Association Experiment Station, Mount Edgecombe, South Africa.

Intel. 2009. Intel Visual Fortran Compiler Professional Edition 11.1 for Windows. Internet. Intel Corporation, Santa Clara, CA, USA. Available from: http://software.intel.com/sites/products/collateral/hpc/compilers/fwin_indepth.pdf. [Accessed: 31 March 2010].

Jones, CA and Kiniry, JR (eds.). 1986. *CERES-Maize, a Simulation Model of Maize Growth and Development*, Texas A&M University Press, College Station, TX, USA.

Jones, JW and Luyten, JC. 1998. Simulation of biological processes. In: eds. Peart, RM and Curry, RB, *Agricultural Systems Modeling and Simulation*, 19-61. Marcel Dekker, Inc., New York, USA.

Jones, JW, Tsuji, GY, Hoogenboom, G, Hunt, LA, Thornton, PK, Wilkens, PW, Imamura, DT, Bowen, WT, Singh, U. 1998. Decision support system for agrotechnology transfer. In: eds. Tsuji, GY, Hoogenboom, G and Thornton, PK, *Understanding Options for Agricultural Production*, 157–177. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Jones, JW, Keating, BA and Porter, CH. 2001. Approaches to modular model development. *Agricultural Systems* 70:421–443.

Jones, JW, Hoogenboom, G, Porter, CH, Boote, KJ, Batchelor, WD, Hunt, LA, Wilkens, PW, Singh, U, Gijsman, AJ and Ritchie, JT. 2003. The DSSAT cropping system model. *European Journal of Agronomy* 18:235-265.

Jones, JW. 2006. Personal communication, University of Florida, FL, USA. October 2006.

Jones, JW. 2007. Personal communication, University of Florida, FL, USA. May 2007.

Jones, MR, Singels, A, van den Berg, M and Wienese, A. 2006. Evaluating production strategies for sugarcane biomass using a crop model. *ISSCT Agronomy Workshop*, Khon Kaen, Thailand.

Jones, MR, Singels, A and van den Berg, M. 2008. *DSSAT v4.5-Canegro Sugarcane Plant Module User Documentation*. International Consortium for Sugarcane Modelling, South African Sugarcane Research Institute, Mount Edgecombe, South Africa.

Jones, MR, Singels, A and Inman-Bamber, NG. 2011. Simulating source and sink control of structural growth and development and sugar accumulation in sugarcane. *Proceedings of the South African Sugar Technologists Association* 84:157-163.

Keating, BA, Carberry, PS, Hammer, GL, Probert, ME, Robertson, MJ, Holzworth, D, Huth, NI, Hargreaves, JNG, Meinke, H, Hochman, Z, McLean, G, Verburg, K, Snow, V, Dimes, JP, Silburn, M, Wang, E, Brown, S, Bristow, KL, Asseng

S, Chapman S, McCown, RL, Freebairn, DM and Smith, CJ. 2003. An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy* 18(3-4):267-288.

Kiker, GA, Inman-Bamber, NG, Hoogenboom, G and McGlinchley, MG. 2001. Further progress in the validation of the Canegro-DSSAT model. In: eds. O'Leary, GJ and Kiker, GA, *Proceedings of the First International Workshop on the Canegro Sugarcane Model.* South African Sugar Association Experiment Station, Mount Edgecombe, South Africa.

Kiker, GA, Clark, DJ, Martinez, CJ and Schulze, RE. 2006. A Java-based, object-oriented modeling system for Southern African hydrology. *Transactions of the ASABE* 49(5):1419-1433.

Kropff, MJ, van Laar, HH, Matthews, RB (eds.). 1994. *ORYZA1: An Ecophysiological Model for Irrigated Rice Production*. SARP Research Proceedings. International Rice Research Institute, Manila, Philippines.

Marin, F, Romon, V, Assad, ED, Pelegrino, GQ and Lopes-Assad, ML. 2009. Impact of climate change scenarios on sugarcane yields in São Paulo State, Brazil. *ISCCT Agronomy Workshop*. Uberlandia, Brazil, p.37.

Martine, JF and Lebret, P. 2001. Modelling the water content of the sugarcane stalk. *Proceedings of the South African Sugar Technologists Association* 75:211-214.

McCown, RL, Hammer, GL, Hargreaves, JNG, Holzworth, DP and Freebairn, DM. 1996. APSIM: a Novel Software System for Model Development, Model Testing and Simulation in Agricultural Systems Research. *Agricultural Systems* 50:255-271.

McGlinchey, MG and Inman-Bamber, NG. 1996. Predicting sugarcane water use with the Penman-Monteith equation. In: eds. Camp, CR, Sadler, EJ and Yoder, RE, *Evapotranspiration and Irrigation Scheduling. Proceedings of the International Conference*, 592-598.

McGlinchey, MG. 1999. Computer crop model applications: developments in Swaziland. *Proceedings of the South African Sugar Technologists Association* 73:35-38.

McGlinchey, MG. 2001. Validation of the Canegro evaporation estimate. In: eds. O'Leary, GJ and Kiker, GA, *Proceedings of the First International Workshop on the Canegro Sugarcane Model*, 15-16. South African Sugar Association Experiment Station, Mount Edgecombe, South Africa.

Monteith, JL. 1965. Evaporation and the environment. In: ed. Fogg, GE, *The State and Movement of Water in Living Organisms*, 205-234. Academic Press, New York, USA.

O'Leary, GJ, Schumann, AW, Kiker, GA and Bezuidenhout, CN. 2001. Adding a soil and plant nitrogen submodel to the Canegro sugarcane simulation model. In: eds. O'Leary, GJ and Kiker, GA, *Proceedings of the First International Workshop on the Canegro Sugarcane Model*, 24-25. South African Sugar Association Experiment Station, Mount Edgecombe, South Africa.

O'Leary, GJ and Kiker, GA (eds.). 2001. *Proceedings of the First International Workshop on the Canegro Sugarcane Model*. South African Sugar Association Experiment Station, Mount Edgecombe, South Africa.

O'Leary, GJ. 2000. A review of three sugarcane simulation models with respect to their prediction of sucrose yield. *Field Crops Research* 68:97-111.

Olivier, FC and Singels, A. 2012. The effect of crop residue layers on evapotranspiration, growth and yield of irrigated sugarcane. *Water SA* 38:77-87

Pickering, NB, Hansen, JW, Jones, JW, Wells, CM, Chan, VK and Godwin, DC. 1994. WeatherMan: A utility for managing and generating daily weather data. *Agronomy Journal* 86:332-337.

Porter, CH, Jones, JW and Braga, R. 2000. An approach for modular crop model development. Internet. International Consortium for Agricultural Systems Applications, 2440 Campus Rd., 527 Honolulu, HI, USA. Available from http://www.icasanet.org/modular/index.html. [Accessed: 15 December 2009].

Porter, CH. 2006. Personal communication, University of Florida, FL, USA. October 2006.

Porter, CH. 2007. Personal communication, University of Florida, FL, USA. August 2007.

Porter, CH. 2010. Personal communication, University of Florida, FL, USA. July 2010.

Priestley, CHB and Taylor, RJ. 1972. On the assessment of surface heat flux and evaporation using large scale parameters. *Monthly Weather Review* 100:81-72.

Reynolds, JF and Acock B. 1997. Modularity and genericness in plant and ecosystem models. *Ecological Modelling* 94:7-16.

Ritchie, JT. 1972a. Model for predicting evaporation from a row crop with incomplete cover. *Water Resources Research* 8:1204-1213.

Rosenzweig, C, Jones, JW, Hatfield, JL, Ruane, AC, Boote, KJ, Thorburn, P, Antle, JM, Nelson, GC, Porter, C, Janssen, S, Asseng, S, Basso, B, Ewert, F, Wallach, D, Baigorria, G, Winter, JM. (2012) The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology* 170: 166-182.

Rostron, H. 1972b. Some effects of environment, age and growth regulating compounds on the growth, yield and quality of sugarcane in South Africa. Unpublished M.Sc. thesis, Leeds University, Leeds, UK.

Schmidt, EJ, Smithers, JC, Schulze, RE and Mathews, P. 1998. *Impacts of sugarcane production and changing land use on catchment hydrology*. Report to the Water Research Commission on the project 'Water Quality and Quantity Assessment in Catchments with Changing Land Use in the Umzinto Coastal Area' by the South African Sugar Association Experiment Station. WRC Project No. 419/1/98, Water Research Commission, Pretoria, South Africa.

Schulze, RE. 1989. *ACRU: Background, concepts, and theory*. Report No. 154/1/89. Water Research Commission, Pretoria, South Africa.

Schulze, RE. 1995. *Hydrology and agrohydrology: A text to accompany the ACRU 3.00 agrohydrological modelling system*. Report No. TT69/95. Water Research Commission, Pretoria, South Africa.

Singels, A and Bezuidenhout, CN. 2002. A new method of simulating dry matter partitioning in the Canegro sugarcane model. *Field Crops Research* 78:151–164.

Singels, A, Donaldson, RA and Smit, MA. 2005. Improving biomass production and partitioning in sugarcane: theory and practice. *Field Crops Research* 92:291-303.

Singels, A and van den Berg, M. 2006. Personal communication. South African Sugarcane Research Institute, Mount Edgecombe, South Africa. April 2006.

Singels, A, Jones, MR and van den Berg, M. 2008. *DSSAT v4.5-Canegro Sugarcane Plant Module Scientific Documentation*. International Consortium for Sugarcane Modelling, South African Sugarcane Research Institute, Mount Edgecombe, South Africa.

Singels, A, Jones, MR, Porter, C, Smit, MA, Kingston, G, Marin, F, Chinorumba, S, Jintrawet, A, Suguitani, C, van den Berg, M, and Saville, G. 2010a. The DSSAT 4.5 Canegro model: a useful decision support tool for research and management of sugarcane. *Proceedings of the International Society of Sugar Cane Technologists Congress*, 27:438-443.

Singels, A, van den Berg, M, Smit, MA and Jones, MR. 2010b. Modelling water uptake, growth and sucrose accumulation of sugarcane subjected to water stress. *Field Crops Research* 117:59-69.

Singh, G, Chapman, SC, Jackson, PA, Lawn, RJ, Hogarth, DM. 1999. Yield accumulation in sugarcane under wet tropical conditions – effect of lodging and crop age. *Proceedings of the 1999 Conference of the Australian Society of Sugar Cane Technologists*, 241-245. PK Editorial Services, Brisbane, Australia, Townsville, Queensland, Australia.

Soil Conservation Service (SCS). 1972. *Hydrology Section 4*. In: National Engineering Handbook, Chapters 4-10. USA.

Smit, MA. 2010. Characterising the factors that affect germination and emergence in sugarcane. *Proceedings of the South African Sugar Technologists Association* 63:230-234.

Thompson, GD. 1988. Consumptive use of water by a plant crops of sugarcane of variety N14. *Proceedings of the South African Sugar Technologists Association* 62:134-139.

van Antwerpen, R, Inman-Bamber, NG and Meyer, JH. 1994. Cane root development and crop modelling. *Proceedings of the South African Society of Crop Production Congress*.

van Antwerpen, R. 1998. Modelling root growth and water uptake of sugarcane cultivar NCo376. Unpublished PhD thesis, University of the Orange Free State, Bloemfontein, South Africa.

van den Berg, M. 2005. Personal communication. South African Sugarcane Research Institute, Mount Edgecombe, South Africa. May 2005.

van der Laan, M, Miles, N, Annandale, JG, du Preez, CC. 2010. Identification of opportunities for improved nitrogen management in sugarcane cropping systems using the newly developed Canegro-N model. *Nutrient Cycling in Agroecosystems* 90(3):391-404.

van Heerden, PDR, Donaldson, RA, Watt, DA and Singels, A. 2010. Biomass accumulation in sugarcane - unravelling the factors underpinning reduced growth phenomena. *Journal of Experimental Botany* 61:2877-2887.

van Kraalingen, DWG, Rappoldt, C and van Laar, HH. 2003. The Fortran simulation translator, a simulation language. *European Journal of Agronomy* 18:359-361.

van Laar, HH, Goudriaan, J and van Keulen, H (eds.). 1997. *SUCROS97: Simulation of Crop Growth for Potential and Water-limited Production Situations*. Quantitative Approaches in Systems Analysis No. 14. C.T. de Wit Graduate School for Production Ecology, Wageningen, The Netherlands.

Villegas, FD, Daza, OH, Jones, JW and Royce, SF. 2005. CASUPRO: An Industry-driven sugarcane model. Paper number 053025, *2005 ASAE Annual Meeting*.

Wilkerson, GG, Jones, JW, Boote, KJ, Ingram, KT, Mishoe, JW. 1983. Modeling soybean growth for crop management. *Transactions of the American Society of Agricultural Engineers* 26:63-73.

11. APPENDIX 1

Table 11.1 lists the parameter values used for all verification and validation runs.

Table 11.1 Genetic parameter names and values used for all verification and validation runs.

Parameter name	**Description**	**Units**	**Parameter value**
APFMX	Maximum fraction of biomass allocated to above-ground parts of the plant	$t\ t^{-1}$	0.88
DELTTMAX	Maximum change in sucrose content per unit change in the unripened section of the stalk	t^{-1} (derived units)	0.07
FTCON	Sucrose accumulation temperature response shape parameter		0.32
PCB	Extinction coefficient determining the rate at which the fraction of daily dry mass increments allocated to below-ground biomass decreases with total crop dry mass		0.6
SUCA	Maximum stalk sucrose content at the base of the stalk, on a dry mass basis	$g\ g^{-1}$	0.58
SURCON	Sucrose partitioning parameter that determines the response time of shifts in partitioning between sucrose and fibre in the stalk due to environmental changes (0-1)		0.99
TBFT	Temperature at which partitioning of unstressed stalk mass increments to sucrose is 50% of the maximum value	°C	25
AREAMX_CF(1)	1st parameter of polynomial function describing maximum leaf area per leaf rank		0
AREAMX_CF(2)	2nd parameter of polynomial function describing maximum leaf area per leaf rank		27.2
AREAMX_CF(3)	3rd parameter of polynomial function describing maximum leaf area per leaf rank		-20.8
EXTCFN	Maximum canopy PAR extinction coefficient		0.84
EXTCFST	Minimum canopy PAR extinction coefficient		0.58
LFMAX	Maximum number of green leaves per shoot	Leaves/ shoot	12
LFNMXEXT	Threshold above which canopy light extinction coefficient is equal to EXTCFN	Leaves/ shoot	20

LG_AMBASE	Above-ground fresh mass at which lodging starts	t ha^{-1}	220
LG_AMRANGE	Above-ground fresh mass range (above LG_AMBASE) over which lodging occurs	t ha^{-1}	30
LDG_FI_REDUC	Maximum decrease in fractional light interception caused by lodging		0.1
LG_GP_REDUC	Maximum decrease in gross photosynthesis rate caused by lodging		0.28
LMAX_CF(1)	1st parameter of polynomial function describing maximum leaf length per leaf number		-0.376
LMAX_CF(2)	2nd parameter of polynomial function describing maximum leaf length per leaf number		12.2
LMAX_CF(3)	3rd parameter of polynomial function describing maximum leaf length per leaf number		21.8
MAXLFLENGTH	Maximum leaf length	cm	100
MAXLFWIDTH	Maximum leaf width	cm	3.5
MXLFAREA	Maximum leaf area	cm^2	360
MXLFARNO	Leaf number at which maximum area is reached		15
WIDCOR	Parameter affecting width of the leaves		1
WMAX_CF(1)	1st parameter of polynomial function describing maximum leaf width per leaf number		-0.0345
WMAX_CF(2)	2nd parameter of polynomial function describing maximum leaf width per leaf number		2.243
WMAX_CF(3)	3rd parameter of polynomial function describing maximum leaf width per leaf number		7.75
dPERdT	Change in plant extension rate per °C d (at base temperature TBASEPER)	cm °C d^{-1}	0.176
CHTCoeff	Coefficient determining canopy height in relation to stalk height and leaf length		0.864
CHUPIBASE	Thermal time after which stalks start to elongate	°C d	1050
PI1	First phyllocron interval (thermal time between consecutive leaf tip appearance)	°C d	69
PI2	Second phyllocron interval (thermal time between consecutive leaf tip appearance)	°C d	169
PSWITCH	Number of leaves at which phyllocron interval increases from PI1 to PI2	leaves	18
TTBASEEM	Base temperature for emergence	°C d	10
TTBASELFEX	Base temperature for leaf extension	°C d	10
TTBASEPOP	Base temperature for stalk population development	°C	16
TTPLNTEM	Thermal time requirement from planting for	°C d	428

	a plant crop to emerge		
TTRATNEM	Thermal time requirement from harvest for a ratoon crop to emerge	°C d	203
TT_POPGROWTH	Thermal time at which peak tiller population is reached	°C d	600
MAX_POP	Maximum tiller population	1000 shoots ha^{-1}	300
TBASEPER	Base temperature for plant elongation / extension	°C	10.057
POPCF(1)	1st stalk population development coefficient, describing stalk population increase		1.826
POPCF(2)	2nd stalk population development coefficient, describing stalk population increase		-0.00201
POPCF(3)	3rd stalk population development coefficient, describing stalk population increase		866.7
POPDECAY	Fraction of shoot population above the final population that senesce per unit thermal time.		0.004
POPTT16	Stalk population at 1600 °C d	shoots ha^{-1}	133000
RWUMX	Maximum rate of water uptake per unit root length	cm^{3} cm^{-1}	0.07