Evaluating and improving crop growth models for simulating genotype-by-environment interactions in sugarcane

by

Matthew Robert Jones

Submitted in fulfilment of the requirements for the degree

Philosophiae Doctor in Agronomy

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

January 2023

Supervisor:

Abraham Singels, University of Pretoria, South Africa

Co-supervisors:

John Annandale, University of Pretoria, South Africa

Graeme Hammer, University of Queensland, Australia

Declaration

I, Matthew Robert Jones, declare that the thesis/dissertation, which I hereby submit for the degree Philosophiae Doctor in Agronomy at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE:

DATE: 8 February 2023

Acknowledgements

I would like to express my grateful thanks to the following:

- My supervisors Abraham Singels, John Annandale and Graeme Hammer, for their support, assistance, advice, patience and generosity.
- The International Consortium for Sugarcane Modelling, and specific member organisations the South African Sugarcane Research Institute (SASRI), the Zimbabwe Sugar Association Experiment Station (ZSAES), the Florida Sugarcane Growers Co-operative (SGC), and the French Agricultural Research Centre for International Development (CIRAD), for their financial support for this study.
- Geoff Inman-Bamber, Gerrit Hoogenboom, Ken Boote, Philip Jackson, Mathias Christina, Jean-Francois Martine and Christophe Poser for their detailed insights, advice and assistance.
- The experimentalists and technical staff who conducted the field experiments and assembled the datasets used in this study.
- My colleagues at SASRI, particularly Belinda Naidoo, Riekert van Heerden, Marvellous Zhou, Derek Watt and Carolyn Baker, for their support, assistance and encouragement.
- My wife, Sarah, for her patience, support, encouragement and love.

Summary

Sugarcane is a large, perennial, tropical grass which accumulates sucrose (sugar) in stem tissues. Breeding of sugarcane cultivars follows a traditional approach of crossing parent genotypes (Gs) and assessing progeny over multiple selection stages. This is time-consuming, resource intensive and costly.

Crop growth simulation models have the potential to assist sugarcane breeding, by characterising environments (Es) meaningfully to improve selection efficiencies; by identifying breeding targets via dissection of complex traits (such as yield) into simpler component traits; and, by predicting yields in target Es from simple phenotypic trait or genotypic information.

For credible application in breeding, crop model users should have confidence in the models' abilities to predict effects, on important traits such as yield, of differences in Es, Gs and interactions between Es and Gs (GxE). Many general-purpose sugarcane crop models have been developed, but none have been rigorously evaluated for their capabilities to predict E, G and GxE interaction effects.

The broad objective of this study was to evaluate and improve sugarcane crop models for supporting breeding of irrigated sugarcane. Using data from an international multi-environment trial (three Gs at four sites), the suitability of key sugarcane crop model concepts and input parameter definitions was assessed; three leading sugarcane crop models were then critically assessed for their abilities to predict E, G and GxE interaction effects. Findings from these assessments, along with insights gleaned from literature, informed the development of a new sugarcane crop growth model. The value of the new model was demonstrated in a case study exploring the impacts of genotypic adaptations to temperature sensitivity, canopy development and canopy senescence on radiation interception and yield accumulation at each of the four sites.

The assessment of model concepts and input definitions revealed that solar radiation intensity, in addition to temperature, needed to be considered in predicting tillering rates and date of onset of stalk growth. The thermal time model of predicting the duration of the germination phase was found to be inadequate. The three models assessed gave satisfactory performance for predicting E and G effects, but could not predict GxE effects accurately.

The new model developed, CaneGEM, makes use of a relative canopy growth rate concept, links canopy development with biomass accumulation, simulates a transition from tillering to stalk growth in response to radiation interception, and uses this transition to regulate canopy development and biomass partitioning; biomass partitioning of above-ground parts of the crop is determined according to relative sink strengths driven by structural growth, and the accumulation of stalk sugars is a passive consequence of source and sink strengths.

Evaluation of the new model revealed similar performance to the DSSAT-Canegro model in general, and improved abilities to predict GxE interaction effects in radiation interception and biomass yields at harvest. However, this improvement required the specification of dates of emergence, as the germination timing algorithm remains inadequate. Furthermore, date of emergence was found to be a powerful determinant of GxE interactions in biomass yields. The case study revealed that genotypic temperature adaptations can increase biomass yields, especially in cooler Es, but can also present a risk of penalising sucrose yields, particularly for late-season crops.

List of tables

Table 5-1. CaneGEM model input parameters and their values for calibrated for the standard reference cultivar NCo376. Explanations of how values were derived are provided in the text in the following sections.....48

Table 5-2. Rate and state variables used in the description of the CaneGEM model in the sections that follow. 50

Table 5-3. Detail of experimental datasets used for model calibration (CaneGEM model) and validation (CaneGEM and DSSAT-Canegro models), and for assessment of E, G and GxE interaction effects (CaneGEM and DSSAT-Canegro models). “CC” refers to crop class (P = plant crop, R1 = 1st ratoon, R2 = 2nd ratoon, etc). “Task”: C = calibration, V = validation, GE = G / E/ GxE interaction evaluation. “RS” is row-spacing (m). 67

Table 5-4. Combinations of crop growth simulation model, dataset, source of germination phase duration (i.e. date of primary shoot emergence) data, the purpose of the model assessment and corresponding scenario and task codes used in this chapter.....70

Table 5-5. Duration of germination (days from crop start to 50% primary shoot emergence) for crops used for CaneGEM model calibration, validation and assessment of E/G/GxE effects. “Obs. DoGP” indicates date of emergence values recorded by experimentalists, while “Est. DoGP” indicates DoGP estimated from shoot density observations; “Crop” refers to ratoon number, where R0 is a plant crop, R1 is the first ratoon and R2 the second ratoon. 78

Table 5-6. Trait parameter values calibrated for R570, N41 and CP88-1762 necessary for capturing improvements in G, E and GxE interaction effects in seasonal radiation interception, with NCo376 values shown for reference.....79

Table 5-7. CaneGEM model performance statistics, for calibration with NCo376 data and predetermined germination phase duration (scenario code “GEM_376C_P” in the text). N = “number of observations”, “Y-int” is the y intercept of linear regression between observed and simulated values, RMSE is the root mean squared error, and APE is the average prediction error (explained in the text).....81

Table 5-8. CaneGEM and DSSAT-Canegro model performance statistics, for validation with NCo376 data. p = “p-value”; “Y-int”, “Slp.” = y-intercept and slope respectively of linear regression between observed and simulated values; “p” indicates the p-value (statistical strength of linear regression between observed and simulated values); “RMSE” and “APE” = root mean squared error and average prediction error respectively between observed and simulated values. Model assessment scenario codes “GEM_376V_P”, “GEM_376V_S” and “DC_376V_S” are included here for correspondence with the presentation of results and discussion in the text. 86

Table 5-9. DSSAT-Canegro model accuracy for predicting environmental (E), genotypic (G) and genotype-by-environment interaction (GE) effects on seasonal fractional interception of radiation (FIPARa) and aerial dry biomass at harvest (ADMh, t/ha). ICSM IGEP dataset, simulated duration of germination phase (model assessment scenario code “DC_IGEP_S”). “Y-int” = the y-intercept of the linear regression between observed and simulated values; “N” = the number of data points considered. 94

Table 5-10. CaneGEM model accuracy for predicting environmental (E), genotypic (G) and genotype-by-environment interaction (GE) effects on seasonal fractional interception of radiation (FIPARa) and aerial dry biomass at harvest (ADMh, t/ha), for three cultivars (R570, N41 and CP88-1762), for the ICSM IGEP dataset. Durations of germination phase were predetermined (model assessment scenario code "GEM_IGEP_P1"). NCo376 trait parameter values were used. 96

Table 5-11. CaneGEM model (with predetermined duration of germination and G-specific calibration of canopy development parameters, model assessment scenario code "GEM_IGEP_P") accuracy for predicting environmental (E), genotypic (G) and genotype-by-environment interaction (GE) effects on seasonal fractional interception of radiation (FIPARa) and aerial dry biomass at harvest (ADMh, t/ha). 98

Table 5-12. CaneGEM model (with simulated duration of germination phase and G-specific calibration of canopy development parameters, model assessment scenario "GEM_IGEP_S") simulation accuracy for predicting environmental (E), genotypic (G) and genotype-by-environment interaction (GE) effects on seasonal fractional interception of radiation (FIPARa) and aerial dry biomass at harvest (ADMh, t/ha). "N" = number of observations. 99

Table 6-1. Trait parameter values used for adapted genotypes in the case study..116

Table 6-2. Harvest months and year ranges simulated at each site. 117

Table 11-1. CaneGEM model performance statistics, for calibration with NCo376 data and simulated durations of growth phase. N = "number of observations" , "p" = p-value, "Y-int" is the y intercept of linear regression between observed and simulated values, RMSE is the root mean squared error, and APE is the average prediction error (explained in the text). 166

Table 11-2. Mean simulated duration of germination (DoGP, d after crop start), date of onset of stalk growth (doOSG, d after crop start), seasonal photosynthetically-active radiation interception fraction (FIPARa), seasonal apparent radiation use efficiency (RUEa, g/MJ), above-ground dry biomass at harvest (ADMh, t/ha), stalk dry mass at harvest (SDMh, t/ha) and sucrose mass at harvest (SUCh, t/ha), for the baseline cultivar NCo376, and three classes of genetic adaptations: modification of temperature sensitivity thresholds, relative canopy growth rate, and the green leaf area index threshold at which senescence starts. Data are shown for three times of harvest (early-, mid- and late-season) and the average of the three ('All'). 177

List of figures

Figure 1-1. Flowchart illustrating the basic daily calculations made during each day of sugarcane crop growth simulation. Rectangles indicate processes and rounded rectangles are variables. Solid lines indicate the flow of data calculated on the same day, while dashed lines indicate input from values calculated the previous day. All masses are dry masses. Acronyms: Fi is 'fractional interception of radiation', and 'ADM' is 'Above-ground dry biomass'. 5

Figure 5-1. Simplified representation of CaneGEM model canopy development. Note that the calculation of source strength from reserves is not detailed in this diagram. 46

Figure 5-2. Relationship between the FI_OSG (the inter-row fractional radiation interception value at which 50% of shoots have transitioned to stalk growth) and row-spacing. 54

Figure 5-3. Onset of stalk growth fraction (OSGfrac) as a logistic function of fractional interception of photosynthetically-active radiation (PAR). Parameter values are explained in the text; parameter values FI_OSG and $OSGfrac_{c1}$ are 0.70 and 20 respectively in this example. 54

Figure 5-4. Example of fitting values for parameters $OSGfrac_{c1}$ and FI_OSG, used in the model equation for calculating the transition from tillering to stalk growth. The data shown are for cultivar R570 at five irrigated sites, 1.5 m row-spacing. The "Belle Glade" crops included leaf sheath dry mass with stalk dry mass observations, resulting in overestimated values at low fractional radiation interception values. 55

Figure 5-5. Temperature factor relationships with daily mean air temperature for the control of leaf area expansion (FTlai) and photosynthesis (FTphoto). 56

Figure 5-6. Example relationship showing source-unlimited relative growth rate of green leaf area index (RGRlai), as a function of OSGfrac (top), which is itself a function of fractional interception of radiation (FIPAR) (bottom). Parameter names are explained in the text. 58

Figure 5-7. Relationship between the relative fraction of green leaf area index (GLAI, m^2/m^2) senesced each day (RSRlai) and GLAI. Parameter GLAIst is explained in the text. 59

Figure 5-8. Example relationship between canopy extinction coefficient and physiological age expressed as primary shoot leaf number. 60

Figure 5-9. Example (using standard parameter values: $RTPF_{c1}$ = 0.95, $RTPF_{c3}$ = 0.6, $RTPF_{c2}$ = 0.88) of root biomass partitioning fraction as a function of total crop dry mass. 61

Figure 5-10. Relative stalk sugars partitioning capacity ($SKSPF_{pot}$) vs number of leaves appeared after the start of stalk growth. Parameter values are explained in the text. 64

Figure 5-11. Relationship between total stalk sugar mass (SKS, t/ha) and the fraction of sugar that is required to be hexose (SKHfrac, g/g), for two daily fibre synthesis (dSKF, t/ha/d) rates: low (dSKF = 0.015) and high (dSKF = 0.150). 65

Figure 5-12. Polynomial functions (orange solid lines) fitted to shoot density vs thermal time from crop start data (black lines and points) to estimate dates of emergence for NCo376 calibration and validation datasets. The vertical green lines indicate the date

of emergence; blue solid lines the estimated final stalk density; and the dashed red line 50% of the final value. 73

Figure 5-13. Scatter plot of time-series observed and CaneGEM model-simulated data for crop phenology- and biomass-related variables, for the calibration dataset for NCo376. Durations of germination phases were predetermined (scenario code "GEM_376C_P" in the text). 80

Figure 5-14. Scatter plots of age of onset of stalk growth (days since crop start) and seasonal PAR interception, simulated (CaneGEM model) and observed, for the calibration dataset for NCo376. Durations of germination phases were predetermined (scenario code "GEM_376C_P" in the text). 82

Figure 5-15. Scatter plot of time-series data for crop phenology- and biomass-related variables, for the CaneGEM model and the validation dataset for NCo376. Durations of the germination phase were predetermined (model assessment scenario code "GEM_376V_P"). 83

Figure 5-16. Scatter plot of time-series data, showing CaneGEM model-simulated and observed biomass fractions, for the validation dataset for NCo376. Durations of the germination phase were predetermined (model assessment scenario code "GEM_376V_P"). "ADM" is above-ground dry biomass (t/ha), while "Trash" refers to senesced leaf and stalk dry mass. 84

Figure 5-17. CaneGEM model-simulated vs observed duration of tillering phase (labelled as "OSG", i.e. onset of stalk growth, left), and seasonal PAR interception (FIPARa, right), for the NCo376 validation dataset. Durations of the germination phase were predetermined (model assessment scenario code "GEM_376V_P"). The shaded region shows values within one standard error range. 84

Figure 5-18. Scatter plot of simulated (DSSAT-Canegro) and observed time-series data for crop phenology- and biomass-related variables, for the NCo376 validation dataset. Durations of germination phase were simulated by the model (model assessment scenario code "DC_376V_S"). "PAR Frac. Int [%]" is fractional interception of photosynthetically-active radiation. 89

Figure 5-19. Scatter plot of time-series data, showing simulated (DSSAT-Canegro) and observed biomass fractions, for the validation dataset for NCo376. Durations of germination phases were simulated by the model (model assessment scenario code "DC_376V_S"). "ADM" is above-ground dry biomass (t/ha), while "Trash" refers to senesced leaf and stalk dry mass. 90

Figure 5-20. Simulated (DSSAT-Canegro) vs observed duration of tillering phase (DoTP, labelled as "OSG date", left), and seasonal PAR interception (FIPARa, right), for the NCo376 validation dataset. Durations of germination phases were simulated by the model (model assessment scenario code "DC_376V_S"). The shaded region shows values within one standard error range. 90

Figure 5-21. Comparison of R^2 values for observed vs simulated time-series data, simulated by CaneGEM and DSSAT-Canegro for the NCo376 validation dataset, with durations of germination phase simulated by the models (model assessment scenario codes "GEM_376V_S" and "DC_376V_S", explained in the text). 91

Figure 5-22. Simulated (by DSSAT-Canegro, red line) and observed (black points and linearly-interpolated lines) fractional interception of photosynthetically-active radiation

(FIPAR, %), for the non-stressed ICSM IGEP experiments, with simulated duration of germination phase (model assessment scenario code “DC_IGEP_S”). 92

Figure 5-23. Simulated (by DSSAT-Canegro, red line) and observed (black points and linearly-interpolated lines) above-ground dry biomass (t/ha), for the non-stressed ICSM IGEP experiments, with simulated duration of germination phase (model assessment scenario code “DC_IGEP_S”).....93

Figure 5-24. E, G and GxE interaction effects on seasonal photosynthetically-active radiation interception (FIPARa) and above-ground dry biomass at harvest (ADM), simulated (DSSAT-Canegro model) vs observed. ICSM IGEP dataset, simulated duration of germination phase (model assessment scenario code “DC_IGEP_S”). ..94

Figure 5-25. CaneGEM-simulated (red line) and observed (black points and linearly-interpolated lines) FIPAR, with predetermined durations of germination phase and NCo376 cultivar parameters for the ICSM IGEP dataset (model assessment scenario code “GEM_IGEP_P1”). Inter-sample period average daily air temperature is also shown (stepped blue line). Differences in simulated values between Gs are due to G-E differences in predetermined durations of germination phase. 95

Figure 5-26. E, G and GxE interaction effects, simulated vs observed, with predetermined durations of germination phase and NCo376 cultivar parameters for the ICSM IGEP dataset (model assessment scenario code “GEM_IGEP_P1”). Trait parameter values are otherwise identical (NCo376 parameter values used).....96

Figure 5-27. Simulated (CaneGEM model, with predetermined date of emergence and G-specific calibration of canopy development parameters, model assessment scenario code “GEM_IGEP_P”) vs observed crop development and biomass components, for the ICSM IGEP dataset. 97

Figure 5-28. E, G and GxE interaction effects, CaneGEM-simulated vs observed, with predetermined duration of germination and G-specific calibration of canopy development parameters, for the ICSM IGEP dataset (model assessment scenario code “GEM_IGEP_P”). 98

Figure 5-29. E, G and GxE interaction effects, CaneGEM-simulated vs observed, with simulated duration of germination phase and G-specific calibration of cardinal temperatures controlling canopy development and germination rate (model assessment scenario “GEM_IGEP_S”). 99

Figure 5-30. Observed (solid lines) and CaneGEM-simulated (dashed lines) genotype-by-environment (GxE) crossover effects for seasonal intercepted photosynthetically-active radiation fraction (FIPARa, left) and above-ground dry biomass yields at harvest (ADMh, right), for two Gs (R570 and CP88-1762) and two Es (La Mare, Reunion Island; and Belle Glade, Florida, USA).....102

Figure 6-1. Summary of monthly long-term weather data for the four study sites. ‘aTMIN’ and ‘aTMAX’ are average daily minimum and maximum air temperature (°C) repectively; ‘aSRAD’ is average daily global solar radiation (MJ/m^2/d). The vertical green and red bars indicate the start and end months respectively of typical milling seasons at each site. 118

Figure 6-2. Daily average (±1 standard deviation, shaded) mean daily air temperature (C°), photosynthetically-active radiation interception fraction, above-ground dry biomass (t/ha), stalk dry mass (t/ha), and sucrose mass (t/ha) and for Belle Glade, 2002-2020, baseline cultivar (NCo376). 121

Figure 6-3. Percentage change in above-ground dry mass (ADMh), stalk dry mass (SDMh) and sucrose mass (SUCh) at harvest due to trait differences in temperature sensitivity (Tmp), relative canopy expansion rate (Cpy) and senescence thresholds (Sen) for Belle Glade, 2003-2020. 123

Figure 6-4. Daily average (±1 standard deviation, shaded) mean daily air temperature (C°), photosynthetically-active radiation interception fraction, above-ground dry biomass (t/ha), stalk dry mass (t/ha), and sucrose mass (t/ha) and for Chiredzi, 2000-2021, baseline cultivar (NCo376).....125

Figure 6-5. Boxplots of biomass variables for Chiredzi, 2000-2020, expressed as percentage changes from baseline cultivar, for early- (blue), mid- (black) and late-season (red) harvested crops. ADMh = 'Above-ground dry biomass at harvest' (t/ha), SDMh = 'Stalk dry mass at harvest' (t/ha), SUCh = 'Sucrose dry mass at harvest' (t/ha). Genetic adaptations to temperature sensitivity, canopy growth rate and senescence threshold are illustrated. 126

Figure 6-6. Daily average (±1 standard deviation, shaded) mean daily air temperature (C°), photosynthetically-active radiation interception fraction, above-ground dry biomass (t/ha), stalk dry mass (t/ha), and sucrose mass (t/ha) and for La Mare, 1994-2020, baseline cultivar (NCo376).....127

Figure 6-7. Boxplots of biomass variables for La Mare, 1994-2020, expressed as percentage changes from baseline cultivar, for early- (blue), mid- (black) and late-season (red) harvested crops. ADM = 'Above-ground dry biomass at harvest' (t/ha), SDM = 'Stalk dry mass at harvest' (t/ha), SUC = 'Sucrose dry mass at harvest' (t/ha). Genetic adaptations to temperature sensitivity, canopy growth rate and senescence threshold are illustrated. 129

Figure 6-8. Daily average (±1 standard deviation, shaded) mean daily air temperature (C°), photosynthetically-active radiation interception fraction, above-ground dry biomass (t/ha), stalk dry mass (t/ha), and sucrose mass (t/ha) and for Pongola, 1998-2020, baseline cultivar (NCo376).....131

Figure 6-9. Daily mean photosynthetically-active radiation interception for the baseline cultivar (NCo376) and genotypes adapted to lower ("Tmp -2 C", blue, short-dashed series) and higher ("Tmp +2 C", red, long-dashed series) temperatures, at Pongola, South Africa, 1998-2020. Shading indicates 1 standard deviation above and below the mean.....132

Figure 6-10. Boxplots of biomass variables for Pongola, 1998-2020, expressed as percentage changes from baseline cultivar, for early- (blue), mid- (black) and late-season (red) harvested crops. ADMh = 'Above-ground dry biomass at harvest' (t/ha), SDMh = 'Stalk dry mass at harvest' (t/ha), SUCh = 'Sucrose dry mass at harvest' (t/ha). Genetic adaptations to temperature sensitivity, canopy growth rate and senescence threshold are illustrated 133

Figure 10-1. Coefficient of variation (%) of thermal time from crop start to emergence (TTem_X) for ICSM cultivars and sites (Belle Glade, Chiredzi, La Mare and Pongola), at different base temperatures. Soil temperatures were simulated by DSSAT-Canegro, while air temperatures were recorded.....164

Figure 11-1. Scatter plot of time-series data for crop phenology- and biomass-related variables, for the CaneGEM model, calibration dataset for NCo376. Durations of germination phases were simulated. 165

Figure 11-2. Scatter plots of age of onset of stalk growth (days since crop start) and seasonal PAR interception, simulated (CaneGEM model) and observed, for the calibration dataset for NCo376. Durations of germination phases were simulated. 166

Figure 11-3. Scatter plot of time-series data for crop phenology- and biomass-related variables, for the CaneGEM model and the validation dataset for NCo376**,** simulated durations of germination phase (model assessment scenario code “GEM_376V_S”). 167

Figure 11-4. Scatter plot of time-series data, showing simulated (CaneGEM model) and observed biomass fractions, for the validation dataset for NCo376**,** simulated duration of germination phase (model assessment scenario code “GEM_376V_S”). “ADM” is above-ground dry biomass (t/ha), while “Trash” refers to senesced leaf and stalk dry mass. 168

Figure 11-5. Simulated vs observed duration of tillering phase (labelled as date of onset of stalk growth, “OSG date”, left), and seasonal PAR interception (FIPARa, right), for the NCo376 validation dataset**,** simulated duration of germination phase (model assessment scenario code “GEM_376V_S”). The shaded region shows values within one standard error range. 168

Figure 11-6. CaneGEM-simulated (red line) and observed (black points and linearly-interpolated lines) FIPAR, with predetermined duration of growth phase and G-specific calibration of canopy development parameters (model assessment scenario “GEM_IGEP_P”). Inter-sample period average daily air temperature is also shown (stepped blue line). 169

Figure 11-7. Simulated (CaneGEM model, with simulated emergence dates and G-specific calibration) vs observed crop development and biomass components, for the ICSM IGEP dataset (model assessment scenario “GEM_IGEP_S”). 170

Figure 11-8. Simulated (CaneGEM model, with simulated duration of germination phase and G-specific calibration of canopy development parameters, model assessment scenario code “GEM_IGEP_P”) vs observed crop development and biomass components, for the ICSM IGEP dataset. 171

Figure 11-9. Daily weather data for Belle Glade, Florida. Observed data (“Obs”, black points) were replaced with MERRA synthesized data (red points) where observed data were missing or out of expected ranges. 172

Figure 11-10. Solar radiation data for Chiredzi, Zimbabwe, 2000-2020, observed, estimated from sunshine hours, and estimated by NASA Power (MERRA). 173

Figure 11-11. Daily weather data for Chiredzi, Zimbabwe. Observed data (“Obs”, black points) were replaced with MERRA synthesized data (red points) or bias-adjusted MERRA data (“MERRA t.”, orange points) where observed data were missing or out of expected ranges. 174

Figure 11-12. Daily weather data for La Mare, Reunion Island, France.....175

Figure 11-13. Daily weather data for Pongola, South Africa. 176

Figure 11-14. Daily fractional interception of photosynthetically-active radiation, for early-season (E), mid-season (M) and late-season (L) crops at Belle Glade, 2003-2020. Impacts of frost events, which reduce green leaf area index, are evident, e.g. 2011.....182

TABLE OF CONTENTS

1. Introduction....3
1.1 Introduction....3
1.2 Problem statement....6
1.3 Research questions and hypotheses....6
1.4 Objectives....7
1.5 Thesis structure....7
2. Literature review....9
2.1 Introduction....10
2.2 Sugarcane....10
2.3 Crop improvement and G, E and GxE interaction effects....14
2.4 Applications of crop modelling to support crop improvement....17
2.5 Sugarcane crop models....25
2.6 Summary and Conclusion....38
3. Exploring process-level genotypic and environmental effects on sugarcane yield using an international experimental dataset....40
4. Evaluating process-based sugarcane models for simulating genotypic and environmental effects observed in an international dataset....41
5. A sugarcane model for improved simulation of GxE interactions of canopy cover and biomass accumulation....42
5.1 Introduction....43
5.2 Model description....44
5.3 Methodology....65
5.4 Results....77
5.5 Discussion....100
5.6 Conclusion....109
6. Model application case study: assessing the impacts of canopy development traits for four environments....111
6.1 Introduction....111
6.2 Methodology....114
6.3 Results....119
6.4 Discussion....133
6.5 Conclusion....137
7. Answers to research questions, and Recommendations....139
7.1 Answers to research questions....139
7.2 Recommendations....142

8. Conclusion 145
9. References 147
10. Appendix A – Investigating the use of simulated soil temperature in driving germination 163
10.1 Methodology 163
10.2 Results 163
10.3 Conclusion 164
11. Appendix B – Supplementary data 164
11.1 Supplementary CaneGEM calibration results 164
11.2 Supplementary CaneGEM validation results 167
11.3 Supplementary CaneGEM results for the ICSM IGEP dataset 169
11.4 Weather data graphs for case study 171
11.5 Results for case study 176
11.6 CaneGEM model source code file structure and operation 182

1. INTRODUCTION

TABLE OF CONTENTS

1.1 Introduction....3
1.2 Problem statement....6
1.3 Research questions and hypotheses....6
1.4 Objectives....7
1.5 Thesis structure....7

1.1 Introduction

Sugarcane (interspecific hybrid of *Saccharum* spp.) is a large, perennial tropical/sub-tropical grass, that accumulates sucrose in stem tissues (Moore et al., 2013). Approximately 26.7 million ha of sugarcane were harvested worldwide in 2019 (FAO, 2021), primarily for the production of sugar and bioethanol. The largest producers are Brazil, India, Thailand and China, with sugarcane forming important parts of the local economies of smaller industries such as South Africa, Zimbabwe, the USA and French overseas territories.

Expected increases in demand for food, biofuels and fibre in future, especially where climate change threatens yields in current production regions, requires increases in the rate of improvement of genetic potential of crop species (Diepenbrock et al., 2021).

Many sugarcane industries worldwide, including those of South Africa, Zimbabwe, USA and France, have well-developed sugarcane breeding programmes (Baucum and Rice, 2009; Dumont et al., 2021; M. Zhou, 2013; Zhou, 2005). Sugarcane is predominantly bred using traditional methods of crossing and selection (Balsalobre et al., 2017; Yadav et al., 2020), taking 10-15 years for the development of a new released cultivar. Sugarcane cultivars are bred to give superior performance for target environments, formally termed “target populations of environments” (TPEs, Hammer & Jordan (Hammer and Jordan, 2007)).

The performance of new sugarcane genotypes (Gs) is assessed in multi-environment trials (METs, (Hammer and Jordan, 2007; Ramburan, 2012)), which are intended to be representative of the corresponding TPEs. Where the ranking of Gs is stable across environments (Es), those Gs are considered broadly-adaptable; conversely, where the relative performance of Gs changes between Es, genotype-by-environment (GxE) interactions are present (de Leon et al., 2016).

The resource-intensive nature of sugarcane breeding necessarily results in the annual release of small numbers of broadly-adaptable cultivars (Ramburan, 2012). This “one size fits all” approach implies potential for genetic improvement, to develop niche-bred cultivars optimised to more narrowly-defined TPEs. Genetic gain in sugarcane is reported to have plateaued over the last 20 years in many countries (Yadav et al., 2020), while genetic yield potentials have improved in other crops (Grassini et al., 2013; Rizzo et al., 2022; Xu et al., 2017; Yadav et al., 2021). The development of

technologies to reduce the costs and time requirements of sugarcane breeding will help to close these potential yield gaps and raise sugarcane productivity overall.

Crop growth simulation modelling is one such technology that has been used in other crops to assist plant breeding, in the following ways (Hammer and Jordan, 2007):

- to characterise environments in TPEs and METs, allowing appropriate selection weighting of Es;
- to dissect complex traits (such as yield), leading to enhanced understanding of the mechanisms by which adaptability to Es is conferred, and thus informing breeding by providing genotypic or phenotypic targets for breeders; and,
- to predict yield and other complex trait outcomes for the range of Es in the TPE, using low-level genotypic or phenotypic information, potentially allowing early screening of genotypic suitability to Es.

If G and E effects on yield and other complex traits of interest were entirely additive, it would be sufficient to use linear statistical models to achieve these insights. The presence, however, of non-linear GxE interaction effects requires the use of mechanistic crop growth models that embed knowledge of crop physiology and respond dynamically to E conditions (Cooper et al., 2020a). Such models should: emulate, in biologically-realistic ways, the development and growth of the plant; predict complex phenotypes as emergent outcomes, the consequences of lower-level processes responding to E drivers and regulated by stable (across Es) G trait parameters; and be as simple and computationally-undemanding ("parsimonious") as possible (Boote et al., 2021; Hammer et al., 2019b; Hammer and Jordan, 2007; Singels, 2014). In order for crop models to support trait dissection and phenotypic prediction applications, the models must be able to predict GxE interactions in yields and other important agronomic traits accurately.

Several process-based sugarcane crop growth simulation models have been developed. Of these, the DSSAT-Canegro (Inman-Bamber, 1991; Jones and Singels, 2018; Singels et al., 2008), APSIM-Sugar (Dias et al., 2019; Inman-Bamber et al., 2016; Keating et al., 1999) and Mosicas (Martiné and Todoroff, 2004) models are amongst the most widely-used. These models follow a broadly similar framework of simulating on a daily basis, integrated over the duration of the crop, (1) the growth of the crop canopy in order to determine the daily leaf area index; (2) the daily fraction of solar radiation intercepted by green leaves (*FI*), via a canopy radiation extinction coefficient (*Ke*); (3) accumulation of dry biomass yield via photosynthesis, calculated as the product of solar radiation, *FI*, and radiation use efficiency (*RUE*); (4) dividing accumulated dry biomass between roots (*RDM*), and above-ground parts of the crop (*ADM*); (5) dividing *ADM* between leaves (*LDM*) and stalks (*SDM*), based on a stalk partitioning fraction (*SKPF*); and (6) dividing *SDM* between sucrose (*SUC*), fibre and other sugars. This is illustrated in Figure 1-1.

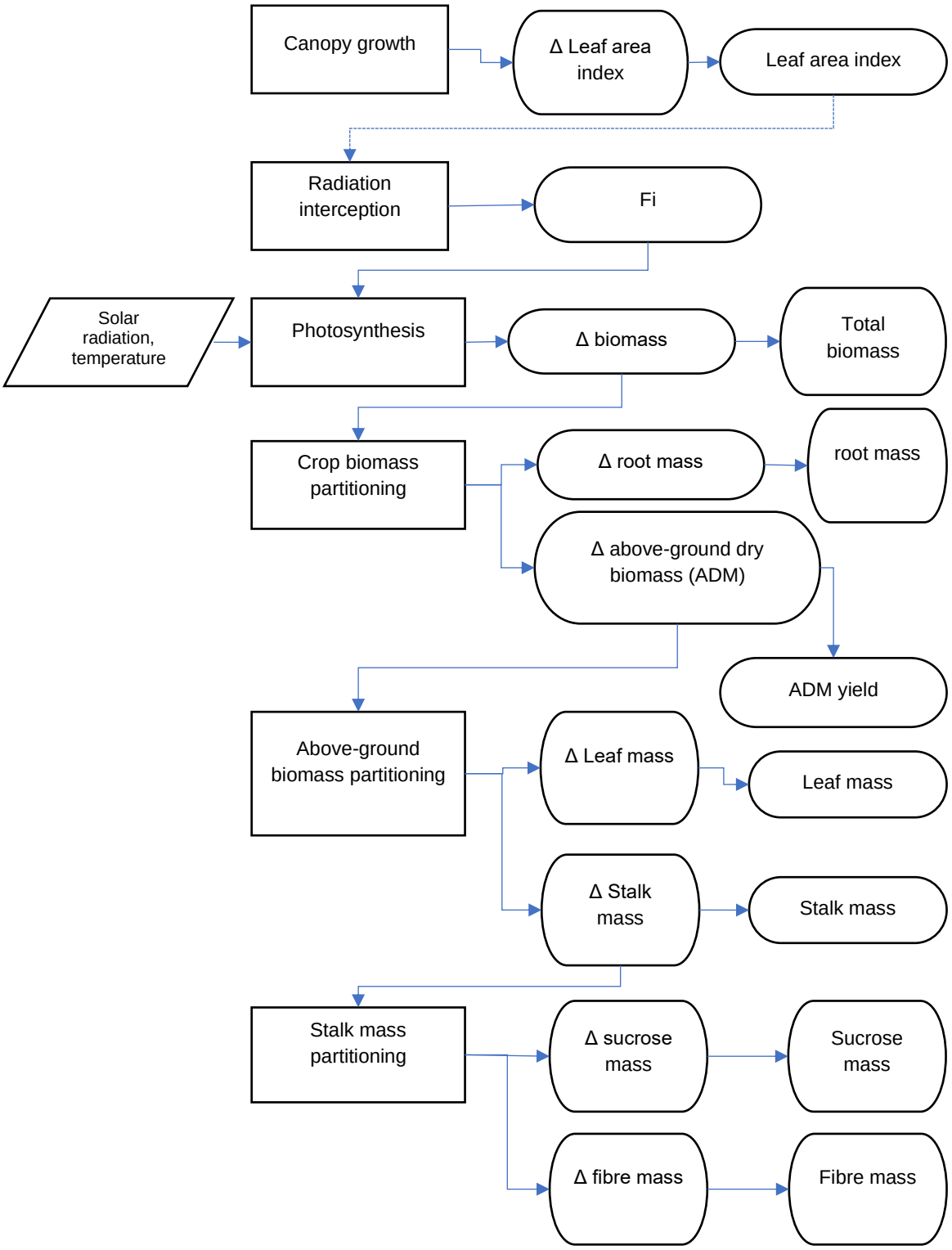

Figure 1-1. Flowchart illustrating the basic daily calculations made during each day of sugarcane crop growth simulation. Rectangles indicate processes and rounded rectangles are variables. Solid lines indicate the flow of data calculated on the same day, while dashed lines indicate input from values calculated the previous day. All masses are dry masses. Acronyms: Fi is 'fractional interception of radiation', and 'ADM' is 'Above-ground dry biomass'.

Applications of these models have included harvest management (Inman-Bamber, 1991; Singels et al., 2005c), irrigation scheduling (Inman-Bamber et al., 2007; Singels, 2007; Singels and Donaldson, 2000), yield forecasting (Bezuidenhout and Singels, 2007; Pagani et al., 2017), estimating yield gaps (Jones and Singels, 2015; Van den Berg and Singels, 2013), and assessing climate change impacts (Everingham et al., 2015; Jones et al., 2015; Marin et al., 2011; Singels et al., 2013). Only a small number of studies have considered cultivar differences in sugarcane crop models (Dias et al., 2020; Hoffman et al., 2018; Sexton et al., 2017). These models provide G-specific model input parameters for many of their constituent simulated plant processes, such as germination, leaf development, tillering, photosynthesis, and biomass partitioning. Nevertheless, no studies prior to this project had set out to evaluate models' abilities to predict GxE interactions in yields or their component processes. It is also not known how appropriate existing model input parameters are in representing G differences.

In recognition of these knowledge gaps, and the potential value of suitable crop models for supporting sugarcane breeding, the International Consortium for Sugarcane Modelling (ICSM) launched and funded the "International GxE modelling project" (IGEP). A fully-irrigated multi-environment, multi-genotype trial was conducted as part of this project, before the start of this PhD study. This ICSM IGEP dataset, which consists of growth analysis data for four sites (Pongola, South Africa; Chiredzi, Zimbabwe; La Mare, Reunion Island; and Belle Glade, USA) and three common cultivars (N41, R570 and CP88-1762), forms the foundation for the research work conducted in this study.

1.2 Problem statement

It is not known if sugarcane crop growth models exist that are capable of credible application in supporting plant breeding, as none have been evaluated for, or improved in respect of, their abilities to predict GxE interaction effects in complex traits such as biomass yields.

1.3 Research questions and hypotheses

The research questions posed at the outset of this study are as follows:

1. Are observed genotypic differences adequately captured by process concepts and their respective G-specific input parameters in existing sugarcane models, and what are the shortcomings in these respects?
2. To what extent are existing sugarcane crop growth models suitable for supporting plant breeding applications, as evidenced by their ability to predict GxE interaction effects on seasonal radiation interception, radiation use efficiency and above-ground biomass yields accurately, and what conceptual shortcomings prevent existing models from predicting these GxE interactions more accurately?
3. To what extent can the prediction of these GxE interaction effects be improved if the identified shortcomings are addressed in a revised or new sugarcane model?
4. How would such an improved model deliver value to sugarcane breeding?

It was hypothesised that crop simulation modelling capacity to support breeding of irrigated sugarcane could be enhanced by:

- evaluating the strengths and weaknesses of existing sugarcane models and/or their constituent process-level concepts for predicting GxE interaction effects, observed in the ICSM IGEP trials, on radiation interception, radiation use efficiency and biomass yields; and
- if necessary, developing an improved (new or revised) sugarcane crop model to address identified weaknesses

1.4 Objectives

The broad objective of this project was to improve the ability to model sugarcane crop growth and development for application in sugarcane breeding. The specific objectives were to address the research questions, as follows:

- to explore and assess the adequacy of process-level concepts and input parameter definitions from existing sugarcane models for predicting genotypic differences, using the ICSM IGEP dataset;
- to evaluate existing sugarcane models for their abilities to predict E, G and GxE interactions in seasonal radiation interception, radiation use efficiency and biomass yields observed in the ICSM IGEP dataset ;
- to develop and evaluate an improved sugarcane model, capable of accurate prediction of biomass, stalk and sucrose yields, and more accurate prediction of GxE interactions in seasonal radiation interception and biomass yields compared to existing sugarcane models, while also fulfilling other requirements for models to support plant breeding, such as biological realism, emergent prediction of complex traits and model parsimony;
- to apply the model in a case study to demonstrate its value in sugarcane breeding; and,
- to make recommendations for further research.

The envisaged outcomes of this project were:

- an improved understanding of the capability of existing sugarcane models for supporting irrigated sugarcane breeding;
- improved understanding of the physiological bases for GxE interactions in irrigated sugarcane yields;
- an improved crop growth simulation model for supporting irrigated sugarcane breeding;
- An improved understanding of prospects for practically incorporating crop modelling into sugarcane breeding programmes.

1.5 Thesis structure

This thesis is divided into seven chapters.

- A literature review (Chapter 2) follows this introductory chapter, in order to provide the scientific context and background for this work, in terms of sugarcane physiology and management, sugarcane modelling approaches, and the use and nature of crop growth models in supporting plant breeding.
- The third chapter addresses the first research question: it provides an analysis of the ICSM IGEP dataset, assesses common plant process simulation concepts (and their G parameters) from several sugarcane crop growth models against observations, and makes recommendations for addressing apparent

shortcomings in these concepts. This work was published as a scientific paper (Jones et al., 2019) and is presented as such.

- Chapter 4 addresses the second research question. It provides a detailed evaluation of three sugarcane models for their abilities to predict E, G and GxE interaction effects, and in so doing provides insight into their value as tools for supporting plant breeding. Recommendations for model improvements are made. This is also presented as a published scientific paper (Jones et al., 2021).
- Chapter 5 addresses the third research question, and describes a new sugarcane model, CaneGEM, that sets out to achieve improvements in predictions of GxE interactions in seasonal radiation interception, and, consequently, biomass yields, by implementing the recommendations identified in Chapters 3 and 4.
- Chapter 6 addresses the final research question by providing a case study demonstrating the value of the improved model for plant breeding. The case study methodology is applied at the each of the sites included in the original ICSM IGEP dataset.
- Chapter 7 provides discussion that addresses the research questions posed in this introductory chapter, and provides recommendations for future work.
- Finally, Chapter 8 presents a summary and conclusion.

Two appendices are included for additional data.

2. LITERATURE REVIEW

TABLE OF CONTENTS

2.1 Introduction 10
2.2 Sugarcane 10
2.3 Crop improvement and G, E and GxE interaction effects 14
2.3.1 Introduction 14
2.3.2 Sugarcane breeding programmes and selection strategies 15
2.4 Applications of crop modelling to support crop improvement 17
2.4.1 Crop growth simulation models 17
2.4.2 Crop modelling as part of an integrated system of crop improvement 17
2.4.3 Representing genetic traits in CGMs 18
2.4.4 Environmental characterisation using CGMs 19
2.4.5 Trait dissection 20
2.4.6 Phenotypic prediction in the TPE 23
2.4.7 The nature of crop models to support crop improvement 24
2.5 Sugarcane crop models 25
2.5.1 Introduction 25
2.5.2 Germination 25
2.5.3 Root development and growth 26
2.5.4 Leaf appearance 27
2.5.5 Tiller development and senescence 27
2.5.6 Leaf expansion and canopy formation 28
2.5.7 Leaf senescence 29
2.5.8 Radiation interception, photosynthesis and respiration 30
2.5.9 Stalk growth 33
2.5.10 Biomass partitioning 34
2.5.11 Water relations 35
2.5.12 Soil temperature 35
2.5.13 Nutrition 36
2.5.14 Discussion and key challenges 36
2.6 Summary and Conclusion 38

2.1 Introduction

The ultimate goal of this work was to improve the ability to model sugarcane crop growth and development for application in sugarcane breeding. The broad objective of this review is to report on science relevant to breeding/crop improvement, sugarcane crop models, and how crop models (in sugarcane and other crops) have been used to support crop improvement. This informed the exploration of the ICSM IGEP multi-environment trial dataset, the evaluation of existing sugarcane models for predicting G, E and GxE interaction effects, and the development of an improved model to support sugarcane breeding.

The specific objectives of this literature review were to:

- describe the sugarcane plant and the physiological processes that sugarcane crop models emulate;
- introduce the concepts of crop improvement and genotype-by-environment (GxE) interaction effects;
- report on the state of the science regarding the use of crop growth models to support crop improvement; and
- describe commonly-used sugarcane models, their underlying concepts and genotype-specific trait parameters, and identify knowledge gaps, shortcomings and opportunities with respect to the application of models for crop improvement applications.

This study focussed on irrigated sugarcane, and the sugarcane-specific aspects of this review are accordingly mostly limited to reporting on previous work relevant to irrigated sugarcane production and breeding.

The information presented in this chapter provides the scientific context and background to the research activities reported on in the following chapters.

- The assessment of model concepts and input parameter definitions reported in Chapter 3 draws on Sections 2.3 and 2.5;
- Sections 2.2, 2.3 and 2.5 inform the work presented in Chapter 4, which provides a critical evaluation of existing sugarcane crop growth models for predicting E, G and GxE interaction effects on seasonal radiation interception and biomass yields;
- Sections 2.4 and 2.5 provide the scientific basis for the aims and approaches for the development of a new sugarcane crop growth model, intended to support breeding of irrigated sugarcane, described in Chapter 5;
- Context for the case study presented in Chapter 6 is presented in Sections 2.3 and 2.4

2.2 Sugarcane

Sugarcane (interspecific hybrid of *Saccharum* spp.) is a large, perennial tropical/sub-tropical grass, that accumulates sucrose in stem tissues (Moore et al., 2013). Approximately 26.7 million ha of sugarcane were harvested worldwide in 2019 (FAO, 2021), primarily for the production of sugar and bioethanol. The largest producers are Brazil, India, Thailand and China, with sugarcane forming important parts of the local economies of smaller industries such as South Africa, Zimbabwe, the USA and French overseas territories.

Sugarcane is established with stalk cuttings in “plant” crops; following harvest, the crop regrows from the root stubble, termed “ratoon” crops. The number of ratoon crops per replant cycle ranges from zero (in Hawaii, Anderson et al. (2015)) to ten or more (e.g. Reunion island, Dumont et al. (2021)) depending on industry conventions in different parts of the world.

Ten developmental phases are defined for sugarcane by Bonnett (2014): germination, leaf development, tillering, stem elongation, development of harvestable parts, emergence of inflorescence, flowering, fruit development, seed ripening and senescence. Germination is the process of underground bud swelling and the subsequent elongation of coleoptiles (which begins on the date of “sprouting”) toward the soil surface, leading to “emergence”. These phases often overlap to some extent. Flowering, which is generally not desirable in commercial situations, requires specific environmental conditions, and is usually sporadic (and is not further discussed in this review). Ramburan (2012) considered three development/growth phases in an analysis of sugarcane cultivar trial data: establishment, elongation and ripening, although no clear definition for these phases was provided. From an agronomic perspective, sugarcane can be considered to have three key development/growth phases: germination, tillering (the development of shoots and leaves to build a complete canopy) and stalk growth (which includes stalk elongation and sucrose accumulation) (e.g. Liu & Bull (2001)).

The early growth dynamics of plant and ratoon crops are different. Sugarcane primary shoots develop from underground buds. The emergence of primary shoots takes place over time (usually several weeks) (Poser et al., 2019). Plant crops take longer to germinate and emerge (‘sprout’), and the bud density of ratoon crops is usually higher (Bezuidenhout et al., 2003). Not all buds successfully germinate and those that do produce shoots that emerge; sett/bud germination is determined by soil temperature and soil moisture content (Keating et al., 1999; Poser et al., 2019; Smit, 2010). Although plant crops are slower to develop canopy cover (due to the smaller number of buds at crop start), plant crops often yield higher than ratoon crops, due to the build-up of pathogens and damage during harvest prior to ratooning of the subsequent crop.

Rooting fronts penetrate rapidly and can reach depths of several meters, although the greatest density of roots is generally found in the top 50 cm of the soil profile. Sugarcane is generally not tolerant of waterlogged conditions, although some cultivars (e.g. NCo310) are noted for their ability to withstand waterlogging. Sugarcane roots are also considered to be intolerant of high soil acidity.

Leaf tips appear at increasing thermal time “phyllocron” intervals as the shoot ages, explained by increasing lamina length for the first 15 leaves; reasons for further increases in phyllocron intervals are not known. Leaf appearance rate is understood to be driven by temperature and restricted by water stress, and is strongly genetically-determined. (Bonnett, 1998).

Secondary shoots develop from the primary shoots, a process termed ‘tillering’. Each shoot supports leaves, and the combination of expanding leaf area and increasing tiller population contributes to the development, over several months, of a large canopy (in the absence of growth-limiting stresses). Tillering is strongly influenced by temperature.

The combination of leaf area index and leaf angle determine canopy-level radiation interception. Radiation interception is in turn a major determinant of photosynthesis and biomass accumulation rates, and therefore yield. Leaf size tends to increase with

the age of the shoot. Leaf elongation is understood to be driven by temperature (thermal time) and limited by water availability (Inman-Bamber, 1994; Robertson et al., 1998; Singels and Inman-Bamber, 2011), and to a lesser extent (due to variation in specific leaf area) carbohydrate availability (Keating et al., 1999; Singels et al., 2005c). Expansive growth of leaves is relatively sensitive to water stress (Singels et al., 2010b).

After a vegetative canopy has developed, shoots start to form above-ground fibrous stalk material; this occurs at approximately the time of peak population in a well-watered crop (Inman-Bamber, 1994), and is termed, the "onset of stalk elongation" (Dias et al., 2019), or the "stem elongation" stage (Bonnett, 2014). After this point, weaker (shaded) shoots begin to senesce (Bonnett, 2014), until a stable final stalk population is reached (Bell and Garside, 2005; Inman-Bamber, 1991). Final stalk population density has been shown to be sensitive to peak population (Singels and Smit, 2009), and peak population increases with narrower row-spacing and increased bud density within the row; environmentally, greater shoot population density is permitted by less-limiting soil moisture conditions (Smit and Singels, 2006).

Leaf senescence is driven by thermal time (Bonnett, 1998; Inman-Bamber, 1991), water stress (Smit and Singels, 2006), frost (Keating et al., 1999) and/or the leaf's carbon balance (photosynthetic income vs respiration costs) (Liu and Bull, 2001).

Photosynthetically-active radiation is intercepted and absorbed by green leaves, and this drives photosynthesis (Inman-Bamber and Thompson, 1989; Muchow et al., 1994; Robertson et al., 1996). Shading, and changes to the red-far red radiation ratio, affect tiller development and senescence, and the number of green leaves per stalk (Bezuidenhout et al., 2003; Donaldson, 2009; Singels and Smit, 2009). Leaves are capable of adapting/acclimatising to different radiation environments, via plasticity in specific leaf area (Keating et al., 1999). Photosynthesis rates are closely correlated with transpiration: as leaf stomata open to facilitate gas exchange for photosynthesis, water is lost (Bezuidenhout and Singels, 2007; Singels and Paraskevepoulos, 2017; Steduto et al., 2009).

Internode elongation creates capacity for sucrose storage. When growth conditions are less favourable for expansive growth (during cool and/or moderately dry conditions, or if flowering is initiated), photo-assimilate is stored as sucrose in the stalk rather than being synthesised into stalk and leaf fibre (Singels and Bezuidenhout, 2002; Singels and Inman-Bamber, 2011). Stalk growth requires warmth (daily average temperature generally above 16 °C (Lingle, 1999; Smit and Singels, 2007)), abundant soil moisture, and a healthy canopy intercepting photosynthetically-active radiation and thus supplying photoassimilate. Mild water stress signals the plant to reduce stalk elongation rates (Singels et al., 2010b), and water stress reduces cane yields (Basnayake et al., 2012; Singels et al., 2010b).

Sugarcane converts stored sucrose in buried setts and stools to leaf and root tissue during germination. During early growth, relatively more biomass is partitioned to roots than above-ground organs. Young shoots consist mostly of leaf sheath and leaves, and biomass is allocated to fibrous stalks after the shoot reaches that stage. (Bonnett, 2014; Robertson et al., 1996; van Dillewijn, 1952). Sucrose is converted into hexoses (mostly fructose and glucose) as part of the fibre synthesis process (Singels and Inman-Bamber, 2011), although the ratio of sucrose and hexoses changes dynamically (Inman-Bamber et al., 2005). It is understood that the partitioning of biomass is determined by dynamic and competitive source (supply of photoassimilate) and sink

(demand for carbohydrates) processes (Inman-Bamber et al., 2010, 2009; Singels and Inman-Bamber, 2011). Under favourable growth conditions, sink strengths for leaf and stalk fibre are high, leaving little excess available for sucrose accumulation. As photosynthesis is less sensitive to water stress and low temperatures than expansive growth, excess source can accumulate during periods of mild water stress and/or low temperatures (Inman-Bamber, 2004; van Heerden et al., 2013). This excess source can be stored as sucrose in the stalk. Sucrose content at harvest averages 13 % on a fresh mass basis (Cursi et al., 2022; Jackson, 2005; Zhou and Gwata, 2015) but can reach 17-18% (Jackson, 2005), and up to about 50% on a dry mass basis (Inman-Bamber et al., 2002; Waclawovsky et al., 2010)). Other significant stalk components include water, hexoses (glucose and fructose) and stalk fibre (cellulose, hemicellulose and lignin). Commercial millable stalk yields average 84 t/ha/an (fresh weight, about 20 t/ha dry mass), with commercial maximum yields of 148 t/ha across South Africa, Australia and Columbia (Waclawovsky et al., 2010), although commercial rainfed yields are generally lower (e.g. in South Africa, 40-70 t/ha on average, Van den Berg and Singels (2013)); in Brazil, yields recently (2019/20) averaged 76 t/ha (Cursi et al., 2022).

The allocation of biomass to different organs is controlled genotypically (e.g. some cultivars are noted to be 'high-sucrose' types, while others are noted for being relatively low in sucrose but produce greater biomass) (Singels and Inman-Bamber, 2011), but is also understood to have management and environmental influences (e.g. applying nitrogen near to harvest affects the ratio of sucrose to hexoses, and if water is withheld for a few weeks before harvest in an irrigated crop ('drying-off'), sucrose content tends to increase (Dias and Sentelhas, 2018; Robertson and Donaldson, 1998; Van Heerden et al., 2015a)).

Sugarcane grows throughout the year. Water use is dependent on the environments in which it is grown, but sugarcane potential evaporation is higher than a reference short grass (≈ 25% higher, e.g. McGlinchey & Inman-Bamber (2002)). Although the species is drought-tolerant in the sense that total crop failure due to drought is very rare, medium to high rainfall (> 1000 mm) and/or irrigation is required to produce economical high biomass crops. Sugarcane appears to be subject to maximum transpiration rates of about 8 mm/day, and has been shown to exhibit transient midday water stress even under generous soil water conditions when vapour pressure deficit is very high (Inman-Bamber et al., 2016). Mild water stress results in reductions in expansive growth rates, while more severe water stress also limits photosynthesis rates (Singels et al., 2010b). Water availability affects biomass partitioning (Singels and Bezuidenhout, 2002).

The ideal climate for sugarcane is one where abundant water is available to support structural growth during warm periods, and where conditions are cooler (but not freezing) and drier prior to harvest. This allows the crop to develop its canopy (photosynthetic apparatus) and sucrose storage capacity, and then use the canopy to produce sucrose which is then stored in the stalk. Sugarcane does not tolerate frost – leaves can be damaged by frost, and the growing point can be killed in the case of a severe frost; this results in a loss of apical dominance, followed by a rapid decrease in cane quality as the plant starts to use stored sucrose to build side shoots that develop from buds on the stalk. Apart from the loss of sucrose and reduced sucrose content, side shoots make harvesting difficult and transport to the mill inefficient. Sugarcane generally thrives under warmer conditions (as long as soil moisture is available), but

extreme high temperatures are likely to be damaging. (Singels et al., 2021; Van Heerden, 2014; Van Heerden et al., 2015b).

Very large (high biomass) crops can lodge (fall over), especially if there are strong winds and if the soil/canopy are wet (Van Heerden et al., 2015b). This interferes with the harvesting process and can result in side-shooting, which reduces cane quality (i.e. the fraction of sucrose in dissolved solids in the stalk juice).

Harvest age is usually between 12 and 24 months. Plant crops are in some cases grown for longer periods than ratoon crops as canopy development can be slower in plant crops. Irrigated cane is normally harvested at 12 months. In high-altitude regions, such as the KwaZulu-Natal Midlands in South Africa, rainfed cane is grown on 15-24 month cycles. The production system in Hawaii is somewhat unique, being characterised by 24-month irrigated cane that is not ratooned but replanted every season.

Sucrose content is lowest in warm and wet times of year, so sugarcane mills generally have a milling season that avoids these periods. The length of milling season (LOMS) in South Africa, for example, is usually about 9 months, starting in autumn (March) and completing in late spring/early summer (November/December). The actual LOMS each season is determined by estimated total crop size and by delays due to mill breakdowns and excessive rain interrupting harvesting. Harvest age needs to be managed such that harvesting occurs during the milling season and that cane is supplied by growers to the mill at a consistent rate, to avoid deterioration of cane due to excessive harvest-to-crush delays. Cane that cannot be harvested before mill closure is 'carried over' and harvested at the start of the following milling season (Moor and Wynne, 2001).

2.3 Crop improvement and G, E and GxE interaction effects

2.3.1 Introduction

Crop improvement refers to the breeding of superior cultivated varieties (cultivars). Hammer and Jordan (2007) summarise crop improvement as a '*search strategy on a complex adaptation or fitness landscape, which consists of the phenotypic consequences of genotype (G) and management (M) combinations in target environments (E)*' with the objective of producing superior genotypes (varieties) for the 'target population of environments', or TPE (Hammer and Jordan, 2007; Ramburan, 2012). Superiority is defined usually as either (high) yield or yield stability (Cooper and Byth, 2002), or a selection index based on economic value (Jackson et al., 2021).

Different production environments have different yield potentials, due to characteristics of soil and climate. The influence of the environment on yield, i.e. the variation in yield due to E characteristics, can be considered the "E effect". Different Gs, on average across Es, can produce higher or lower yields; the variation in yield explained by G is termed the "G effect". In some cases, however, genotype G1 might out-yield genotype G2 in environment E1, but not in E2. This is termed the "GxE interaction effect".

Genotype-by-environment (GxE) interaction is a term that formally refers to '*a statistical decomposition of variance and provides a measure of the relative performance of genotypes grown in different environments*' (de Leon et al., 2016, p. 2082). In some cases, and in the context of this project, management (M) is generally considered part of E. G, E and GxE interaction effects are assessed for multi-environment trials (METs) using statistical techniques. In cases where GxE interaction effects are insignificant,

high-ranking genotypes are said to be 'broadly adaptable'. In cases with significant GxE interaction effects, the ranking of genotypes changes depending on the environment, permitting the identification of genotypes that are specifically-adapted ('niche-bred') for particular Es (Cooper and Byth, 2002).

GxE interaction can alternatively be considered conceptually as a '*measurement of the relative plasticity of genotypes in terms of the expression of specific phenotypes* [e.g. yield] *in the context of variable environmental influences*' (de Leon et al., 2016, p. 2082). Phenotypes refer to physical traits of the plant when grown – e.g. yield, sucrose yield, maximum leaf area index – and so represent the outcome of the genotype responding to the environment in which it is grown over the duration of the season.

Values for a given trait (such as cane yield at harvest) for two genotypes might:

- be completely unaffected by a change in environment (no E effect, so no GxE interaction effect);
- change similarly for the change in environment (i.e. changes in the same direction and of similar magnitude) (no GxE interaction effect);
- show unequal changes for the change in environment, but the ranking remains unchanged (i.e. the direction of change is the same, but the magnitudes of the change differ) (GxE interaction effect evident);
- respond in opposite directions, so-called 'crossover interactions' – where the ranking of genotypes in terms of trait value changes – when moving from one environment to another. (de Leon et al., 2016; Ramburan, 2012) (GxE interaction effect evident)

Using ANOVA (analysis of variance) techniques, it possible to express yield (Y_{ij} for genotype *i* and environment *j*) in terms of additive main effects, determined by G and E, and non-additive effects due to interactions between G and E (i.e. GxE):

$$Y_{ij} = \mu + G_i + E_j + GE_{ij} + e_{ij} \qquad (2\text{-}1)$$

where μ represents the grand mean (over all genotypes and environments), and G_i, E_j and GE_{ij} represent the effects of the genotype, environment and GxE interaction respectively. The error term (capturing influences not accounted for by *G*, *E* or *GE*) is e_{ij}. (Ramburan, 2012). Other statistical techniques for evaluating GxE interactions in MET datasets are described by Ramburan (2012), and include stability analysis (whereby yield means per G are regressed against marginal environmental yield means as a surrogate for the environmental influences on yield), and various analyses based on principal component analyses (PCA, whereby high-dimensional data are projected into a (usually) 2-dimensional sub-space orthogonal to the eigenvector calculated by minimising the sums of squares of data points along each axis) combined with analysis of variance.

2.3.2 Sugarcane breeding programmes and selection strategies

An overview of sugarcane breeding is provided by Yadav et al. (2020). The "vast majority" of sugarcane breeding programmes worldwide make use of a fairly standard approach of selecting parents, creating genetic variability through cross-pollination, then assessing progeny through three (or more) stages of assessment and selection, followed by cultivar release (and possible inclusion of that cultivar as a parent in future crosses). The selection of parents is based on identified desirable agronomic traits (e.g. cane yield, estimated recoverable sucrose, disease resistance). The first stage

of selection is undertaken at family level, with subsequent phenotypic assessment (based on selection indices) being undertaken at individual genotype level; approximately 5-10% of the top-performing families or genotypes are selected for consideration at the next stage. This breeding approach was effective in the 20th century, driving significant genetic gain in sugar industries around the world. Although sugarcane genotypes are pure hybrids (as each is clonally propagated from a single plant after growing from seed following crossing), this approach is nevertheless extremely costly and time-consuming (requiring in excess of 10 years per released cultivar). These resource constraints suggest that breeders aim to release broadly-adaptable varieties with good performance across all Es in the TPE, and so with minimal GxE interaction effects. GxE interaction effects can however reduce the rate of genetic improvement in sugarcane, by decreasing trait heritability.

Short descriptions of sugarcane breeding programmes in three industries (South Africa, Zimbabwe and Reunion) are presented below.

The South African Sugarcane Research Institute's breeding programme consists of crossing of parent sugarcane varieties performed under artificial environmental conditions, after which seed is germinated in a glasshouse and then sent to seven locations representing different agro-climatic regions of the South African sugar industry, for selection. Approximately 250 000 seedlings are assessed, from which one or two cultivars are released each year, 10-14 years later. In early stages, selection is based on visual inspection of disease symptoms and yield estimated from stalk length and diameter. Selection in later stages is based on "Recoverable Value" (RV) yield, based on a formula that considers juice purity, fibre and ash contents, and pest/disease resistance. An extensive post-release cultivar evaluation programme is in place, which includes trials on commercial farms. (Ramburan, 2011).

The objective of the Zimbabwe Sugar Association Experiment Station's breeding programme is to develop adaptable, smut-tolerant/resistant varieties with high cane and sucrose yields, and reliable ratooning abilities (Zhou, 2005) for an industry that produces cane almost exclusively under irrigation. After early-stage selection, stools are planted into single lines, then selected lines are assessed in replicated cultivar observation trials up to the third ratoon crop, as well as Smut inoculation trials (SITs). Selected genotypes are then planted to advanced cultivar trials for a further three ratoon crops, after which they are divided between four locations (with differing representative soil types) for pre-release evaluation up to the third ratoon. Evaluation includes yield, sucrose content, Smut and Eldana (stalk borer) and ratooning ability assessments. (Zhou, 2005).

Sugarcane breeding in Reunion is conducted by eRcane (Dumont et al., 2021). Reunion is characterised by widely-varying environmental conditions (including irrigated production) across a relatively small area, and so after conducting crosses and generating 100 000 seedlings at the hybridisation facility at La Bretagne, genotype selection is conducted at seven locations representing different agro-climatic production zones on the island. Agronomic trait priorities include high sugar productivity (cane yield x sucrose content), disease resistance, low propensity to flower, and high ratoonability (with 10-12 ratoon crops typically harvested per replant cycle). Selection is undertaken over five stages in a 14-year period and includes inoculated disease resistance assessments.

Sugarcane breeding is time-consuming (10-14 years per cultivar) and resource-intensive. The annual budget for the South African breeding programme is ≈ ZAR60

million (US$4million at the time of writing), so at a rate of 1-2 cultivars released per year, the cost of single N cultivar is US$2-4million. Considerable genetic gains have been achieved since 2000 in most TPEs in South Africa (Zhou, 2017; M. M. Zhou, 2013), although worldwide sugarcane yields have plateaued (Yadav et al., 2020) and the rate of genetic gain in sugar yields from genetic improvement has declined (Acreche et al., 2015), driving interest in model-assisted sugarcane breeding.

2.4 Applications of crop modelling to support crop improvement

2.4.1 Crop growth simulation models

Crop growth simulation models (CGMs) are computer programs that represent simplified mathematical analogues of cropping systems. Crop models can be used as research tools for testing hypotheses relating to crop physiology, genetic effects, environmental influences and management approaches (Singels, 2014). They can also be used in a more tactical, applied manner, such as for developing irrigation schedules (Inman-Bamber et al., 2007; Singels, 2007) or forecasting yields (Bezuidenhout and Singels, 2007). Some crop models attempt simply to describe relationships between inputs (generally descriptions of management, soil and climate) and desired outputs (yield, water use, etc), and such models are often based wholly or partially on empirical relationships. These models are often fit for purpose – they can provide accurate results for the regions and environments for which they have been calibrated. More sophisticated crop models are described as 'mechanistic' or 'process-based' (Jones and Luyten, 1998). These models attempt to emulate the internal functioning of a plant, in a physiologically-realistic manner.

Process-based crop modelling emerged in the mid-1960s, with discipline-based quantitative theories of radiation interception and photosynthesis, crop water relations, evaporation and transpiration, and plant phenology eventually becoming integrated into general purpose simulation models in the 1980s – examples include the CERES-Maize (Jones et al., 1986), Canegro (Inman-Bamber, 1991) and CropGro (Boote et al., 1998) models. Many of these models were included in modular software suites offering simulation capacity for multiple crops and cropping systems (e.g. sequences and rotations) in the late 1990s (e.g. Decision Support System for Agro-technology Transfer (DSSAT, Jones et al. (2003), Agricultural Production Simulator (APSIM) (McCown et al., 1996)).

2.4.2 Crop modelling as part of an integrated system of crop improvement

Hammer & Jordan (Hammer and Jordan, 2007) outlined an integrated systems approach to crop improvement, in which crop modelling can assist in three broad ways: (i) characterisation of environments; (ii), understanding and dissecting complex trait physiology and genetics; and (iii) predicting complex phenotypes (such as yield) in the TPE. One of the core tenets of this approach is changing the breeding paradigm from a search for superior genotypes to a search for superior combinations of genetic regions and packaging these as released cultivars.

In reviewing model-assisted breeding, several themes emerge, as the development and ongoing improvement of:

- the representation of genetic effects in crop growth models (CGMs), and the use of crop models as gene-to-phenotype translation functions;
- the application of CGMs for characterising environments to increase the precision of selection;

- the use of crop models to dissect complex traits, and the use of sub-trait phenotypic values to assist breeding;
- techniques for predicting sub-trait and complex phenotypic values from DNA sequence data;
- defining breeding targets for TPEs;
- using genomic selection techniques in breeding; and
- recognising the importance of crop management-by-genotype interactions and addressing yield gaps as part of the breeding process

There is a degree of overlap between these themes due to the presence of shared technologies. Each of these is explored in the sections that follow.

This review concludes with an exploration of characteristics of crop models suited for different applications of crop modelling in breeding/crop improvement.

2.4.3 Representing genetic traits in CGMs

White & Hoogenboom (2003) defined six levels of genetic detail that can be accommodated in crop growth models:

1. Generic plant model not representing any one species
2. Plant model representing a species (no notion of cultivars)
3. Genetic differences represented with cultivar-specific parameters
4. Genetic differences captured with specific alleles, with the actions and effects of these genes represented with linear effects on model parameter values
5. Genetic differences represented by genotypic information, with gene actions being simulated on the basis of knowledge of gene expression and gene product effects
6. Genetic differences represented by genotypic information, with gene actions simulated at a metabolic level.

Most crop models represent genetic effects via cultivar-specific parameters (level 3) (Wang et al., 2019). Most sugarcane crop models (including DSSAT-Canegro (Inman-Bamber, 1991; Jones and Singels, 2018; Singels et al., 2008), APSIM-Sugar (Inman-Bamber et al., 2016; Keating et al., 1999), Mosicas (Martiné and Todoroff, 2004) and Canesim (Singels and Donaldson, 2000; Singels and Paraskevepoulos, 2017)) represent genetic effects in this manner. Most simulation studies conducted with sugarcane models have used the models' default cultivars (Q117 for APSIM-Sugar, NCo376 for Canegro and Canesim, R570 for Mosicas), rather than calibrating these models for other cultivars (Dias and Inman-Bamber, 2020). In this sense, relatively little use has been made of the models' abilities to simulate different cultivars. Where models are applied to represent systems containing many cultivars using a single cultivar parameter set, the models serve to explore the ExM landscape. Some examples of such studies have included yield gap analysis (Jones et al., 2015; Van den Berg and Singels, 2013), irrigation scheduling (Inman-Bamber et al., 2007; Singels et al., 1998), yield forecasting (Bezuidenhout and Singels, 2007), and exploring climate change impacts (Jones et al., 2015; Knox et al., 2010; Marin et al., 2013). Dias & Inman-Bamber (2020) provide a comprehensive review of historical applications of sugarcane models, most of which do not explore cultivar differences.

A key application of crop growth models at this level of genetic representation (i.e. without specific genotypic effects), to support breeding, is environmental characterisation. This is reviewed in Section 2.4.4.

The DSSAT system divides genetic parameters into “species”, “ecotype” and “cultivar” parameters. Species parameters are the same for all cultivars, cultivar parameters can be different for each cultivar, and ecotype parameters are those genetic parameters which have the same values for groups of related cultivars. For the CropGro template model (Boote et al., 1998) that runs within DSSAT, each (Boote et al., 2021) a crop species is represented by a set of parameters which configure the model to represent that species, and with the ecotype and cultivar parameters to represent cultivar differences. Several other modelling systems, such as AquaCrop (Steduto et al., 2009) and STICS (Brisson et al., 2003; Buis et al., 2011), use a similar concept of configuring a generic model of plant growth to represent different species and cultivars.

Some exploration of the level 4 representation of genetic effects in CGMs has been conducted. White & Hoogenboom (1996) developed the “GeneGro” model, which used linear combinations of input gene presence/absence and coefficient values to calculate parameter values internally for processes such as photosynthesis and photothermal time to flowering, for deterministic crops like wheat and soybean. The same gene input value is used in several process rate equations, capturing pleiotropic effects, where one gene affects multiple processes, and epistatic effects, where the presence/absence of one gene can affect the actions of another. Zheng et al. (2013) modified the APSIM-Wheat model to predict vernalisation date and photoperiod sensitivity from the presence/absence of several genes known to effect control over these processes. Oliveira et al. (2021) demonstrated the inclusion of a gene-based module in a traditional process-based CGM: associated QTL are used in a linear mixed model to predict time to flowering for in a modified version of the DSSAT CROPGRO-Drybean model. This hybrid model was able accurately to predict E, G and GxE interaction effects on time to flowering, as well as yields and other consequent outputs from the original Drybean model. The effectiveness of this modelling approach is limited by availability of knowledge of gene actions and quantified phenotypic responses (White and Hoogenboom, 2003).

An alternative, and far more common, framing of the level 4 representation is the use of Quantitative Trait Loci (QTL) rather than known gene actions. QTL are molecular markers, the presence or absence of which can be correlated via linear regression with, and significantly explain, variation in a quantitative phenotype, either a complex trait such as yield or a simpler sub-trait such as leaf elongation rate (White and Hoogenboom, 2003). This is discussed further in Section 2.4.5. Where sub-trait phenotypic values can be used as, or translated into, CGM input parameter values, the QTL predictions of these values can be used as CGM input parameter values and the CGM then used to predict complex phenotype outcomes for that genotype in the TPE (see Section 2.4.6).

No current crop growth models have achieved genetic representation at levels 5 and 6 (Boote et al., 2021). The current “Crops *in silico*” initiative is to develop a multi-scale simulation platform capable of simulating plant growth from gene-level to ecosystem scale (Marshall-Colon et al., 2017). Tardieu et al. (2020) do however argue that natural selection has constrained combinations of metabolic processes to ‘meta-mechanisms’ that can be adequately modelled at the higher integrated levels used by crop growth models.

2.4.4 Environmental characterisation using CGMs

Environmental characterisation is critical to understanding GxE interactions: in order to build a physiological understanding of E effects on G, which might lead via indirect

selection to improved G, the E must be defined sufficiently distinctly (Cooper and Byth, 2002). The term “Enviromics” (Cooper and Messina, 2021) collectively refers to techniques for characterising and clustering environments into categories (“envirotyping”). Ramburan (2012) reports on the use of approaches ranging from relatively simple models for characterising environments (such as seasonal soil water satisfaction, a measure of seasonal water availability; and seasonal transpiration relative to potential evapo-transpiration), to approaches making use of fully-fledged crop growth simulation models. An example of the latter is the CGM-based identification of categories of water-limited environments for sorghum in Australia (Chapman, 2000). By weighting genotype performance according to the extent to which the selection environments matched the target environments, greater breeding efficiencies could have been achieved. Chenu et al. (2009a) used a CGM to simulate wheat growth over a 100-year period for representative sites, soils and management systems in North Eastern Australia; the growth environments were characterised into five types based on the simulated patterns of crop water stress around flowering. Cultivars with different maturation characteristics were simulated at each environment type and their performance analysed; simulated yields were consistent with observed outcomes. Further analysis on this basis yielded insights into low-level traits (flowering date and pre-/post-anthesis water use) that conferred favourable yield performance in each environment type. Ramburan (2012) used growth phase-linked water stress indices derived from Canesim model simulations to assist with the characterisation of environments in sugarcane MET datasets in South Africa. This information was used to attribute causes of GxE interactions in post-release METs in South Africa. Hammer and Jordan (2007) suggest that many conventional crop growth models are sufficient and appropriate for characterising environments, particularly abiotic stress patterns experienced by crops.

2.4.5 Trait dissection

CGMs can be used in breeding to dissect traits – to identify and characterise process-level mechanisms that explain complex phenotypes (Hammer and Jordan, 2007), which can then provide insights into breeding priorities.

Hammer et al. (2010) used the APSIM model to explain G variation in sorghum grain yield as consequence of G control over plant height: compared to shorter plants, taller plants partitioned more N into immobilisable stem tissues; during grain filling, more N was consequently translocated from the leaves to support grain fill, which led to faster leaf senescence and reductions in photosynthesis rates, which resulted in lower grain yields. In a more recent study, Hammer et al. (2019a) showed that the “stay-green” trait in sorghum, which is associated with delayed canopy senescence in the post-anthesis period and higher yields in water-stressed Es, was not itself a G trait, but a consequence of G-regulated low-level processes: increased transpiration efficiency (with limited maximum transpiration rate), and reduced tillering; both of these traits reduced pre-anthesis water use (by directly or indirectly reducing water uptake) and retained sufficient water in the soil profile to maintain green leaf area for longer, with yield consequences consistent with observations of the “stay-green” cultivars. This trait is advantageous in water-stressed situations but incurs a yield penalty – by reducing photosynthesis rates per unit intercepted radiation, or reducing radiation interception – in non-stressed Es. Where mechanisms like these are not understood, considerable GxE interaction might lead to inefficient selections. Tsutsumi-Morita et al. (2021) dissected tomato yield traits using two very simple static yield component models, and associated QTL with these component traits as well as associating QTL

directly with yield. Although prediction performance for QTL associated with component yield traits was similar to when yields were predicted directly from QTL, heritability of the component traits was considered to be greater. Additionally, some of the component traits could be determined well before harvest, potentially leading to faster breeding.

Although the majority of sugarcane model applications have explored E and M issues, some work has focussed using sugarcane CGMs to explore genetic effects on complex traits. Sexton et al. (2017) analysed the APSIM-Sugar model's sensitivity to cultivar input parameters for two environments in Australia. Inman-Bamber et al. (2012) used the APSIM-Sugar model to explore emergent drought tolerance traits in sugarcane, finding that increased transpiration efficiency, increased rooting depth and reduced stomatal conductance lent adaptation advantages to sugarcane genotypes under most mild to moderately water-stressed environments, and could be used as selection indices for breeding for such environments. This was followed up with work to adapt the APSIM-Sugar to accommodate directly the physiological mechanisms explored, essentially adapting the model to support breeding for drought adaptation in sugarcane (Inman-Bamber et al., 2016). Zhou et al. (2003) calculated Canegro model parameter values for four southern African cultivars grown under irrigation in Zimbabwe, indicating that tillering parameters had a greater bearing on PAR interception than leaf parameters. Ngobese et al. (2018) assessed several Canegro model tillering and stalk elongation-related process parameters for 12 South African sugarcane cultivars for irrigated and rainfed conditions, and highlighted stable parameters that could be used in breeding. Hoffman et al. (2018) found significant differences in leaf-level photosynthesis rates, thermal time from primary shoot emergence to the onset of stalk growth, and stalk partitioning fraction for 14 genotypes in a well-watered pot trial. These were used as the basis for parameter values for the Canesim model, which was then used to simulate cultivar differences in yield for a field trial; emergent simulated yield rankings matched observations, confirming the role of sub-traits in regulating yield (Singels et al., 2016). Singels & Inman-Bamber (2011) used data from four clones and a modelling framework to conclude that stalk biomass partitioning to sucrose is a consequence of source strength and sink demands for structural growth of leaf and stalk fibre.

Chenu et al. (2009b) demonstrated how a process-based CGM (a modified version of the APSIM Maize model) could be used to translate QTL controlling leaf elongation rate (and known to co-localise with anthesis-to-silking interval under drought) to predict yield under different stress environments, for a simulated recombinant inbred line. A key finding was that traits increasing leaf elongation rate increased yield under low-stress environments, but decreased yield under water-stressed environments; these yield effects were entirely emergent outcomes rather than explicitly-programmed behaviours in the model. Gu et al. (2014) estimated QTL for seven trait input parameters for the GECROS rice model, for 94 introgression lines from parents varying in their drought response characteristics. Five QTLs significantly affected yield. The model-based trait dissection approach was then applied to 251 recombinant inbred lines from the same parents, and more significant markers were found for the sub-traits than for yield. Following model sensitivity analysis, sub-trait based ideotypes were defined that predicted yields 10-36% higher than ideotypes defined on basis of markers for yield only.

For sugarcane, Singels et al. (2010a) found stable QTLs for fully-expanded leaf area per leaf and a leaf-level photosynthetic capacity based on chlorophyll-*a* fluorescence

measurements, across three fully-irrigated field experiments for a mapping population of 80 sugarcane clones. This work did not include crop growth modelling, but the intention was to determine QTL for sub-traits that could be transformed into more complex phenotypes using a sugarcane CGM. Other QTL studies in sugarcane have generally focussed on predicting complex traits directly, e.g. Pastina et al. (2012) who found QTL for sugarcane yield, sucrose yield, sucrose content and fibre content, using novel methodologies for accommodating the interspecific polyploid nature of sugarcane as well as harvest year (as sugarcane is perennial crop grown over several harvests) in the supporting statistical analyses.

Genomic selection (Lorenz et al., 2011; Meuwissen et al., 2001) uses genotypic and phenotypic data to train statistical models, which can then predict phenotypic outcomes. In this approach, field trials are used for generating model training datasets rather than for selection; crosses are genotyped and this information used as input to the statistical models to predict yields; and selection is based on these predicted yields (Lorenz et al., 2011). Statistical models that predict yields from complete sequence data are termed “Whole-Genome Prediction” (WGP) models. Once trained, and where parents have been genotyped, WGP models can be used to evaluate crosses performed *in silico* (on computer), because frequency distributions of allelic combinations can be predicted statistically and WGP used to estimate phenotypic outcomes. The development of new statistical techniques (Heslot et al., 2012), rapid phenotyping techniques, and significant reductions in the cost of sequencing has resulted in dramatic change in plant breeding over the last 20 years: for most crop species some form of WGP approach is used, and only a small proportion of crosses are physically evaluated in METs (Messina et al., 2018). The genetic complexity (highly polyploid and aneuploid, with ≈ 120 chromosomes, Moore et al. (2013)) of sugarcane, along with the high cost of sequencing chips for sugarcane, however, means that applications of genomic tools in sugarcane breeding are delayed compared to other crops (Balsalobre et al., 2017; Yadav et al., 2020).

Where gene effects are additive, WGP methods can predict yields accurately; where gene effects are non-additive, or where there are GxE interaction effects, however, prediction with WGP methods is more challenging (Technow et al., 2015). Given the challenges associated with GxE interactions within the TPE, WGP models require enormous and prohibitively expensive training datasets in order to be accurate (Diepenbrock et al., 2021). If WGP models are combined with CGMs, where WGP outcomes predict sub-trait values that can be represented as CGM input parameter values, the CGM can be used to augment the training dataset with knowledge of physiological processes embedded in the CGM, and also unravel some of the more challenging GxE interaction effects and non-additive gene effects. The yield prediction accuracy for a biparental diploid maize inbred lines study was considerably higher for a WGP-CGM than for WGP alone in a simulation experiment (Technow et al., 2015). For sugarcane, Yadav et al. (2020) recommend applying genomic selection methods to simple traits, determined via high-throughput precision phenotyping and then associated with QTL, given GxE interactions in complex traits. The integration of newer enviromics techniques has the potential further to assist model-assisted breeding (Crossa et al., 2021).

Yield gaps measure the difference between actual yields, attainable yields and climatic potential yields (Van Ittersum et al., 2013). Yield gaps are attributed to agronomic management quality, for a given G and E. Co-selection of G and agronomic management characteristics has the potential to discover not only the optimal G for the

TPE, but also the optimal GxM. This is possible using WGP-CGM frameworks because the CGMs can account mechanistically for management approaches, complementing the WGP's determination of the genetic control of plant processes for that E-M combination. (Cooper et al., 2020b; Diepenbrock et al., 2021). The MET network in a breeding programme needs to be representative of the TPE, which is characterised by variation in soils and climate. Generally, rainfall is more variable that temperature and solar radiation. It is possible that the particular combination of Es (where an E is a site-year combination) encountered during selection and evaluation in METs is not properly representative of the TPE in terms of the progression of water stress. The use of managed stress Es, that impose particular stress patterns on selected Gs, can elucidate stress responses for selection or for training the WGP model in a representative manner. (Cooper et al., 2014; Diepenbrock et al., 2021). Cooper et al. (2021) describe a framework for managing the complexity of GxExM interactions in crop improvement, particularly in the context of climate change.

2.4.6 Phenotypic prediction in the TPE

Determining phenotypic performance – i.e. yield – of a new genotype in the TPE is of prime interest to a breeder (Hammer et al., 2019b). In traditional sugarcane breeding, the process of assessing and selecting candidate genotypes takes 10-15 years. If it were possible to predict final yields accurately from genotypic information, the time taken to breed new cultivars would be greatly reduced. This would free up resources to breed cultivars more optimally-adapted for smaller, 'niche', TPEs, which would likely increase overall yields. CGMs have great utility acting as gene-to-phenotype (Cooper et al., 2020a; Technow et al., 2015) "multi-trait link functions" for translating genetic information into complex phenotypic information, where the mechanistic process-level crop physiology embedded in the CGM can account for GxE interactions and non-linear effects more accurately than pure statistical methods.

The development of 'ideotypes' (Donald, 1968) is a key breeding application of robust phenotypic prediction in the TPE. The complex phenotypic outcomes of different sets of model trait parameters, representing genetic traits, can be evaluated using a CGM. Combinations of trait parameters that maximise [simulated] yield across the TPE, or trait parameters that consistently result in increased yield in the TPE, can provide breeders with sub-trait phenotypic targets for the TPE. This could inform the selection of parents, as well as assist in selection. It can also direct phenotyping resources to focus on specific traits. (Hammer and Jordan, 2007). Peng et al. (2008) reported on the development of highly-productive "super" rice varieties, based in part on ideotyping of physiological and morphological traits over a period of two decades. Tao et al. (2017) designed barley ideotypes for future environments, using ensembles of climate projection models for managing uncertainty in future climate and barley CGMs for managing uncertainty in understanding of crop physiology.

As described in previous sections, phenotypic prediction is a key component supporting several pathways to model-assisted plant breeding. To summarise:

- in the simplest cases, phenotyping in terms of low-level traits can be translated into model inputs, and the CGM used to estimate the complex phenotypic consequences (e.g. yield) of these traits across the TPE;
- ideotyping via iterative CGM runs with different trait parameter combinations can identify desirable phenotypes, and with QTL, desirable combinations of genetic regions;

- if QTLs can be robustly associated with sub-trait values, crosses can be sequenced and CGMs used to predict yield in the TPE, potentially accelerating selection; and,
- crosses conducted *in silico* can predict new genotypes, which in WGP-CGM hybrid frameworks can be translated into phenotypic low-level model input trait parameters (via WGP) and then into complex phenotypes with emergent GxE(xM) interaction effects (via the CGM).

2.4.7 The nature of crop models to support crop improvement

The requirements for 'credible' (Singels, 2014) gene impact models are described by Hammer and Jordan (2007), Hammer et al. (2019b) and Hammer (2020). The user must have confidence in the model in terms of its abilities to separate genetic, environmental and management influences on complex traits (i.e. to simulate GxE interactions accurately, Boote et al. (2021)), and environmental inputs (soil and climate) must be accurately specified. Model and dataset capabilities must be evaluated (Stöckle and Kemanian, 2020) in this regard. The models need to emulate biological systems in a realistic, 'biologically robust' manner, and generate accurate qualitative responses to key factors (such as N and irrigation) (Hammer, 2020). They need to capture important physiological linkages and interactions, and should be implemented as a hierarchy of physiological processes and input variables (following Tardieu (2003)); the simulation of complex phenotypes should be the 'emergent consequence' of lower-level processes and their differential responses to environmental stimuli and hazards (Hammer and Jordan, 2007), to facilitate better connection of complex traits with their genetic regulation (Hammer et al., 2010). Singels (2014) and Zhou et al. (2003) suggested that genetic parameters in the models should have clear physiological meanings, and defined such that they are stable across environments (i.e. determined primarily by genetic factors with either minimal or very predictable environmental plasticity); they should be easily measured or derived from measurements (Parent and Tardieu, 2014); and definitions should cater for the interdependence of genes (capturing epistatic and pleiotropic effects). Trait parameter values should be estimated objectively and should be limited to realistic ranges (e.g. by using Bayesian approaches with a-priori distributions of parameter values, such as in Generalised Likelihood Uncertainty Analysis (GLUE), Jones et al. (2011)). Hammer et al. (2010) advised in favour of model parsimony, advocating the simplest possible modelling approaches. Hammer et al. (2019b) reiterated this view in exploring what might initially appear as a trade-off between biological realism and parsimony. Model-assisted genomic selection frameworks require biological realism to capture physiological mechanisms that drive GxE interaction effects, but also need to be as computationally undemanding as possible and require the fewest input data for operation as possible. The authors advocate multi-scale modelling platforms built on an hierarchical basis that allows simple algorithms to be replaced with more complex algorithms operating at a lower organisational level where necessary and/or possible.

The simulation performance of sugarcane models has been reported for individual cultivars (e.g. Keating et al. (1999)). The only study identified where the ability of a sugarcane model to differentiate between Gs in a multi-environment trial was described by Dias et al. (2020). In this study, the APSIM-Sugar model was used to predict G yield differences arising from canopy development traits for 21 Brazilian varieties grown at two sites in Brazil. Salmerón et al. (2017) used a stability index to assess the DSSAT CropGro-Soybean model's ability to predict GxE interaction effects. Rotili et al. (2020) evaluated the APSIM-Maize model for four hybrids across nine on-farm sites

in the Northern Grains region of Australia, and found that the model captured 88% of the variability across Gs and Es; no attempt was made to assess GxE interaction effect predictions specifically, however. The literature is more richly endowed with reports of model modifications better to support plant breeding applications (and by implication, prediction of GxE interaction effects), where the modifications are driven by hypotheses regarding modes of action rather than poor GxE prediction performance per se. For example, Inman-Bamber et al. (2016) modified the APSIM-Sugar model to support drought tolerance traits such as transpiration efficiency; Chenu et al. (2009b) modified the APSIM Maize model to use parameter values derived from QTL to regulate leaf elongation rate and anthesis-to-silking interval in response to drought, as well as kernel development. Wang et al. (2019) recommended more generic improvements to crop models for predicting GxExM interactions to support plant breeding, with a view to increasing the representation of genetic effects to levels 4 and 5. These include: improving physiological understanding of key plant processes; linking model input parameters to effects of allelic variation in genes, determined via QTL identification; and modifying model structures to account for gene expression on physiological processes. These recommendations are partly based on the recognition of uncertainties in physiological process understanding, where the model intercomparisons conducted within the Agricultural Model Intercomparison and Improvement Project, AgMIP, (Rosenzweig et al., 2012) revealed considerable uncertainties with respect to crop physiology in crops such as wheat. Boote et al. (2021) also make the point that models must be realistically sensitive to the types of breeding issues under exploration; the exploration of genetic factors controlling N use efficiency, for example, would require a model with well-proven N dynamics simulation capabilities. Models ought to be evaluated for their sensitivity to the required G and E factors and simulation capability in this regard before application in breeding.

2.5 Sugarcane crop models

2.5.1 Introduction

Singels (2014) provides a comprehensive review of several sugarcane simulation models (APSIM-Sugar, DSSAT-Canegro, QCane (Liu and Bull, 2001), Mosicas, Casupro (Villegas et al., 2005) and Canesim. Sugarcane crop simulation capacity is also offered by the STICS, AquaCrop and other generic crop models, and modified versions of standard models are frequently used for specific applications (e.g. GTP-Canegro, Jones et al. (2011)). To avoid confusion, “DSSAT-Canegro” is used to refer to various versions of the Canegro model, even versions prior to it being incorporated into the DSSAT framework.

The sections that follow describe simulation approaches for key plant processes, their known or likely GxE interactions, strengths and weaknesses of relevant simulation approaches, and their abilities to accommodate genetic differences. The analysis focusses mainly on three models: DSSAT-Canegro, APSIM-Sugar and Mosicas.

2.5.2 Germination

Germination involves swelling of underground buds on the sugarcane sett (plant crops) or stool (ratoon crops), followed by coleoptile elongation towards the soil surface, culminating in the emergence of primary shoots. The date of emergence is generally defined as the date when the 50th percentile viable bud emerges. Germination rate is most commonly (e.g. DSSAT-Canegro, APSIM-Sugar, Mosicas, Canesim) simulated on the basis of thermal time (calculated from air temperatures), with plant crops

requiring more thermal time than ratoon crops (Keating et al., 1999) from crop start to primary shoot emergence. APSIM-Sugar and DSSAT-Canegro consider thermal time delay from crop start to the start of the linear coleoptile elongation phase (i.e. sprouting), followed by a depth-dependent thermal time period for the shoot to reach the soil surface. In APSIM-Sugar, this elongation can be delayed due to insufficient soil moisture: in APSIM v7.8 (2017), germination is inhibited below 0.05 mm/mm plant-available soil moisture content, and reaches a maximum germination rate at 0.70 mm/mm soil moisture content. DSSAT-Canegro simulates the appearance of primary shoots over a thermal time window, while APSIM-Sugar, Mosicas and Canesim assume a single date of primary shoot emergence, defined as the date at which 50% of viable buds have emerged. APSIM-Sugar, Casupro and DSSAT-Canegro are sensitive to planting density in addition to bud depth.

Genetic trait parameters controlling germination include the thermal time from crop start to primary shoot emergence for plant and ratoon crops separately. The DSSAT-Canegro and Mosicas models permit setting of G-specific cardinal temperatures for germination thermal time accumulation, while APSIM-Sugar uses fixed values (across all cultivars and most plant processes).

Many simplifications are made in the simulation of germination in these models – most models assume a single date of emergence, some do not take into account planting depth or planting density (other than the implications of changing row-spacing), and only APSIM-Sugar considers soil moisture content. Air, rather than soil, temperatures are used to drive germination; soil temperatures are not routinely measured, although soil temperature is often simulated. These weaknesses detract from the biological realism of model operation and may have a negative impact on simulation accuracy.

Singels & Bezuidenhout (2002) forced the DSSAT-Canegro model to use predetermined observed dates of primary shoot emergence in order that their calibration and assessment of biomass partitioning algorithms were not confounded by errors in simulated date of emergence.

2.5.3 Root development and growth

The rooting front deepens either at a fixed rate per day (APSIM-Sugar) or per unit thermal time (DSSAT-Canegro, Casupro). Root density distribution with depth is an input parameter in DSSAT-Canegro. DSSAT-Canegro considers soil hydraulic conductivity (k) and root length density (L) separately (soil and species parameters respectively), while until recently APSIM-Sugar considered a combined soil-specific kL value as this is a more pragmatic, and more easily-determined, value. Root density is simulated as functions of root mass and root length density parameters (APSIM-Sugar, DSSAT-Canegro, Casupro). Root senescence is either not simulated or is simulated as a constant fraction of root mass (DSSAT-Canegro, QCane) or root length density and water stress (Casupro). Roots are not simulated directly by Canesim, Mosicas or QCane.

Simulation work by Inman-Bamber et al. (2012) using APSIM, and Singels et al. (2016) using DSSAT-Canegro, suggests that additional investment in root systems generally did not noticeably improve the fitness of genotypes under most water-limited environments. Inman-Bamber et al. (2016) developed a version of the APSIM-Sugar model, in which the soil k and G-specific L parameters are separated to support exploration of G and GxE effects in drought adaptation.

2.5.4 Leaf appearance

The APSIM-Sugar model predicts primary shoot leaf number based on leaf phyllocron intervals and accumulated thermal time. The phyllocron intervals change as a function of leaf number, using linear interpolation between leaf number-phyllocron pairs provided as input trait parameters. The DSSAT-Canegro and Casupro models predict leaf number using two leaf phyllocron intervals, a shorter interval for young shoots and a longer interval for older shoots, where the phyllocron interval changes at a certain leaf number. G-specific phyllocron intervals can be specified. Cardinal temperatures for thermal time accumulation driving leaf appearance can be specified in DSSAT-Canegro and Mosicas, but not in APSIM-Sugar. Canesim, QCane and Mosicas do not simulate individual leaves. Bonnett (1998) presented a continually variable power-law model of leaf appearance as a function of leaf appearance (base temperature 9 °C), with calibrated parameters for nine sugarcane cultivars. Sinclair et al. (Sinclair et al., 2004) reported biphasic model phyllocron interval values for four USA cultivars, with a base temperature of 10 °C. Leaf phyllocron intervals are considered stable across Es, implying G control over this trait.

2.5.5 Tiller development and senescence

Tiller development in DSSAT-Canegro (Inman-Bamber, 1991) was originally described by a parabolic equation of thermal time defined descriptively for each genotype. Bezuidenhout et al. (2003) developed a process-based tillering model for the DSSAT-Canegro model, which considered bud population and light competition above a radiation interception threshold. DSSAT-Canegro v4.6 (Jones and Singels, 2018) uses similar principles to calculate tiller development based on a G-specific maximum tillering rate per unit thermal time per primary stalk , with new tillers appearing at thermal time-determined intervals (based on Singels et al., 2005), and with the interval increasing in response to increasing radiation interception and/or water stress. The radiation interception effect is meant as a proxy for the red:far-red radiation ratio which is understood to have a controlling effect on tillering rates (pers. comm. Inman-Bamber, 2010). Tillering is not related in any way to source or sink strengths, and is insensitive to radiation intensity. Alam et al. (2014) found that E variation in tillering in sorghum, also a graminaceous C4-photosynthesis species, was determined by a photosynthetic supply/demand ratio, implying that tillering is also a growth process . Sorghum Gs were however also found to differ in their propensity to tiller, indicating G control as well.

Leaf phenology in DSSAT-Canegro is calculated independently for each tiller cohort.

APSIM-Sugar simulates a ‘tillering factor’, which increases (mimicking tiller development) and then decreases (mimicking tiller senescence) leaf area index as a function of leaf number, but secondary shoots are not simulated as such. The set of values defining the tillering factor in response to leaf number is considered G-specific. Stalk density at harvest is determined by the bud density at crop start (M input).

Other models (e.g. Canesim, QCane, Mosicas) do not simulate shoots/stalks at all, opting instead to simulate canopy cover directly. Canesim and Mosicas consider only temperature (via thermal time) for describing canopy development under non water-stressed conditions.

Tiller senescence in DSSAT-Canegro starts after a predetermined thermal time period (G trait input), with shoot population decreasing by a fraction each degree day thereafter to a predetermined (G input) final shoot population value; senescence rate

is increased with water stress. The extinction rate parameter is a G trait input value. APSIM-Sugar is very similar – the tillering factor is reduced according to leaf number, and leaf number is determined by thermal time; the tillering factor is G-specific. Ngobese (2018) found that the thermal time to peak population was not consistent between Es and Gs, in a study of 11 South African cultivars grown for plant and ratoon crops. This parameter was, on average, also smaller for ratoon crops – a dynamic also not generally reflected in model parameter values.

No sugarcane models simulate tiller senescence in a truly mechanistic manner. Bezuidenhout et al. (2003) defined a 'minimum sustainable leaf area' (MSLA) with the concept of senescing tillers whose leaf area fell below this threshold. This did not however consider source and sink strengths, and was calibrated empirically. Smit & Singels (2006) reported clear G differences in tiller senescence rate response to water stress.

All sugarcane CGMs essentially treat tillering as a developmental process, determined by temperature and (in the case of DSSAT-Canegro) the light environment via shading. Tillering could be viewed instead (or in addition) as a growth process, driven (or limited) by temperature and carbon availability, and subject to G control. Kim *et al.* (2010) found that tillering rates in sorghum were greater under high source:sink conditions. Tardieu *et al.* (1999) asserted that leaf expansion is largely independent of the carbon budget, suggesting that sugarcane source-strength responses in canopy cover are manifested as changes in tillering rates.

2.5.6 Leaf expansion and canopy formation

DSSAT-Canegro, APSIM-Sugar and Casupro estimate leaf area per leaf, driven by temperature and limited by water stress, and scale this to canopy level using shoot population (DSSAT-Canegro) or the primary stalk population and tillering factor (APSIM-Sugar). In DSSAT-Canegro, the number and size of leaves is calculated independently for each shoot cohort, and so total leaf area index reflects this heterogeneous population of shoots. In models that simulate individual leaves, G parameters determine the maximum leaf size profile. The DSSAT-Canegro model scales this internally according to a maximum leaf area G trait parameter value, while APSIM-Sugar reads in a series of leaf area per leaf number combinations and interpolates linearly between these. Leaf area expansion is calculated by translating leaf elongation (driven by temperature) into area using a shape factor (typically 0.71).

Casupro and APSIM-Sugar couple leaf expansion rate to carbon (photo-assimilate) availability; APSIM-Sugar restricts leaf expansion if the daily allocation of biomass to the leaf is insufficient for full expansion at the maximum permitted specific leaf area (SLA; the SLA range is not considered G-specific). Marin & Jones (2014) developed a simple process-based model of sugarcane growth, which uses a quadratic function of leaf number to vary SLA between 6 (units not specified) at crop start, a peak of 9.3 at 15 leaves, and 4 after 35 leaves have appeared. These are fixed due to "a lack of sufficient evidence of how other factors affect SLA". In both their model and APSIM-Sugar, SLA is not considered a G-specific parameter. Limited published SLA data for sugarcane are available (Sebastião de Oliveira Maia Júnior et al., 2019; Terauchi and Matsuoka, 2001; Venkataramana et al., 1984). Robertson et al. (1998) reported SLA of 70-130 cm^2/g under field conditions, while in a controlled environment facility SLA decreased consistently from ≈ 160 cm^2/g at crop start to between 50 and 100 cm^2/g at 25-40 leaf age depending on temperature treatment (26 °C produced the most leaves and highest SLA; 14 °C the fewest leaves and lowest SLA). The APSIM-Sugar model

permits SLA in the range 85-115 cm^2/g. In Maize, SLA has been found to decrease with cumulative thermal time (Zhou et al., 2020).

Canesim, Mosicas and QCane simulate the canopy directly, rather than dividing canopy development into separate shoot population and leaf growth/senescence processes. Canesim simulates canopy cover (fractional interception of photosynthetically-active radiation) as a function of thermal time and water stress, while Mosicas estimates leaf area index in a similar manner. QCane estimates leaf area index as a function of biomass partitioned to leaves and specific leaf area.

Singels et al. (2005c) found that for crops grown under irrigation, fully-expanded leaf size was correlated with the green leaf area index of the canopy at the time of leaf appearance. They hypothesised that leaf size is determined by source strength: when the canopy is larger, source strength is higher (all else being equal). Gs with larger initial leaf size would then develop canopy cover faster; Sinclair et al. (2004) reached a similar conclusion, adding that early leaf size could be used as a phenotypic marker for high yield potential. Leaf size was consistently larger for summer-start crops compared to winter-start crops, perhaps as a result of higher temperatures and solar radiation during the partial canopy period in the summer-start crop. These relationships were consistent across the two Gs tested (NCo376 and CP66/1043), although leaf sizes were different between the Gs. Singels et al. (2005c) conclude by suggesting that the absence of coupling between biomass accumulation and canopy growth processes in sugarcane models is a shortcoming that hinders application of the models for optimising radiation interception. Zhou et al. (Zhou et al., 2003) however noted in a study of four southern-African cultivars (NCo376, N14, ZN6 and ZN7) grown under irrigation in Zimbabwe that tiller development and senescence traits were more important in determining PAR interception than leaf size and canopy architecture.

Relative growth rate (rate of mass increase per unit time, expressed as a fraction of the existing mass) is considered an important parameter for characterizing plant growth (Hoffmann and Poorter, 2002; Hunt, 1982). A similar concept could be applied to characterize dynamics of green leaf area index (GLAI, m^2/m^2). GLAI increases with net availability of biomass (from photosynthesis); but increasing GLAI – particularly in a young crop with a relatively sparse canopy – also increases radiation interception and therefore the biomass available to support leaf expansion.

In a young sugarcane crop, the relative rate of growth of leaf area is high; this rate slows as light competition restricts the marginal photosynthetic benefit to additional leaf area (as the crop nears canopy closure), and then slows further as the crop starts partitioning above-ground biomass to stalks in addition to leaf blades. The daily increase in leaf area under well-watered conditions will be restricted by (i) temperature and (ii) availability of source photo-assimilate (Singels et al., 2005b).

2.5.7 Leaf senescence

Leaf senescence in APSIM-Sugar is driven by frost, light competition, maximum number of green leaves, and water stress. The daily reduction in GLAI due to frost starts at 10% loss if the air temperature falls below 0 °C to 100% loss at -5 °C. The light competition effect is based on the extent to which current GLAI exceeds a threshold value at which light-induced senescence starts. Parameters controlling these drivers of senescence are not G-specific. The green leaf number restriction limits the number of green leaves on the shoot to a maximum number, which is also the approach to leaf senescence used in DSSAT-Canegro. The maximum number of

green leaves per shoot is specified as a trait parameter in DSSAT-Canegro and APSIM-Sugar. The findings of Singels & Smit (2009) suggest that maximum green leaf number is not stable, varying with row-spacing. Maximum green leaf number is probably also determined by canopy architecture, so is not independent of the radiation extinction coefficient (discussed in Section 2.5.8). DSSAT-Canegro does not simulate frost or light competition effects on leaf senescence. The QCANE model simulates layered canopy radiation interception and leaf senescence is based on shading: when the photosynthesis rate of leaves in a canopy layer decrease below the maintenance respiration rate, leaves in that canopy layer are senesced by an amount proportional to the daily increase in stalk mass. This reflects the observation that green leaf number is determined by radiation intensity (more green leaves in summer) and light penetration into the canopy (more green leaves on field edges). The linkage to stalk mass increment is based on the concept that leaf senescence occurs when its internode has matured; the faster the cane grows, the faster the rate of senescence. (Liu and Bull, 2001).

The Canesim model does not simulate senescence as such: green canopy cover reaches a maximum value and does not increase further, although it can decrease transiently in response to water stress (Singels and Paraskevepoulos, 2017).

A significant G effect controlling the acceleration of leaf senescence rate under water stressed conditions has been reported (Smit and Singels, 2006). Faster leaf senescence in cultivar N22 (compared with NCo376) was coordinated with faster tiller senescence as well. The G effects of water stress on tiller senescence rate and leaf senescence rate are independent of each other in the DSSAT-Canegro and APSIM sugar models, but it is likely that a single trait input parameter for controlling canopy senescence response to water stress would be appropriate.

2.5.8 Radiation interception, photosynthesis and respiration

The interception of radiation (either global or photosynthetically-active) drives photosynthesis, directly or indirectly, in all sugarcane models. Radiation interception is either calculated directly (as in the Canesim and AquaCrop models) or is estimated, using a Beer's Law equation and a radiation extinction coefficient (Ke), from GLAI. The radiation extinction coefficient is G-specific in DSSAT-Canegro (minimum and maximum values linked to leaf number) and Mosicas, and is fixed in APSIM-Sugar. Ke reflects the canopy architecture (leaf area and angle, specifically (Luo et al., 2013)), and has an influence on radiation interception especially during the partial canopy phase. Ke values for PAR are higher than for SRAD; 0.65 (for PAR) provides an equivalent interception fraction to 0.48 (for global radiation) (Jovanovic and Annandale, 1998), although it is noted that this equivalence is limited to interception fraction and not to the conversion of intercepted radiation in biomass.

Dias et al. (2020) calibrated APSIM-Sugar Ke values (on a global shortwave radiation (SRAD) basis) for 27 Brazilian sugarcane varieties at two sites. Ke values differed (within narrow ranges, 0.45-0.59 and 0.55-0.77) at each of the sites, and were very well correlated between the sites (R^2 = 0.99), but were about 20% higher at the second site compared to the first, indicating a significant E effect on Ke (not considered in any sugarcane models). Luo et al. (2013) reported Ke values for 17 Chinese cultivars grown in an irrigated plant crop in south-eastern China, and found no significant differences between the Gs; Ke was found to be ≈0.65 overall. De Silva and de Costa (2012) reported Ke values in the range 0.22-0.28 for eight cultivars grown as a plant crop under irrigation in Sri Lanka, corresponding with leaf angles in the range 66-76°.

The water-stressed rainfed treatment had higher leaf angles (presumably due to shorter leaves as a consequence of water stress) and higher Ke values (0.31-0.47). Water stress effects on Ke have not been taken into account with sugarcane models. Zhi et al. (2021) found that leaf angle in sorghum has G control, and explains 36% of variation in Ke. Ke appears to be somewhat dynamic, which may play a role in determining GxE interactions in yield. Default Ke values for different models are: DSSAT-Canegro, 0.58-0.86 (PAR basis, cultivar NCo376); Mosicas, 0.48 (PAR basis, cultivar R570); APSIM-Sugar (SRAD basis, 0.38). Although DSSAT-Canegro and Mosicas models permit G-specific Ke values, none of the models attempt any kind of dynamic prediction of Ke. Ke was not measured or calculated in the experiments reported-on in this thesis. It should be noted that errors in Ke values can be compensated (in a misleading manner) by equivalent errors in specific biomass accumulation rates (see following paragraph). Also, mistakenly recording interception of radiation by stalks and dead leaves, with little or no photosynthetic capacity, can lead to unintentional variation in reported Ke values.

The simulation of photosynthesis is conducted broadly in two ways: radiation-use efficiency (RUE) and transpiration efficiency (TE) approaches. In the RUE approach, photosynthesis rate (P, g/m^2/d) is a function of intercepted radiation intensity:

$$P = PAR * FIPAR * RUE_d \quad (2\text{-}2)$$

where PAR is incident daily photosynthetically-active radiation (MJ/m^2/d), $FIPAR$ is the daily fractional interception of PAR, and RUE_d is the daily radiation use efficiency (g/MJ)), and factors that affect conversion efficiencies:

$$RUE_d = F_T * F_W * RUE_0 \quad (2\text{-}3)$$

where F_T and F_W reflect the temperature and water influences on photosynthesis, with values between 0 (fully-limiting) and 1 (not-limiting), and RUE_o is a maximum theoretical gross photosynthetic conversion efficiency parameter defined for ideal conditions of temperature, water and nutrient availability for a healthy fully-canopied crop (Inman-Bamber, 1991; Inman-Bamber and Thompson, 1989; Jones and Singels, 2018; Singels et al., 2005b).

Considering the TE approach, biomass accumulation is function of plant water uptake and atmospheric vapour pressure deficit (VPD_d, kPa):

$$P = T_d * \frac{TE_o}{VPD_d} \quad (2\text{-}4)$$

where T_d is daily transpiration rate (mm/d) and TE_o is a transpiration efficiency coefficient (g/kg/kPa) (Keating et al., 1999; Sinclair, 2012).

DSSAT-Canegro, APSIM-Sugar, Mosicas, Canesim, QCane, and Casupro make use of the RUE approach to calculate photosynthesis rates. APSIM-Sugar uses the TE approach to calculate transpiration rate (see Section 2.5.11). The AquaCrop model uses a TE-type approach (Vanuytrecht et al., 2014), basing biomass accumulation rate on a water productivity coefficient (effectively TE_o) normalised with respect to reference potential evapotranspiration (a proxy for VPD_d) and atmospheric CO_2 content.

Inman-Bamber et al. (2012) found that TE_o was likely to be genetically-controlled and strongly determined environmental fitness (in terms of yield at harvest), but that high TE_o values were favourable across most mild- to moderately water-stressed environments. This suggests that a high TE_o value would contribute to broad adaptability for cultivars bred for a range of rainfed environments. The sugarcane TE

parameter in APSIM-Sugar was not (as at v7.10) considered G-specific, although Inman-Bamber et al. (2016) described changes to APSIM-Sugar including G-specific TE_o. Basnayake (2015) found G variation in canopy-level stomatal conductance. Jackson et al. (2014) found considerable G variation in TE_o, and presented a framework for including TE as a basis for selection in sugarcane breeding programmes. AquaCrop considers the water productivity parameter "conservative" (not G-specific).

RUE_o values are highly dependent on specific definitions for different models. DSSAT-Canegro v4.5 (Singels et al., 2008) assumes a maximum gross (i.e. before losses to respiration) theoretical value of 9.9 g of dry biomass per MJ of photosynthetically-active radiation (PAR), while the DSSAT-Canegro v4.6 (Jones and Singels, 2018), which has slightly different respiration and photosynthesis algorithms, assumes RUE_o values of approximately 5 g/MJ. Canesim uses an RUE_o parameter that reflects net above-ground biomass accumulation rate per unit of intercepted PAR (G-specific). APSIM-Sugar assumes 1.8 g/MJ for plant crops and 1.6 g/MJ for ratoon crops, defined in terms of above-ground biomass, after deductions by growth and maintenance respiration, on the basis of global shortwave radiation (SRAD, approximately double PAR). RUE_o is G-specific in DSSAT-Canegro, but not in APSIM-Sugar.

Anderson et al. (2015) reported seasonal average above-ground SRAD RUE of 1.15-1.24 g/MJ for two fully-irrigated commercial plant crops (cultivar H65-7052) grown over 24-month period in Hawaii. Biomass accumulation in this case was determined using eddy-covariance flux towers, which can have some limitations, e.g. difficulty in accounting for night-time respiration. SRAD RUE was reported for an irrigated plant crop in Sri Lanka in the range 1.63-2.09 g/MJ, for eight cultivars (six Sri Lanka, one Indian and one Mauritian) (De Silva and De Costa, 2012). RUE was reported to be consistent (i.e. no G differences) across a set of 13 cultivars (seven Argentinian, five USA and one South African) grown in rainfed plant (RUE ≈ 1.0 g/MJ) and first ratoon (RUE ≈ 1.7 g/MJ) crops, although water-use efficiencies differed (Acreche, 2017). Dias et al. (2020) calibrated APSIM-Sugar for 27 Brazilian varieties without changing RUE_o, suggesting that RUE is not G-specific. Marin et al. (2011) and Coelho et al. (2020) calibrated RUE_o values for several Brazilian cultivars nearly 25-50% higher than NCo376 values for the DSSAT-Canegro model. Singels & Inman-Bamber (2011) argued in favour of G-specific RUE_o, and recommended this for crop modelling, based on a glasshouse experiment with four clones. Small (≈1-8%, statistical significance not shown) differences in RUE were reported by Donaldson (2009) for varieties N25, N26 and NCo376. Hoffman et al. (2016) demonstrated that leaf-level photosynthetic efficiency rankings amongst a set of 14 sugarcane cultivars grown in a pot trial were robust across sampling events. They also showed that the ranking of simulated yield of a field trial, using Canesim calibrated with (canopy-level) radiation use efficiency (RUE) values based on the leaf-level photosynthetic rate values determined in the pot trial, closely matched observed yield rankings. This suggests that RUE may be G-specific.

DSSAT-Canegro, QCane and Casupro simulate growth (Rg) and maintenance respiration (Rm), where RUE reflects gross rather than net photosynthetic efficiency. Different versions of DSSAT-Canegro have assumed: a fixed Rm rate of 0.004 g/g total crop dry mass and a Rg rate of 0.242 g/g (Inman-Bamber, 1991); Rm rate as a function of temperature and total crop biomass (Singels and Bezuidenhout, 2002); and Rm calculated using different reference rates to each respirable C pool (leaves, roots and sucrose) (Jones and Singels, 2018). Gifford (2003) argued against the concept of Rm,

preferring that all respiration be calculated as part of growth respiration. This is corroborated by Anderson et al. (2015), who found that average observed RUE did not decrease over the course of 24-month plant crops grown in Hawaii, where temperatures varied in the range 20-27 °C throughout the growing season. On the other hand, Dias et al. (2019) developed a 'high biomass slowdown' feature for APSIM-Sugar, which reduces net relative biomass accumulation rates linearly from 100% at 20 leaves to 50% from 50 leaves onward, to accommodate observed productivity dynamics at high-potential sites in Brazil. The empirical nature of this slow-down feature is acknowledged by the authors, but reflects the 'reduced growth phenomenon" (RGP) in high-potential sugarcane (Park et al., 2005). Causes of RGP are not yet fully determined, although lodging, declining leaf N content, and negative feedback inhibition of photosynthesis by high stalk sucrose content are suspected (Van Heerden et al., 2010).

2.5.9 Stalk growth

The date of onset of stalk growth (OSG) in sugarcane crop models is generally determined for the single purpose of partitioning above-ground biomass increments towards stalks (Inman-Bamber, 1991; Keating et al., 1999; Martiné and Todoroff, 2004). The transition from the tillering to stalk growth phases is simulated in DSSAT-Canegro and APSIM-Sugar as a single event, after a predetermined G-specific thermal time period has elapsed since primary shoot emergence. This is supported by Singels & Inman-Bamber (2011), who showed that internode elongation, the appearance of which is coordinated with leaf appearance, occurs after the appearance of 4-5 immature leaves and 4-5 fully-expanded leaves; and leaf appearance appears to be very well-predicted by thermal time accumulation (Bonnett, 1998; Inman-Bamber, 1994). In Mosicas, the crop transitions from tillering to stalk growth as a function of aerial dry mass accumulation above a threshold value.

In unstressed crops, OSG occurs at approximately 70% inter-row FIPAR (measured across the space between two adjacent cane rows), 90% intra-row FIPAR (measured underneath the canopy, between the outer edges of the canopy column), and coincides with peak shoot population (Allison et al., 2007; Inman-Bamber, 1994; Singels and Smit, 2009). No sugarcane models simulate OSG this way. Cessation of tillering has been attributed to radiation interception in sorghum (Kim et al., 2010). Singels & Smit (2009) showed that tillering ceases at 90% intra-row PAR interception, and the date of cessation (i.e the date of peak shoot population) was insensitive to row-spacing. OSG date was not reported, directly, but may be imputed from the SDM/ADM fractions; these are consistent with the hypothesis that OSG is linked to the date of peak population and 90% intra-row PAR interception. The authors also showed that 90% intra-row PAR interception coincided with the achievement of maximum number of green leaves, implying that leaf/shoot senescence also starts at or near to OSG. Inman-Bamber (1994) reported that the phyllocron interval lengthened for cultivar NCo376 after 14 leaves had appeared, at a total thermal time age very similar to the thermal time delay from primary shoot emergence to OSG, suggesting a physiological linkage. It is speculated that this could be a response to increased sink strength from elongating stalks following OSG, and therefore lower assimilate availability for leaf growth. Bonnett (1998) reported continually variable PI but with a similar shape to the broken stick – short PIs when young, long when old – implying a gradual transition from tillering to stalk growth rather an abrupt event.

Stalk growth itself receives little attention in sugarcane models. Singels (2014) does not even list stalk growth as a process in his comparison of sugarcane models. Stalks are considered a biomass pool, so receive attention in the biomass partitioning processes of sugarcane models. DSSAT-Canegro simulates stalk height as an output (regulated by a genotype parameter), but this has no bearing on yield simulations. The GTP-Canegro model, however, simulates source-sink processes at a stalk level considering structural demands (sink strength) for leaf and stalk fibre based on elongation rates in response to temperature and water status combined with specific densities, and partitions biomass accordingly. Ngobese (2018) reported that stalk elongation rate G rankings were consistent across Es, suggesting that this is a stable G trait, while Hoffman (2017) reported a wide and statistically significant genetic range in the parameter SERo, defined as the stalk elongation rate per unit thermal time, for a pot phenotyping trial with 14 southern African cultivars.

2.5.10 Biomass partitioning

Biomass partitioning is simulated in most sugarcane models via allometric fractions (Singels, 2014). The fractions are either fixed for different phenological phases (O'Leary, 2000), or described in terms of environmental conditions (e.g. low temperatures favour partitioning to stalks rather than leaves in the Canegro and QCane models). This is at odds with reports that base temperatures for leaf expansion are lower than those for stalk expansion. APSIM-Sugar and Casupro consider structural demands (sink strengths) in some biomass partitioning processes.

The DSSAT-Canegro model sucrose partitioning model (Singels and Bezuidenhout, 2002), which is also used in Canesim (Singels and Paraskevepoulos, 2017) uses an empirical source:sink approach. Under irrigated conditions, a temperature factor *FT* is used to determine the daily partitioning between sucrose and fibre+hexose, with a G-specific model input temperature sensitivity parameter. When air temperatures decrease below this parameter value (calibrated as 25 °C for NCo376), more stalk biomass to be partitioned to sucrose than fibre and hexose; and *vice versa*. Earlier versions of DSSAT-Canegro relied on empirical relationships (of stalk dry mass and day of year) to predict sucrose yield (Inman-Bamber et al., 2002).

APSIM-Sugar partitions biomass to sucrose based on water, nitrogen and temperature stress factors (O'Leary, 2000; Singels and Inman-Bamber, 2011), although if the biomass allocation to leaves exceeds the maximum sink strength (determined by the daily change in green leaf area and the minimum specific leaf area parameter value), excess biomass allocated to leaves is transferred to the sucrose pool (Keating et al., 1999); nevertheless, sucrose accumulation in APSIM-Sugar is primarily source-driven.

The GTP-Canegro model calculates sink strengths for leaf and stalk fibre and necessary hexose demands, and bases the partitioning of above-ground biomass on these; a source deficit results in a proportional decrease in growth rates of both stalks and leaves, and excess source is stored as sucrose, subject to sucrose storage capacity being available. The GTP-Canegro model has not proceeded beyond prototype stage, and has some clear shortcomings which need to be addressed. This source-sink approach to biomass partitioning nevertheless shows promise as a physiologically-sound algorithm that is anticipated to be useful for accurate exploration of GxE interactions, with a key challenge being the determination of sink strengths for expanding stalks. Inman-Bamber et al. (2010) found, in a glasshouse experiment, that sucrose content and mass were higher in the cool (15-25 °C) treatment compared to a warm (24-31 °C) treatment and attributed this behaviour to differential structural sink

strengths. Singels & Inman-Bamber (2011) compared four clones (two high-sucrose and two low-sucrose Gs) and showed that biomass partitioning to sucrose in the stalks appears to be a consequence of source and sink strengths: generally, the low-sucrose Gs partitioned more biomass to leaves, leaving less available for sucrose; additionally, high-sucrose Gs had shorter leaf phyllocron intervals, meaning that structural internode development was completed faster and so allowed sucrose accumulation to start sooner. This is supported in broad terms by Lingle (Lingle, 1999) who reported that at a metabolic level, sucrose content was related to sucrose synthase activity (in the cleavage direction) during internode elongation, and declined thereafter.

2.5.11 Water relations

APSIM-Sugar uses the transpiration efficiency (TE) approach to calculate water uptake: transpiration is calculated from daily biomass accumulation rate, TE_o and VPD_d, and limited by root water supply. DSSAT-Canegro and Casupro use FAO-56 or Priestley-Taylor reference evapotranspiration methods, and the Canesim model reads in daily ET values calculated using a Penman-Monteith equation parameterised for sugarcane. DSSAT-Canegro, APSIM-Sugar, Casupro, and Mosicas use CERES-style (Jones et al., 1986) 'tipping bucket' layered soil water balances. Early versions of Canesim had a single-layered water balance, while the current version has a multi-layered soil profile (Singels and Paraskevepoulos, 2017).

Water stress is either calculated as a function of leaf and soil water potentials, or some analogue of these – soil water content and vapour pressure deficit or reference evaporation rate, or even just soil water content by itself. There is little scope for easily capturing G-specific water uptake / stress characteristics in existing models. The DSSAT-Canegro v4.5 model permits G-specific values for the fraction of biomass allocated to roots; other parameters (such as root length per unit of root biomass, root conductivity, and the CERES water uptake model parameter values) are considered 'species' parameters. The version of APSIM published by Inman-Bamber et al. (2016) allows G-specific specification of parameters that affect the TE response to water stress, atmospheric CO_2 concentration, root hydraulic conductivity, and maximum hourly transpiration rate. Singels et al. (2010) identified shortcomings with CERES-based water uptake model in DSSAT-Canegro, and made recommendations for improved approaches. A version of the DSSAT-Canegro model[1] (Jones and Singels, 2018) uses the AquaCrop model's concept of relative soil depletion concept and reference evaporative demand to calculate soil water satisfaction indices for photosynthesis and expansive growth; a G parameter controls the relative soil water depletion fraction at which photosynthesis stress starts, at a reference evaporative demand of 5 mm/day. Although empirical in nature, this approach permits very easy per-G specification of water stress sensitivity; it is an integrated measure, however, summarising the effects of a set of lower-level processes, and may be subject to GxE interactions in itself unless these are generally genetically-coordinated. Zhao et al. (2017) however reported that main G effects were much stronger that GxE interaction effects when comparing Gs under irrigated and mild/moderately water-stressed Es.

2.5.12 Soil temperature

The DSSAT system calculates soil temperature using an approach based on CERES-Maize (Jones et al., 1986), while APSIM uses a more sophisticated approach, whereby

[1] The AquaCrop water uptake feature in DSSAT-Canegro is not yet publicly available.

heat fluxes between the atmosphere and soil surface, and between soil layers, are solved numerically.

2.5.13 Nutrition

DSSAT-Canegro, Canesim, Mosicas and Casupro do not simulate nitrogen dynamics. APSIM-Sugar uses the CERES-N approach (Godwin and Jones, 1991), with plant component-specific critical N-contents considered for modifying process rates in response to N. A similar approach was implemented in DSSAT-Canegro (van der Laan et al., 2010) but has not been released publicly. Critical N content values can be specified for different genotypes, to capture genetic differences in N-use dynamics. N can be translocated from senescing leaves. QCANE uses an empirical approach to modelling N effects.

Accurate, mechanistic simulation of N is probably very important for accurately simulating GxExM interactions in final yield, because of G variation in N-use efficiency, water uptake (affecting N uptake via mass flow) and root growth patterns. Improvements to N modelling in sugarcane is limited by availability of suitable data.

2.5.14 Discussion and key challenges

Sugarcane CGMs have several points in common: germination is driven by thermal time calculated from air temperature; canopy development is (directly or indirectly) a function of thermal time and water status; the onset of stalk elongation is not mechanistically linked to canopy development; biomass accumulation is driven by a radiation use efficiency approach (at least under unstressed conditions) with sensitivity to radiation interception fraction, temperature and water status; biomass is partitioned on the basis of predetermined allometric fractions or prescribed responses to temperature and water stress.

The simulation approaches followed by the different models have their strengths and weaknesses. DSSAT-Canegro is strong on tiller dynamics and phenology, but leaf size is not linked to carbon availability; APSIM links carbon dynamics to canopy development by calculating leaf sink strength and potentially limiting leaf expansion to biomass partitioned to the canopy, but has a weaker simulation of tillering. Canesim, Mosicas and QCane provide elegantly simple approaches to simulating canopy cover / leaf area index, but these approaches are less compatible with dynamic, competitive source-sink biomass partitioning, which is likely to be necessary for a model that is to be used for exploring and understanding GxE interactions in sugarcane. APSIM has a mature nitrogen model, and is therefore able to accommodate N-use efficiency traits in unravelling GxE interactions. QCANE simulates radiation interception, photosynthesis and respiration in a layered fashion and so has very sophisticated prediction of green leaf number and leaf area index, but relies heavily on empirical relationships.

Leaf and tiller senescence appear to be poorly understood under unstressed conditions, with evidence of G and potentially GxE interaction effects; other than QCANE, sugarcane CGMs rely on simplistic empirical descriptions for these. Radiation extinction coefficients (Ke) appear to vary between Gs and Es, but there has been little effort to account for these in model applications. No models attempt to predict Ke dynamically. Despite evidence that stalk elongation rate is a stable G trait, sink strengths from stalk elongation have not been used to drive biomass partitioning in mainstream sugarcane CGMs. Indeed, the use of source:sink relations in determining biomass partitioning has been very limited. APSIM-Sugar considers

source and sink strengths for leaf expansion. The DSSAT-Canegro model simulates sucrose partitioning using an approach based on source:sink principles, but ignore sink strengths from the canopy and has independent temperature parameters from those controlling stalk elongation. These omissions, in the DSSAT-Canegro and Mosicas models, limit their biological realism. None of the models consider tillering a growth process, or link onset of stalk growth to canopy cover. Specification of G controls over water stress sensitivity in crop models is or has been generally difficult: partly because the model user cannot change the values all of the relevant parameters in a G-specific way, and partly because water stress itself is a complex outcome of many traits controlling canopy development, root water uptake and stomatal control, with strong interactions with the hydraulic characteristics of the soil itseld. A recent development in the DSSAT-Canegro model allows the user to modify a G-specific photosynthesis reference soil moisture depletion threshold parameter (Jones and Singels, 2018)). A relatively recent change to the APSIM-Sugar model also allows a G-specific TE parameter values (Inman-Bamber et al., 2016)). These changes make crop model-based exploration of G control over water stress responses more accessible.

Recommendations from a sugarcane trait modelling workshop held at SASRI in 2017 (Singels, 2017) were that crop modelling to support breeding ought to focus on transpiration efficiency, radiation use efficiency, stomatal conductance and canopy development.

Considering these, and the recommended characteristics of CGMs for supporting plant breeding, the following areas should be explored for potentially improving sugarcane crop growth models for such applications for irrigated sugarcane:

- Assessing the value of transpiration efficiency as an explanatory trait for rainfed production scenarios.
- Assessing the role of leaf-level stomatal conductance. For irrigated cane, it might be informative to relate directly to biomass yields. For rainfed production scenarios, it might be more useful to include a physiological photosynthesis model (von Caemmerer, 2021), capable of upscaling from leaf stomatal conductance to canopy conductance and then predicting photosynthesis and transpiration rates, into an existing sugarcane crop model to explore the impacts of this trait in its interaction with water stress.
- Exploring maximum radiation use efficiency as a trait for explaining variation in biomass yields.
- Quantifying genotypic influences of canopy development (leaf and tiller development and senescence characteristics) and linking the development of canopy components with biomass accumulation.
- Ensuring that model parameters have clear physiological meanings that reflect independent genotypic traits whose values are stable across Es. This may ease phenotyping, and may improve the chances of linking back to QTLs.
- Replacing allometric biomass partitioning fractions with dynamic emergent consequences of source and structural sink strengths.
- Linking the onset of stalk elongation with radiation interception.
- Exploration of tillering as a growth process in addition to a developmental process, responsive to source:sink ratio.

2.6 Summary and Conclusion

The past 20 years has seen unprecedented development of plant breeding techniques, driven by rapid advances in and decreasing costs of genetic sequencing, phenotyping and modelling. Breeding programmes for many crops are now based on statistical prediction from genetic sequence data rather than traditional evaluation and selection. Value from including process-based crop growth models into breeding programmes has been clearly demonstrated for crops such as maize, rice and sorghum. Development of model-assisted breeding in sugarcane has however lagged behind due to its genetic complexity; sugarcane breeding programmes around the world still rely on traditional crossing and selection approaches where the development of new cultivars takes 10-14 years, requires large areas of land, is labour-intensive, and is practically limited to development of cultivars for broadly-defined Es. Nevertheless, considering the progress made in other crops, the application of crop modelling in sugarcane breeding has potential for increasing the cost-effectiveness of sugarcane breeding, shortening the time from crossing to release, and possibly developing cultivars bred for niche environments.

Process-based crop growth models (CGMs) that operate in biologically realistic ways represent crop physiology mechanisms. These can predict non-linear G and GxE interaction effects as emergent consequences of process-level responses to environmental drivers. CGMs can be used to translate low-level (sub-) trait phenotypes into complex phenotypic outcomes, such as yield, across the target population of environments (TPE). Such models can be used for environmental characterisation to improve the efficiency of breeding by weighting environments (Es) during selection to be more representative of the TPE. They can also be used to identify breeding priorities or even ideal low-level phenotypic trait combinations. If molecular markers (Quantitative Trait Loci, QTL) in genetic sequence data can be statistically associated with low-level phenotypes, CGMs can predict yield outcomes from genotypic data, on which selection can be based.

CGMs to support model-assisted breeding need to have simple trait parameters that are strongly linked to genetics and are stable across Es, and should predict higher level, complex trait phenotypes as the emergent consequences of low-level processes responding to environmental drivers. Such models should be biologically robust, realistically emulating biological processes rather than simply describing outcomes. Models ought to be as appropriately simple, “parsimonious”, as possible while still being fit for purpose. Users must be confident in the abilities of the CGM to account for G and E differences and predict GxE interaction effects with sufficient accuracy.

Several mature, process-based sugarcane models have been developed and are potentially available to breeders. Many of these models – APSIM-Sugar, DSSAT-Canegro and Mosicas, for example – can be used for environmental characterisation to improve the efficiency of selection. Although these models provide support for simulating different cultivars by being able to read G-specific model input parameters, they have not been comprehensively tested in a multi-environment trial (MET), and understanding their strengths and weaknesses with respect to predicting G and GxE effects needs to be determined before they can be confidently applied to trait dissection and phenotypic prediction in the TPE.

Recommendations for addressing identified weaknesses or gaps include replacing allometric biomass partitioning fractions with dynamic emergent consequences of source and structural sink strengths, linking canopy development (leaf appearance and

growth, and tillering) with biomass accumulation, linking the onset of stalk elongation with radiation interception, and treating tillering as growth process rather than simply a phenological developmental process. A sugarcane model that includes these features would likely operate in more biologically-realistic ways. With appropriate and realistic calibration, such a model may be able to predict GxE interaction effects sufficiently accurately to support sugarcane breeding, via: (1) more accurate environmental characterisation, (2) trait dissection to understand causes of GxE interactions, (3) development of robust linkages with QTL, and (4) yield prediction in the TPE from genotypic data (i.e. derived from genetic sequence analysis) or phenotypic data (i.e. derived from measurements of plant dimensions).

Implementing such a model is expected to achieve the broad objective of developing an improved sugarcane crop growth model capable of accurate prediction of GxE interaction effects on yield; this is presented in Chapter 5. The necessity of doing so, however, requires that existing models be evaluated first; this is described in Chapter 4. Existing models represent collections of process-level simulation approaches, and the performance and appropriateness of individual processes for predicting GxE outcomes might vary. An evaluation of process-level concepts and input parameter definitions, from existing models, for their abilities to capture E, G and GxE interaction effects, also needs to be conducted; this analysis is presented in Chapter 3. Finally, the value of the new model needs to be demonstrated, particularly with an application that goes beyond the capabilities of existing CGMs; such a case study is described in Chapter 6.

3. EXPLORING PROCESS-LEVEL GENOTYPIC AND ENVIRONMENTAL EFFECTS ON SUGARCANE YIELD USING AN INTERNATIONAL EXPERIMENTAL DATASET

Abstract

Crop improvement aims to produce high yielding genotypes for target environments. Crop models simulate yield formation as the outcome of a series of low-level processes, driven by environmental (E) variables and regulated by genetic (G) traits. There is potential for crop models to aid sugarcane breeding, by identifying desirable genetic traits for target environments. The objective of this study was to evaluate existing concepts of G and E control of plant processes for explaining crop development, growth and yield, using an international growth analysis dataset. Crop development, growth and yield were monitored in the plant and 1st ratoon crops for seven cultivars (N41, R570, CP88-1762, HoCP96-540, Q183, ZN7 and NCo376) grown under well-watered conditions at La Mare (Reunion Island, France), Pongola (South Africa (RSA), Chiredzi (Zimbabwe), and Belle Glade (Florida, USA). Weather data were collected and environmental conditions characterized for each experiment. Derived process-level phenotypic parameters, based on concepts from four sugarcane growth simulation models (DSSAT-Canegro, Mosicas, APSIM-Sugar and Canesim), were calculated from observations and used to (1) evaluate current understanding of E drivers of sugarcane growth and development processes, and (2) identify and quantify G control at a process level. Final yields showed significant E and GxE variation; dry above-ground biomass and stalk yields were highest in La Mare and lowest in Pongola. Cultivar rankings in stalk dry mass for the common cultivars (N41, R570, CP88-1762) varied significantly between Es. Significant E variation in phenotypic parameters describing germination, tillering and timing of the onset of stalk growth (OSG) revealed shortcomings in the underlying simulation concepts. Significant G variation was found for germination rate, leaf appearance rate and canopy development rate per unit thermal time (TT), and maximum radiation use efficiency, indicating strong G control of the associated underlying processes. Solar radiation was found to influence tillering rate per unit TT, and TT to OSG, challenging the current theory of TT as the sole driver of these processes. By explaining more of the E variation, more stable and accurate G-specific model parameters can be defined and evaluated. This is anticipated to lead to less GxE confounding of modelled processes, and hence crop models that are better-equipped for supporting sugarcane crop improvement.

Published as a journal article and corrigendum thereto:

Jones, M. R., Singels, A., Chinorumba, S., Patton, A., Poser, C., Singh, M., Martiné, J. F. F., Christina, M., Shine, J., Annandale, J., & Hammer, G. (2019). Exploring process-level genotypic and environmental effects on sugarcane yield using an international experimental dataset. *Field Crops Research, 244*, 107622. https://doi.org/10.1016/j.fcr.2019.107622

Jones, M.R., et al., 2021. Corrigendum to 'Exploring process-level genotypic and environmental effects on sugarcane yield using an international experimental dataset' [Field Crops Research, 244 (2019) 107622]. *Field Crops Research*, 260, 107986. https://doi.org/https://doi.org/10.1016/j.fcr.2020.107986

4. EVALUATING PROCESS-BASED SUGARCANE MODELS FOR SIMULATING GENOTYPIC AND ENVIRONMENTAL EFFECTS OBSERVED IN AN INTERNATIONAL DATASET

Abstract

Crop modelling has the potential to assist plant breeding by identifying favourable genotypic (G) traits for specific environments (Es). Sugarcane crop models have not been rigorously evaluated against a factorial GxE dataset. It is imperative that models are evaluated in this way before they are applied to plant breeding problems. Our objectives were to (1) calibrate, (2) assess, and (3) identify weaknesses and recommend improvements to, three sugarcane models, DSSAT-Canegro, Mosicas and APSIM-Sugar, in relation to their predictions of observed E, G and GxE interaction effects in response to abiotic factors (temperature and solar radiation). Data from an international GxE growth analysis trial were used; these consisted of five irrigated experiments at four sites (Belle Glade, Florida, USA; Chiredzi, Zimbabwe; La Mare, Reunion Island; and Pongola, South Africa), with cultivars N41, R570 and CP88-1762. Observed G and E effects on final above-ground dry mass (ADM) yields were explained in terms of seasonal radiation interception (FIPARa) and seasonal average radiation use efficiency (RUEa). Calibration was undertaken where possible by translating phenotypic parameters derived from observations into model input trait parameter values representing genetic traits. E and G effects on FIPARa were generally simulated satisfactorily, while GxE interaction effects were poorly predicted due to inadequate responses to temperature. E, G and GxE effects on RUEa were poorly predicted by all models, although data shortcomings (arising from uncertainty regarding date of primary shoot emergence and impacts of lodging) prevented us from making strong conclusions in this regard. Models accurately predicted G differences in RUEa during mid-season biomass sampling periods where data confidence was greater. Although the models were able to predict final ADM yield per G and per E reasonably well, none of the models predicted GxE interaction effects well. All models also under-estimated the variation in RUEa and ADM. Recommendations for experimental protocols for exploring RUEa are made. Our key recommendations for future work to improve models for sugarcane breeding applications are to explore G-specific thermal time base temperatures for germination and canopy development processes, and to improve linkages between carbon availability and canopy development.

Published as a journal article:

Jones, M. R., Singels, A., Chinorumba, S., Poser, C., Christina, M., Shine, J., Annandale, J., & Hammer, G. L. (2021). Evaluating process-based sugarcane models for simulating genotypic and environmental effects observed in an international dataset. *Field Crops Research*, *260*(April 2020), 107983. https://doi.org/https://doi.org/10.1016/j.fcr.2020.107983

5. A SUGARCANE MODEL FOR IMPROVED SIMULATION OF GXE INTERACTIONS OF CANOPY COVER AND BIOMASS ACCUMULATION

TABLE OF CONTENTS

5.1 Introduction 43
5.2 Model description 44
5.2.1 Introduction 44
5.2.2 Overview 44
5.2.3 Phenological phases 52
5.2.4 Temperature control factors 55
5.2.5 Canopy expansion 56
5.2.6 Biomass accumulation (photosynthesis) 60
5.2.7 Biomass partitioning to plant components 61
5.3 Methodology 65
5.3.1 Overview and general considerations 65
5.3.2 Models, datasets and model assessment scenarios 69
5.3.3 Statistical measures of model performance 71
5.3.4 Determination of dates of emergence 71
5.3.5 CaneGEM trait parameter calibration with NCo376 data 73
5.3.6 Model validation with NCo376 data 74
5.3.7 CaneGEM model calibration for ICSM IGEP cultivars (R570, N41 and CP88-1762) 75
5.3.8 DSSAT-Canegro calibration for ICSM IGEP cultivars 76
5.3.9 Evaluation of E, G and GxE interaction effects 76
5.4 Results 77
5.4.1 Introduction 77
5.4.2 NCo376 experimental data 79
5.4.3 ICSM IGEP experimental data 91
5.5 Discussion 100
5.5.1 Novel canopy development and biomass partitioning algorithms are sound, and conventionally-assessed performance of the CaneGEM model is similar to that of DSSAT-Canegro 100
5.5.2 Simulation of GxE interaction effects on canopy development and biomass accumulation is improved in the CaneGEM model 101
5.5.3 GxE interactions for date of emergence drive GxE in canopy cover and biomass yields, but prediction of germination rate is inadequate 103

5.5.4 Differential temperature responses appear to drive remaining GxE interactions, and can be inferred from germination rate 104
5.5.5 Model structure and concepts 105
5.5.6 The CaneGEM model is well-suited for supporting plant breeding 107
5.5.7 Additional possible model applications.....108
5.5.8 Recommendations 108
5.6 Conclusion.....109

5.1 Introduction

Three widely used sugarcane crop growth models, DSSAT-Canegro, APSIM-Sugar, and Mosicas were reviewed in Chapter 4 (Jones et al., 2021) for their abilities to simulate E, G and GxE interaction effects on seasonal canopy development (fractional interception of photosynthetically-active radiation, *FIPARa*), radiation use efficiency (*RUEa*) and above-ground dry biomass accumulation (*ADMh*, t/ha), for a multiple-E, multiple-G dataset (Jones et al., 2019). This dataset was collected by the International Consortium for Sugarcane Modelling (ICSM) in the International Genotype-by-Environment Project (IGEP). While the models assessed gave satisfactory and similar performance for predicting E and G effects, they had essentially no ability to predict GxE interaction effects in that study.

Recommendations made in Chapter 4 for improving sugarcane models to support irrigated sugarcane breeding include: (1) simulating germination rate on the basis of simulated soil temperature, rather than observed air temperature, and (2) regulating the onset of stalk growth (*OSG*) according to radiation intensity in addition to temperature, described in Chapter 3 (Jones et al., 2019); as well as (3) regulating germination and canopy development with temperature factors based on G-specific cardinal temperatures and (4) including realistic linkages between carbon availability and canopy growth, as described in Chapter 4 (Jones et al., 2021).

The second part of the hypothesis stated in Section 1.3 asserted that crop simulation modelling capacity to support breeding of irrigated sugarcane could be enhanced by (if necessary) developing an improved (new or revised) sugarcane crop model to address identified weaknesses arising from evaluating existing models and their process-level concepts.

The overall aim of this phase of the work was to develop a parsimonious crop growth simulation model capable of improved predictions of GxE interaction effects on seasonal radiation interception (and dry biomass yields, by implication), that could be of value to irrigated sugarcane crop improvement.

The specific objectives were to:

1. Develop a new sugarcane model with the following features:
    a. an improved germination algorithm, making use of simulated soil temperatures;
    b. a carbon-linked canopy development algorithm, sensitive to temperature and radiation intensity,
    c. an *OSG* algorithm that is responsive to radiation intensity in addition to temperature, and

2. calibrate and validate the new model using existing detailed growth analysis experiment data for cultivar NCo376;
3. calibrate the new model with respect to canopy development for cultivars CP88-1762, N41 and R570 and evaluate its ability to predict G, E and GxE interaction effects on seasonal radiation interception, biomass and stalk mass using data from the ICSM IGEP experiments.

5.2 Model description

5.2.1 Introduction

The model description starts with an overview, followed by details of the following processes and calculations:

- Phenological phases
- Temperature control over plant processes
- Canopy (leaf) expansion and senescence
- Biomass accumulation (net photosynthesis)
- Biomass component growth

Genotype-specific model input parameters are termed “trait parameters” in this text.

5.2.2 Overview

A new sugarcane crop growth model was developed to address shortcomings identified in Jones et al. (2019) and Jones et al. (2021), to incorporate concepts identified in previous research (as outlined in Chapter 2), and to reflect features of models appropriate for model-assisted crop improvement (parsimony, biological realism and emergent complex phenotypes (Hammer et al., 2019b; Hammer and Jordan, 2007)). The model is named “CaneGEM” (sugarcane genotype-by-environment model).

The decision was made to implement the new concepts as a new model, rather than a modification to an existing model, for several reasons. Using an existing codebase might have biased the implementation of concepts towards those that operate more comfortably within that framework. In some cases the development and evaluation of new concepts would have been hindered by the need to disconnect or disable functionality in existing models in order to permit a new feature to operate. The R environment chosen for implementation integrates more easily with data analysis and visualisation tools compared with the Fortran-based codebases of the existing models considered.

The key features differentiating CaneGEM from DSSAT-Canegro, APSIM-Sugar and Mosicas are summarised as follows, with details provided in Section 5.2.3-5.2.7:

- **Canopy development is a source-limited growth process in addition to a developmental process**

  In the DSSAT-Canegro model, unstressed canopy development is determined by temperature. Tillering, leaf appearance and leaf elongation are driven prescriptively by temperature, and then combined to determine canopy cover each day.

  In the CaneGEM model, canopy cover expansion (leaf area growth) is determined by source and sink strengths. The canopy can potentially increase in size each day by a fraction of the existing size of the canopy – linking source

capacity with canopy development rates. Effective temperature (with G-specific base and other cardinal temperatures) regulates the potential growth rate.

Detailed developmental processes, such as the emergence and senescence of individual tillers and leaves, are not simulated (as they are in DSSAT-Canegro), in the interests of parsimony.

The size of the canopy depends on the availability of carbohydrates from photosynthesis. In this way, biomass accumulation and canopy development processes are linked, introducing realistic trade-offs between trait parameter values and greatly enhancing the biological realism of the model. (Note: this feature exists in APSIM-Sugar).

Canopy development rate is therefore an emergent complex trait, governed by lower-level, simpler trait parameters.

A simplified diagram of how the daily increase in green leaf area index is calculated in the CaneGEM model each day is shown in Figure 5-1.

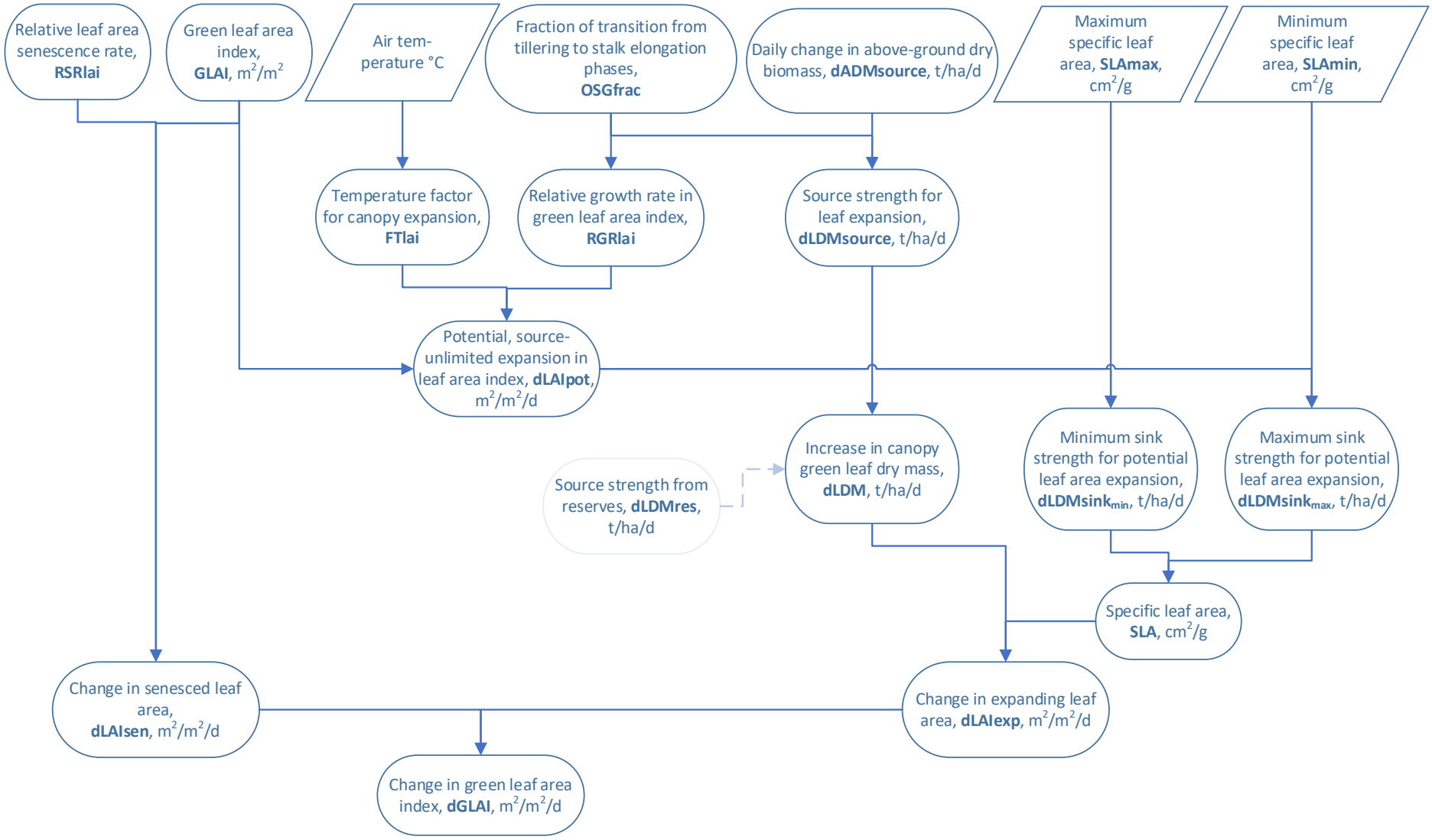

Figure 5-1. Simplified representation of CaneGEM model canopy development. Note that the calculation of source strength from reserves is not detailed in this diagram.

- **Onset of stalk growth is determined by radiation interception, and indirectly by source:sink balance**

  A transition from the tillering to stalk growth phases is simulated, rather than this be treated as a single event, capturing the idea that the population of shoots of different ages and developmental stages reach OSG over a period of time. OSG is a function of photosynthetically-active radiation (PAR) interception rather than elapsed thermal time (DSSAT-Canegro, APSIM-Sugar) or an above-ground dry biomass threshold (Mosicas).

  The time to OSG is therefore dependent on canopy growth rate, which is highly dynamic with respect to E conditions and trait parameters controlling canopy development. OSG is therefore an emergent complex phenotype, and not controlled directly by a trait parameter.

- **Leaf senescence is driven by leaf area index**

  Leaf senescence has in previous models been driven by maximum green leaf number per stalk, but maximum green leaf number is not stable across Es (Singels and Smit, 2009).

  In the new CaneGEM model, leaf senescence starts at a GLAI threshold and senescence rate is a function of the difference between current and target GLAI (these parameters are described in Section 5.2.5). GLAI each day is therefore the emergent consequence of green canopy growth and senescence. Senescence can also occur in response to frost.

- **Partitioning of above-ground biomass is governed by onset of stalk growth**

  The extent of transition from tillering to stalk growth phases (onset of stalk growth – OSG) determines the partitioning of above-ground biomass between leaves and stalks. This process is therefore dynamic and emergent with respect to source:sink drivers of canopy development.

- **Sucrose accumulation is a passive consequence of source:sink balance**

  Canopy growth and stalk elongation (driven by effective temperature and controlled via G-specific specific elongation rate and density parameters, and cardinal temperatures) drive sink demand for structural fibre; any remaining source is stored as sucrose or reserves. Sucrose can only be stored when the plant is physiologically able to do so, and the capacity to accumulate sucrose is linked to internode number, offset from OSG.

  Although this model depends on empirical relationships, it is asserted that the fundamental mechanism is more realistic than the approaches used in DSSAT-Canegro, Mosicas and APSIM-Sugar.

  Sucrose accumulation is also an emergent and highly-complex consequence of E conditions and trait parameters controlling other plant processes.

The descriptions of the CaneGEM model processes in the sections that follow define several model input parameters (whose values are determined before running the model), as well as rate/state variables (whose values are determined daily during the simulation of an individual cropping season). For reference, these parameters are listed and defined in the present section, although they are relevant to, and defined in context, in the sections that follow.

- Model input parameters and their values are listed in Table 5-1.
- Rate and state variables used in the model are listed in Table 5-2 for reference.

Table 5-1. CaneGEM model input parameters and their values for calibrated for the standard reference cultivar NCo376. Explanations of how values were derived are provided in the text in the following sections.

Acronym	Units	Description	Standard value (NCo376)	Source of value
Temperature responses				
Tb_germ_P	°C	Base temperature for germination (plant crops)	16.0	(Jones and Singels, 2018)
Tb_germ_R	°C	Base temperature for germination (ratoon crops)	10.0	(Jones and Singels, 2018)
Tb_lai	°C	Base temperature for leaf area index expansion	13.00	Calibrated
Tb_lfapp	°C	Base temperature for leaf appearance	9.00	(Bonnett, 1998)
Tb_photo	°C	Base temperature for photosynthesis	10.00	(Jones and Singels, 2018)
Tb_sk	°C	Base temperature for stalk expansion	16.00	Calibrated
Tf_germ	°C	Ceiling temperature for germination	41.0	(Jones and Singels, 2018)
Tf_lai	°C	Ceiling temperature for leaf area index expansion	45.00	(Jones and Singels, 2018)
Tf_lfapp	°C	Ceiling temperature for leaf appearance	40.00	(Jones and Singels, 2018)
Tf_photo	°C	Ceiling temperature for photosynthesis	45.00	(Jones and Singels, 2018)
Tf_sk	°C	Ceiling temperature for stalk expansion	48.00	(Jones and Singels, 2018)
To_germ	°C	Optimal temperature for germination	28.0	(Jones and Singels, 2018)
To_lai	°C	Optimal temperature for leaf area index expansion	35.00	Calibrated
To_lfapp	°C	Optimal temperature for leaf appearance	28.00	(Jones and Singels, 2018)
To_sk	°C	Optimal temperature for stalk expansion	35.00	(Jones and Singels, 2018)
To1_photo	°C	Lower bound of optimal temperature range for photosynthesis	20.00	(Jones and Singels, 2018)

Acronym	Units	Description	Standard value (NCo376)	Source of value
To2_photo	°C	Upper bound of optimal temperature range for photosynthesis	40.00	(Jones and Singels, 2018)
Phenology				
$FI\_OSG_{1.5}$		FI_OSG value at a reference row-spacing of 1.5 m.	0.80	Calibrated
LFPI1	°Cd	Short (young crop) leaf phyllocron interval	70	(Singels et al., 2008)
LFPI2	°Cd	Long (mature crop) leaf phyllocron interval	170	(Singels et al., 2008)
$OSGfrac_{c1}$		Slope parameter controlling the rate at which the crop transitions from tillering to stalk growth as a function of radiation interception	20	Calibrated
TTem_P	°Cd	Thermal time from crop start to start of primary shoot emergence, plant crops	150	Calibrated
TTem_R	°Cd	Thermal time from crop start to start of primary shoot emergence, ratoon crops	43	Calibrated
Canopy expansion				
GLAIinitial	m^2/m^2	Starting green leaf area index for a reference crop (1.5 m rows-spacing)	0.15	Calibrated
GLAIst	m^2/m^2	GLAI at which senescence starts.	3.2	Calibrated
Ke_{max}		Maximum PAR interception extinction coefficient	0.86	(Singels et al., 2008)
Ke_{min}		Minimum PAR interception extinction coefficient	0.58	(Singels et al., 2008)
$LAIsen_{c1}$		Slope parameter relating relative senescence rate with GLAI above the GLAIst threshold	0.003	Calibrated
$Lfno_{Kemax}$	l/shoot	Leaf number at which Ke is maximised	20.00	(Singels et al., 2008)
$RGRlai_{max}$		Maximum relative GLAI growth rate	0.250	Calibrated
$RGRlai_{min}$		Minimum relative GLAI growth rate	0.014	Calibrated
SLA_{max}	cm^2/g	Maximum specific leaf area	60	Calibrated
SLA_{min}	cm^2/g	Minimum specific leaf area	20	Calibrated
Biomass accumulation				
RES_{exfrac}	g/g	Maximum fraction of underground reserve that can be used to support structural growth each day.	0.10	Calibrated
RTPFavg	g/g	Growing period average daily partitioning fraction of photoassimilate to root dry mass.	0.19	DSSAT-Canegro and APSIM-Sugar

Acronym	Units	Description	Standard value (NCo376)	Source of value
RUEadm	g/MJ	Theoretical maximum radiation use efficiency, for a young, healthy, unstressed crop, defined for above-ground dry biomass per unit intercepted photosynthetically-active radiation.	2.90	Calibrated
Biomass partitioning: roots				
$RTPF_{c1}$	g/g	Maximum allocation fraction of photosynthetic source to root dry mass	0.95	(Singels et al., 2008)
$RTPF_{c2}$	g/g	Minimum allocation fraction of photosynthetic source to root dry mass	0.88	(Singels et al., 2008)
$RTPF_{c3}$		Exponential parameter controlling the rate at which RTPF decreases as a function of TDM	0.60	(Singels and Bezuidenhout, 2002)
Biomass partitioning: canopy				
$SKPF_{max}$	g/g	Maximum partitioning fraction of the daily change in ADM to stalk dry mass	0.76	Calibrated
Biomass partitioning: stalk fibre				
$SKexp_o$	$cm^3/m^2/d$	Maximum daily stalk expansion rate	215.0	Calibrated
Biomass partitioning: stalk sugars				
$LFNOsks_{max}$		Number of leaves that must appear after $LFNOsks_{min}$ before stalk sugar accumulation capacity is maximised	7.0	Calibrated
$LFNOsks_{min}$		Number of leaves that must appear after start of OSG before stalk sugars can accumulate	3.0	Calibrated
SFHXR	g/g	Stalk fibre synthesis:hexose ratio, i.e. mass of stalk hexoses required to synthesise a unit mass of stalk fibre	30.0	Calibrated

Table 5-2. Rate and state variables used in the description of the CaneGEM model in the sections that follow.

Acronym	Units	Description
		Temperature responses
FT		Temperature control factor
FTlai		Temperature control factor for leaf area expansion
FTlfno		Temperature control factor for leaf appearance rate
FTphoto		Temperature control factor for photosynthesis
FTsk		Temperature control factor for stalk expansion
		Phenology
CTTem	°Cd	Cumulative thermal time since crop start, for driving germination rate

Acronym	Units	Description
$CTTem_{d-1}$	°Cd	Previous day's thermal time since crop start
dEmFrac		Daily change in the fraction of primary shoot emergence completed
FI_OSG		Fractional radiation interception value at which 50% of shoots have transitioned from tillering to stalk growth.
LFNO	leaves/ shoot	Number of leaves per shoot of reference primary shoot.
OSGfrac		Fraction of shoots that have commenced stalk growth
PI	°Cd	Daily leaf phyllocron interval
TTem_X	°Cd	Thermal time from crop start to start of primary shoot emergence, for crop type X (plant or ratoon)
TTlfno	°Cd	Cumulative thermal time for leaf appearance
		Canopy expansion
dGLAI	$m^2/m^2/d$	Daily change in green leaf area index
dLAIem	$m^2/m^2/d$	Daily change in green leaf area index due to emergence of primary shoots
dLAIexp	$m^2/m^2/d$	Daily expansion in GLAI
dLAIsen	$m^2/m^2/d$	Daily senescence of GLAI
$dLDMsink_{max}$	t/ha/d	Maximum leaf sink strength, calculated with SLA_{min}
$dLDMsink_{min}$	t/ha/d	Minimum leaf sink strength, calculated with SLA_{max}
FIinter		Inter-row fractional interception of photosynthetically-active radiation
GLAI	m^2/m^2	Green leaf area index
RSP1		Slope coefficient regulating FI_OSG response to row-spacing
RSP2V		Intercept coefficient regulating FI_OSG response to row-spacing
RGRlai		Daily source-unlimited relative canopy growth rate
RSRlai		Relative GLAI senescence rate
SLA	cm^2/g	Specific leaf area for leaf expansion on a given day
SLA_{var}	cm^2/g	An internal model variable representing an intermediate value of specific leaf area, unbounded by the input minimum or maximum input specific leaf area values.
		Biomass accumulation
ADM	t/ha	Above-ground dry biomass
dRES	t/ha/d	Daily underground reserve dry mass available to support structural growth
Ke		Daily PAR interception extinction coefficient
PAR	MJ/m^2	Daily incident photosynthetically-active radiation, estimated from SRAD.
PHOTO	t/ha/d	Daily photo-assimilate production
RUEo	g/MJ	Theoretical maximum radiation use efficiency, for a young, healthy, unstressed crop, defined for whole-crop biomass per unit intercepted photosynthetically-active radiation.
SOURCE	t/ha/d	Daily dry biomass increase, consisting of photo-assimilate and remobilised underground reserve dry mass.
		Biomass partitioning: roots
dRDM	t/ha/d	Daily allocation of photosynthetic source to root dry mass
RTPF	g/g	Daily photosynthetic source allocation fraction to root dry mass
TDM	t/ha	Total crop dry mass
		Biomass partitioning: canopy

Acronym	Units	Description
dADMsource	t/ha/d	Daily photosynthetic source allocated to above-ground plant components (green leaf canopy and stalks)
dLAIpot	$m^2/m^2/d$	Potential daily expansion in green leaf area index
dLDM	t/ha/d	Daily change in green leaf dry mass, including leaf blades, leaf sheaths and meristems
dLDMres	t/ha/d	Daily mass of reserve used to supplement photosynthetic source to support structural leaf growth
dLDMsf	t/ha/d	Daily photo-assimilate shortfall limiting leaf expansion
dLDMsource	t/ha/d	Source available to support daily leaf area expansion
LDM	t/ha	Green leaf dry mass, including leaf blades, leaf sheaths and meristems
		Biomass partitioning: senescence
dSEN	t/ha/d	Daily change in senesced material dry mass
SEN	t/ha	Dry mass of senesced plant material
		Biomass partitioning: stalk fibre
dSKF	t/ha/d	Daily change in stalk fibre dry mass
dSKres	t/ha/d	Daily reserve dry mass used to supplement stalk fibre expansion
dSKsf	t/ha/d	Stalk fibre biomass allocation shortfall
dSKsink	t/ha/d	Sink strength for stalk fibre
dSKsource	t/ha/d	Daily photo-assimilate partitioned to stalks
SDM	t/ha	Stalk dry mass
dSKpot	$cm^3/m^2/d$	Daily potential stalk volume expansion rate
		Biomass partitioning: stalk sugars
dSKS	t/ha/d	Daily change in stalk sugars mass
dSKSsource	t/ha/d	Photo-assimilate available for storage as stalk sugars
SKH	t/ha	Stalk hexose mass
SKHfrac	g/g	Hexose fraction of stalk sugars
$SKHfrac_{d-1}$	g/g	Previous day's hexose fraction of stalk sugars
SKS	t/ha	Stalk sugars (sucrose and hexose) mass
SKSC	t/ha	Stalk sucrose mass
$SKSPF_{pot}$		Relative stalk sugars storage capacity, a linear function of leaf number between $LFNOsks_{min}$ and $LFNOsks_{max}$
		Biomass partitioning: underground reserve
$dRES_{max}$	t/ha/d	Daily maximum mass of reserves that can be used to support expansive growth
RES	t/ha	Dry mass stored in underground reserve

5.2.3 Phenological phases

Three broad growth phases stages are considered in this model:

- **Germination**: this is the development phase from the date of crop start until primary shoot emergence. The date of emergence (DoE, d) is usually defined as the date when the 50th percentile primary shoot appears. In this model emergence is represented as the appearance of a certain small initial green leaf area index (parameter *GLAIinitial*, m^2/m^2). A gradual transition from germination to tillering is simulated.
- **Tillering**: this is the phase of canopy and root growth before the onset of stalk growth. It is characterized by horizontal growth of the canopy. Biomass is allocated to roots, the green canopy and underground reserve, and not to stalk fibre or sucrose.

- **Stalk growth phase**: following the gradual onset of stalk growth, biomass is allocated to structural stalk fibre and stalk sucrose, in addition to root and leaf fibre.

Shoot emergence starts after *TTem_P* or *TTem_R* thermal time (°Cd, calculated using cardinal temperatures *Tb_germ_P or Tb_germ_R*, *To_germ*, *Tf_germ*, see Section 5.2.4) has elapsed after crop start, for plant/ratoon crops respectively. Crop start is set by the user. For ratoon crops, this is the date of the previous harvest. For plant crops, it is recommended that the date of the first irrigation or significant rainfall event after planting is specified as the crop start date. The daily change in leaf area due to emergence of underground shoots, *dLAIem* ($m^2/m^2/d$) is calculated as:

$$dLAIem = \begin{cases} 0, CTTem < TTem\_X \\ InitialGLAI * dEmFrac \end{cases} \quad (5\text{-}1)$$

where *GLAIinitial* is the starting green leaf area index when fully-emerged, *CTTem* (°Cd) is the thermal time from crop start to date of primary shoot emergence, *TTem_X* is *TTem_P* or *TTem_R*, and *dEmFrac* represents the daily change in the fraction of emergence completed, calculated as:

$$dEmFrac = \left(1 - exp(-0.025 * \max(0, CTTem - TTem\_X))\right) - \left(1 - exp(-0.025 * \max(0, CTTem_{d-1} - TTem\_X))\right) \quad (5\text{-}2)$$

where $CTTem_{d-1}$ is the previous day's cumulative thermal time value. The underlying exponential equation used in Eq. (5-2) is from the DSSAT-Canegro model (Eq. (7) from Jones & Singels (2018)), which describes the cumulative distribution of primary shoot emergence as a function of thermal time. The coefficient (0.025) is an empirical parameter fitted to this DSSAT-Canegro primary shoot emergence function.

dLAIem is integrated with total green leaf area index each day (see Section 5.2.5).

The transition from tillering to stalk growth (referred to as the "onset of stalk growth", OSG) is implemented as a gradual transition (over physiological development period) rather than a single event, in order to take account of the range in developmental stages of tillers and stalks.

OSGfrac is the fraction of shoots that have commenced stalk growth. A logistic function (Eqn. (5-3); Figure 5-3) defines the relationship between fractional interception of photosynthetically-active radiation (*FIinter*) and *OSGfrac*:

$$OSGfrac = \frac{1}{1 + e^{-1 * OSGfrac_{c1} * (FIinter - FI\_OSG)}} \quad (5\text{-}3)$$

where $OSGfrac_{c1}$ is a slope parameter and *FI_OSG* is the *FIinter* value at which 50% of shoots have transitioned to stalk growth. It is understood however that the light environment within the stool – i.e. within-row ("intra-row") radiation interception – determines OSG (Singels and Smit, 2009). For this reason, OSG occurs earlier in *FIinter* terms for crops with wide row-spacing (*RowSpc*, m) and *vice versa*. *FI_OSG* is estimated from row-spacing and a trait parameter, $FI\_OSG_{1.5}$, defined as the inter-row radiation interception value at which 50% of shoots have transitioned to stalk growth, for a reference crop grown at 1.5 m row-spacing. Based on analysis of row-spacing trial data from Singels and Smit (2009), the model assumes that a narrow row-spacing (0.75 m or less) will have *FI_OSG* = 0.85; and that this will decrease curvilinearly with increasing row-spacing

$$FI\_OSG = RSP2 + 0.85 * RowSpc^{RSP1} \quad (5\text{-}4)$$

where *RSP1* is a coefficient regulating the change in *FI_OSG* with row-spacing. Slope and intercept coefficients *RSP1 and RSP2* are calculated dynamically by fitting a curve to cardinal points {(0.75, 0.85); (1.5, FI_OSG$_{1.5}$), and (3.0, $FI\_OSG_{1.5} - (0.9 * (0.85 - FI\_OSG_{1.5}))$)}.

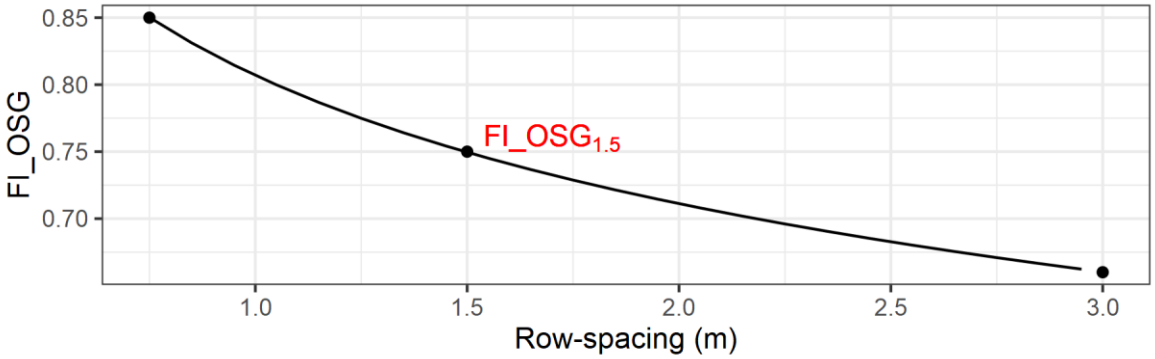

Figure 5-2. Relationship between the FI_OSG (the inter-row fractional radiation interception value at which 50% of shoots have transitioned to stalk growth) and row-spacing.

Typical values for *FI_OSG$_{1.5}$* and *OSGfrac$_{c1}$* are 0.75 and 20.0 respectively, derived by fitting Eqn. (5-3) to observations of SDM/ADM (a measurable proxy for *OSGfrac;* Figure 5-4) and *FIinter* to estimate a value for *FI_OSG* for that row-spacing (1.5 m), and then applying the inverse of Eqn (5-4) to calculate *FI_OSG$_{1.5}$*.

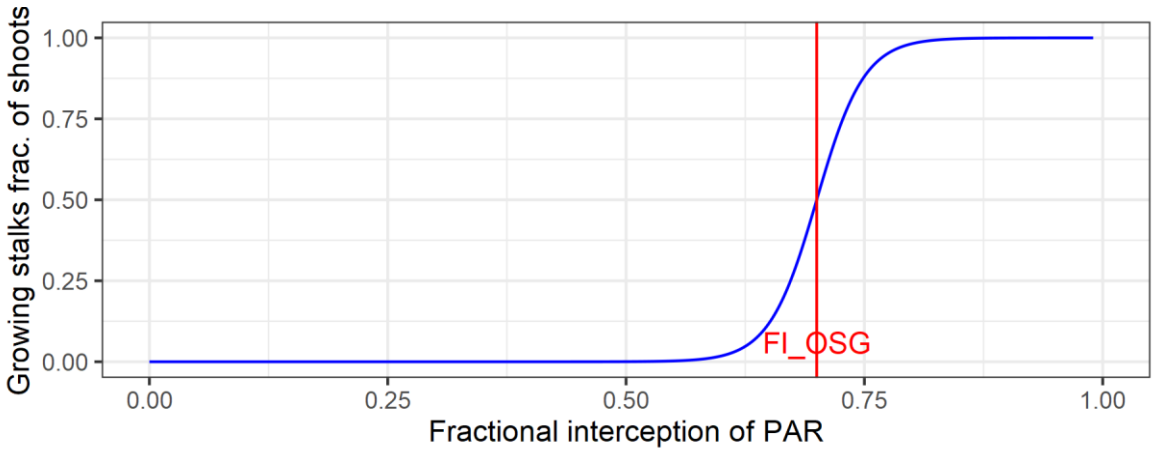

Figure 5-3. Onset of stalk growth fraction (OSGfrac) as a logistic function of fractional interception of photosynthetically-active radiation (PAR). Parameter values are explained in the text; parameter values FI_OSG and OSGfrac$_{c1}$ are 0.70 and 20 respectively in this example.

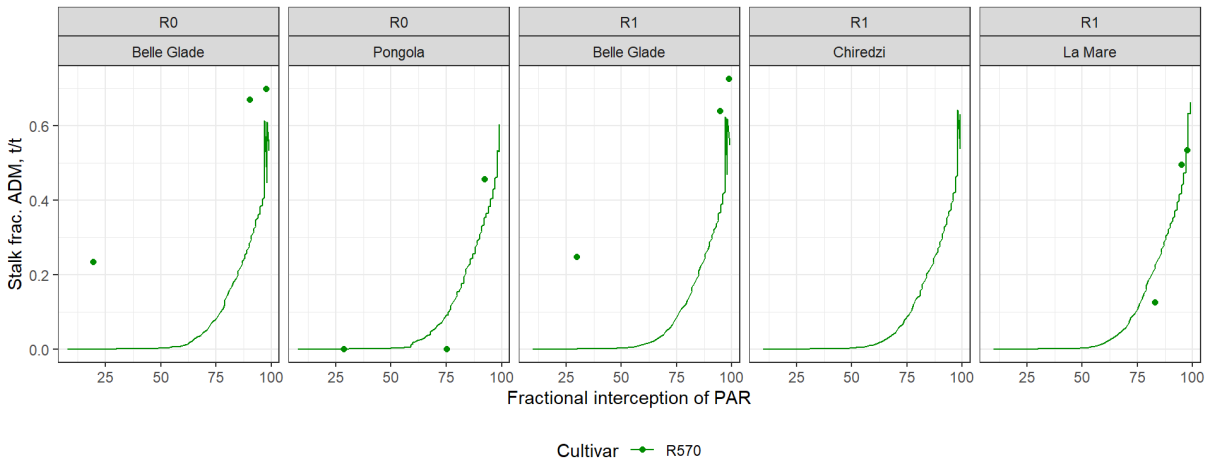

Figure 5-4. Example of fitting values for parameters $OSGfrac_{c1}$ and FI_OSG, used in the model equation for calculating the transition from tillering to stalk growth. The data shown are for cultivar R570 at five irrigated sites, 1.5 m row-spacing. The "Belle Glade" crops included leaf sheath dry mass with stalk dry mass observations, resulting in overestimated values at low fractional radiation interception values.

Although this model does not consider individual leaves or shoots, a nominal number of leaves (*LFNO*, l/shoot) for a representative 50th-percentile primary shoot is maintained and used as a measure of physiological age to regulate two process parameter values: radiation extinction coefficient (explained in Section 5.2.6), and the start of stalk sugar accumulation (Section 5.2.7, Eq. (5-40)).

LFNO is calculated using a daily leaf phyllocron interval *PI* and cumulative thermal time (*TTlfno*, °Cd), as per DSSAT-Canegro (Jones and Singels, 2018). *PI* is calculated from two trait parameters: *LFPI1* and *LFPI2* (°Cd), phyllocron interval values corresponding with rapid leaf appearance in the young crop and slower leaf appearance in the mature crop (Inman-Bamber, 1994). The model switches from the short phyllocron to the longer phyllocron as the crop transitions from the tillering the stalk growth phase, with the following calculation:

$$PI = LFPI1 + (LFPI2 - LFPI1) * OSGfrac \quad (5\text{-}5)$$

Leaf PI values are widely published, and can be determined by via interactive calibration to maximise correlation between simulated and observed leaf number data.

5.2.4 Temperature control factors

Several processes (germination, photosynthesis, canopy and stalk expansion, reference leaf appearance rate) are controlled by temperature, represented in the model with process-specific temperature control factors (*FT*). *FT* values are calculated using cardinal temperatures: *FT* is zero for daily mean air temperatures below the 'base temperature' (*Tb*); increases linearly to a maximum value when temperatures equal or are in the range of 'optimal temperature' (*To*); and decrease linearly to zero at and above ceiling temperature (*Tf*). Different base temperature for germination values are permitted for plant crops (*Tb_germ_P*, °C) and ratoon crops (*Tb_germ_R*, °C). This was done for two reasons, despite it being physiologically likely that these values are the same. Firstly, a wide range of base temperatures for germination have been published, sometimes for the same genotype, so this functionality provided greater flexibility to explore different values. Secondly, as the buds in ratoon crops are closer to the soil surface than for plant crops, soil temperatures fluctuate within a wider range

over the course of the day; this dynamic is lost when mean daily air temperature is used to drive germination, and may have the apparent effect of a different base temperature value.

FT values are calculated for processes listed in Table 5-1; *FT* temperature functions are shown for canopy expansion and photosynthesis (Figure 5-5).

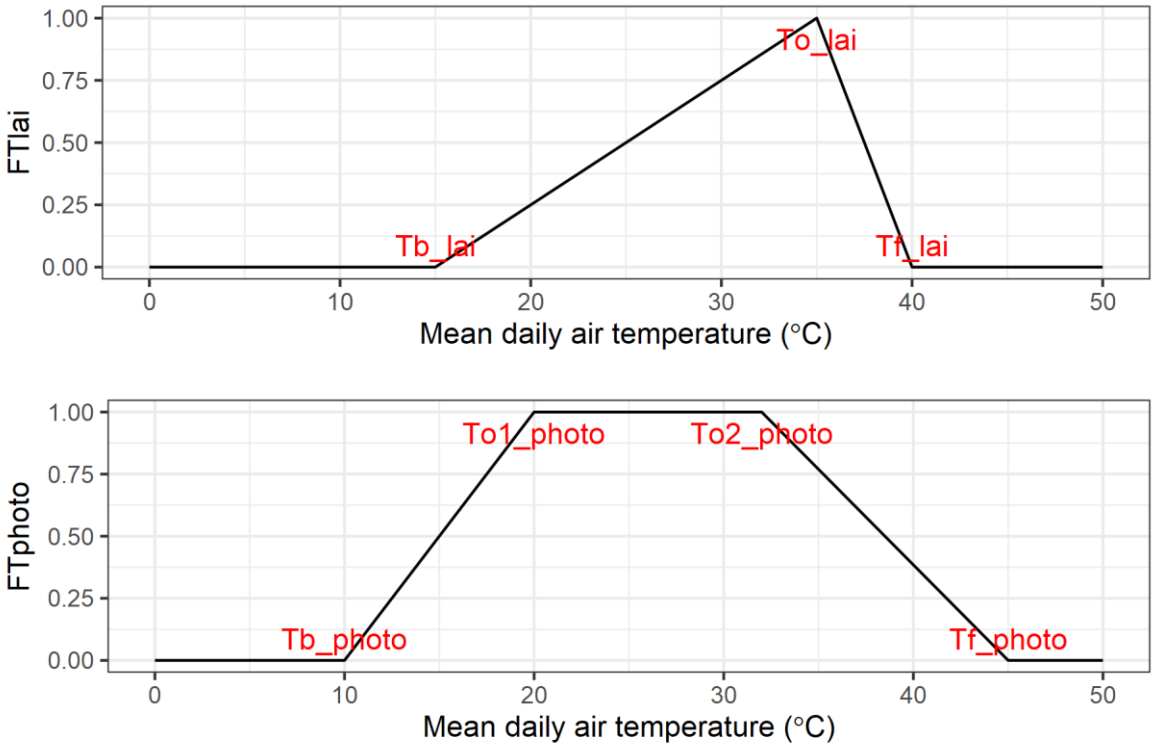

Figure 5-5. Temperature factor relationships with daily mean air temperature for the control of leaf area expansion (FTlai) and photosynthesis (FTphoto).

Values for these parameters were based on those published by Jones & Singels (Jones and Singels, 2018) for the DSSAT-Canegro model.

5.2.5 Canopy expansion

The daily change in green leaf area index (*dGLAI*, $m^2/m^2/d$) is calculated (Eqn. (5-6)) from daily leaf area expansion (*dLAIexp*, $m^2/m^2/d$, Eqn. (5-7)) and daily senescence (*dLAIsen*, $m^2/m^2/d$, Eqn. (5-14)) rates. *dLAIexp* is calculated from the dry mass of new leaf area each day (*dLDM*, t/ha/d) and specific leaf area (*SLA*, cm^2/g).

$$dGLAI = dLAIem + dLAIexp - dLAIsen \quad (5\text{-}6)$$

$$dLAIexp = dLDM * SLA * 0.01 \quad (5\text{-}7)$$

The calculation of *SLA* uses the concept from the APSIM-Sugar model (Keating et al., 1999). *SLA* is permitted to vary daily between G input minimum (SLA_{min}) and maximum (SLA_{max}) *SLA* values, depending on source:sink status. *SLA* is set to SLA_{max} when assimilate partitioned to leaves (*dLDMsource*, t/ha/d, dry mass of leaf blades, sheaths and meristem) is insufficient to fulfil the requirements of the expanding leaf area,

calculated using SLA_{max} ($dLDMsink_{min}$, t/ha/d, Eqn. (5-10)). *SLA* is set to SLA_{min} (i.e. densest leaves) when *dLDMsource* exceeds the biomass required to grow the leaf area that day, calculated using SLA_{min} ($dLDMsink_{max}$, t/ha/d, Eqn. (5-11)).

In cases where the source:sink balance falls between these boundaries, a proportional *SLA* (SLA_{var}, cm^2/g), between SLA_{min} and SLA_{max}, is determined (this is not specified for the APSIM-Sugar model).

This behaviour is described formally as follows:

$$SLA = \begin{cases} SLA_{min}\ , SLA_{var} < SLA_{min} \\ SLA_{var}, SLA_{min} \leq SLA_{var} \leq SLA_{max} \\ SLA_{max}, SLA_{var} > SLA_{max} \end{cases} \quad (5\text{-}8)$$

$$SLA_{var} = SLA_{min} + \left((SLA_{max} - SLA_{min}) * \left(1 - \frac{dLDMsource - dLDMsink_{min}}{dLDMsink_{max} - dLDMsink_{min}} \right) \right) \quad (5\text{-}9)$$

The minimum sink strength for the expanding canopy ($dLDMsink_{min}$, t/ha/d), corresponding with maximum *SLA*, is calculated as:

$$dLDMsink_{min} = dLAIpot * \frac{1.0}{SLA_{max}} * 100.0 \quad (5\text{-}10)$$

where *dLAIpot* ($m^2/m^2/d$) is the potential daily expansion in green leaf area index (Eqn. (5-12)). The maximum daily sink strength for canopy expansion ($dLDMsink_{max}$ (t/ha/d), corresponding with SLA_{min}) is calculated as:

$$dLDMsink_{max} = dLAIpot * \frac{1.0}{SLA_{min}} * 100.0 \quad (5\text{-}11)$$

Minimum and maximum SLA values (i.e. for SLA_{min} and SLA_{max} respectively) were calculated by dividing measurements of green leaf area index by green leaf canopy dry mass (i.e. the sum of leaf blade, meristem and leaf sheath dry masses) and rounded to the nearest 5 cm^2/g. It is noted that the definition of SLA in other crop models, and values reported in the literature, usually excludes leaf sheath and meristem dry mass.

The calculation of *dLAIpot* forms one of defining features of this model: a relative growth rate concept is used, whereby the size of the canopy can potentially increase by a fraction of its current size each day (and where the actual increase is limited by assimilate partitioned to the canopy). This fraction decreases with crop age, reflecting a situation where green leaf area index (*GLAI*, m^2/m^2) can double in a matter of days in a young crop (when transitioning from one to two leaves, for example), but increases by a much smaller proportion each day in a fully-canopied crop. *dLAIpot* is calculated as:

$$dLAIpot = RGRlai * FTlai * GLAI \quad (5\text{-}12)$$

$$RGRlai = RGRlai_{min} + \left((RGRlai_{max} - RGRlai_{min}) * (1 - OSGfrac) \right) \quad (5\text{-}13)$$

where *RGRlai* is the daily source-unlimited relative canopy growth rate; $RGRlai_{max}$ and $RGRlai_{min}$ are respectively maximum and minimum relative *GLAI* growth rate trait parameters. The transition from $RGRlai_{max}$ to $RGRlai_{min}$ (Eqn. (5-13)) is regulated by *OSGfrac*.

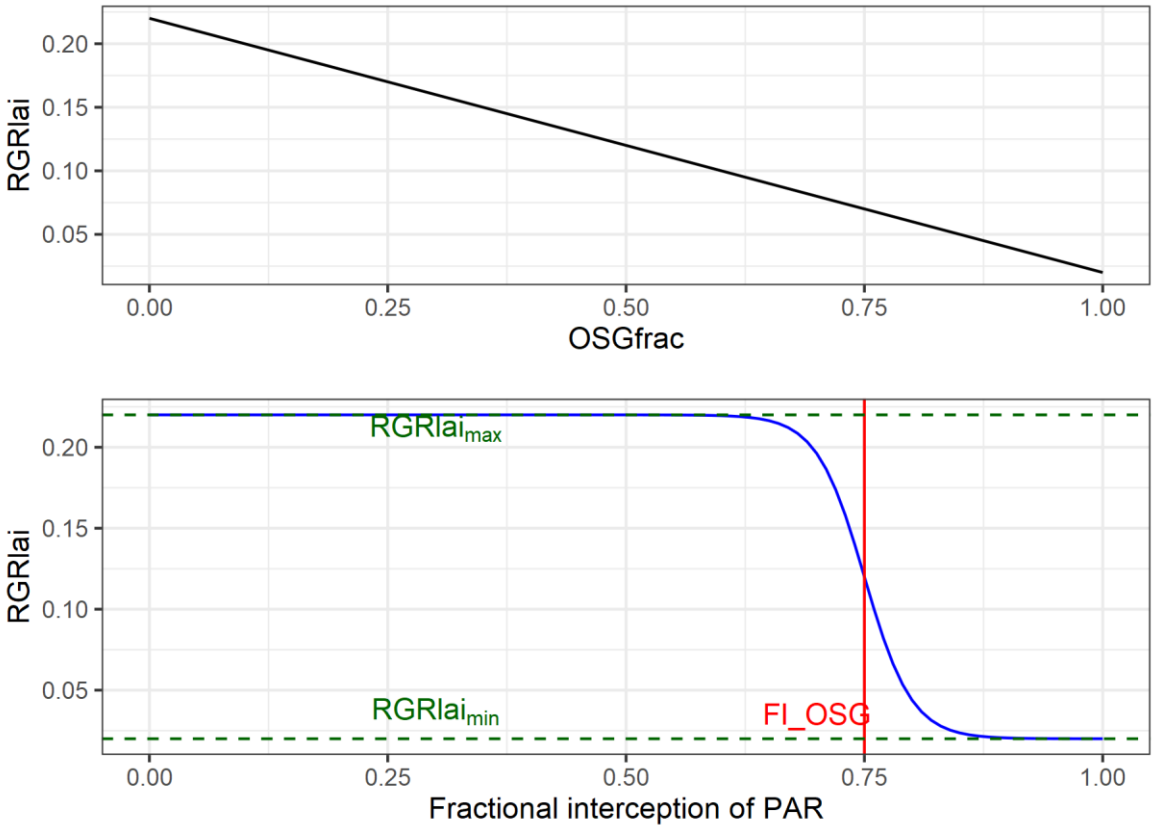

Figure 5-6. Example relationship showing source-unlimited relative growth rate of green leaf area index (RGRlai), as a function of OSGfrac (top), which is itself a function of fractional interception of radiation (FIPAR) (bottom). Parameter names are explained in the text.

Initial guidance for determining values for $RGRlai_{max}$ and $RGRlai_{min}$ came from analysing well-calibrated DSSAT-Canegro simulations (Jones et al., 2021). These values were finalised via interactive calibration with the objective of minimising differences between simulated and observed GLAI and senesced leaf mass.

The calculation of *dLAIsen* is based on the shading-driven aspect of leaf senescence in the APSIM-Sugar model[2].

$$dLAIsen = GLAI * RSRlai \quad (5\text{-}14)$$

$$RSRlai = \begin{cases} 0, GLAI \leq GLAIst \\ LAIsen_{c1} * (GLAI - GLAIst) \end{cases} \quad (5\text{-}15)$$

where *RSRlai* is the daily relative senescence rate (the fraction of *GLAI* that senesces each day), identical to the APSIM variable *slai_light_fac*. Equation (5-15) is illustrated in Figure 5-7. Trait parameter *GLAIst* (m^2/m^2; in APSIM, *lai_sen_light*; typical values 2.3-3.0 m^2/m^2) represents the threshold *GLAI* at which senescence starts. According to Equation (5-15), the relative senescence rate will be higher the larger the difference between *GLAI* and *GLAIst*; this relative senescence rate is zero when *GLAI* ≤ *GLAIst*.

[2] The description of senescence in the APSIM-Sugar model literature reviewed lacks detail. The algorithm was based on the code viewed at:

https://github.com/APSIMInitiative/APSIMClassic/blob/27396788a2c0fcc2750b72700d7fccf35fc7c7ea/Model/CropTemplate/crp_cnpy.f90 (1 June 2021)

Trait parameter *LAIsen*$_{c1}$ (0.007, *sen_light_slope* in APSIM) defines the relationship between *GLAI* above the threshold and *RSRlai*.

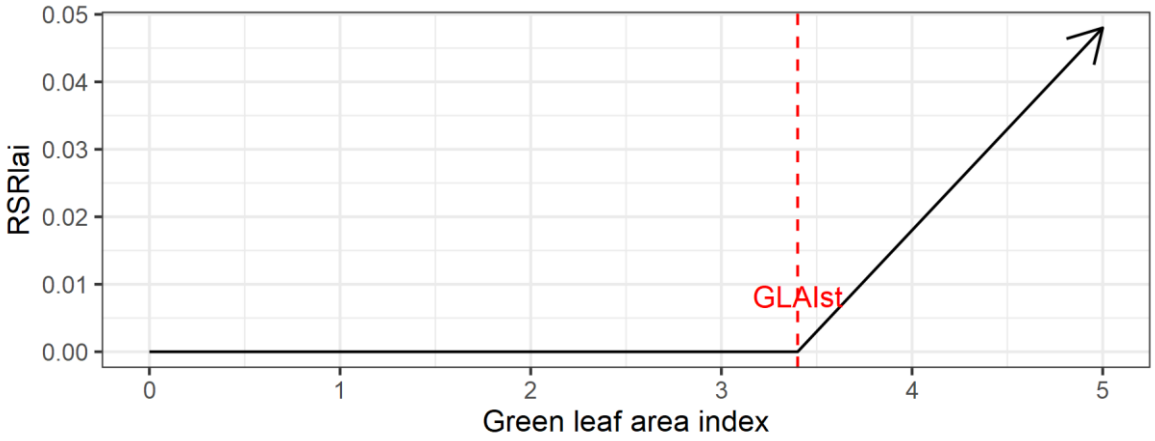

Figure 5-7. Relationship between the relative fraction of green leaf area index (GLAI, m^2/m^2) senesced each day (RSRlai) and GLAI. Parameter GLAIst is explained in the text.

Leaf senescence can be triggered by, or its rate enhanced, by frost. The approach used in CaneGEM for capturing frost impacts on senescence is based on APSIM-Sugar: the leaf area fraction senesced due to frost ranges linearly between 100 % at a minimum air temperature of -5 °C or below, to zero at 0 °C.

No provision is currently made for accelerating senescence with water stress, as the model does not simulate water stress. This could be done by making the value of *GLAIst* dependent on water status, or the ratio of total source to sink strengths.

GLAI is integrated each day as

$$GLAI = GLAI + dGLAI \tag{5-16}$$

Fractional interception of photosynthetically-active radiation by the crop canopy (*FIinter*) is calculated from green leaf area index (GLAI, m^2/m^2) as

$$FIinter = 1 - e^{-Ke*GLAI} \tag{5-17}$$

where *Ke* is a photosynthetically-active canopy radiation extinction coefficient that varies with leaf number (Singels et al., 2008). This is the standard Beer's Law relationship, used in many crop models including DSSAT-Canegro.

Ke is calculated by interpolating linearly between minimum (*Ke*$_{min}$) photosynthetically-active canopy radiation extinction at leaf 1, and maximum extinction (*Ke*$_{max}$) at and after a defined maximum leaf number (Lfno$_{Kemax}$). This is illustrated in Figure 5-8. Values for DSSAT-Canegro (Jones and Singels, 2018) were used.

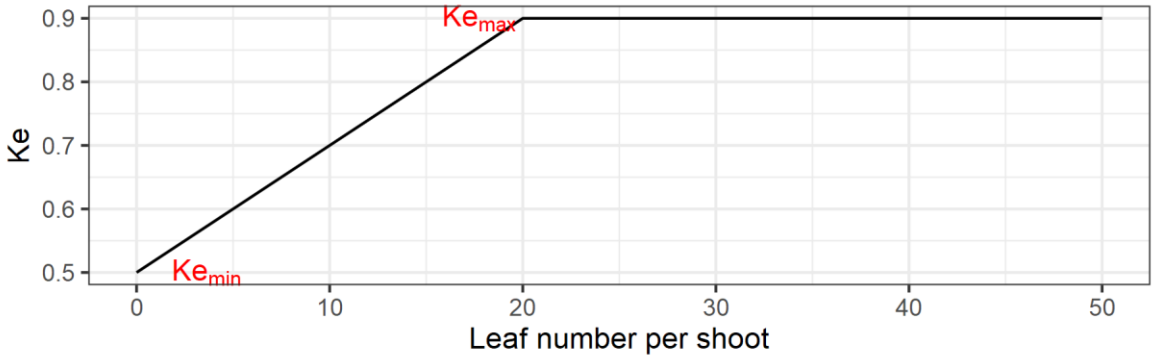

Figure 5-8. Example relationship between canopy extinction coefficient and physiological age expressed as primary shoot leaf number.

5.2.6 Biomass accumulation (photosynthesis)

Source strength (*SOURCE*, t/ha/d) includes daily photosynthesis rate (*PHOTO*, t/ha/d) and the daily provision of carbohydrate from the underground reserve (*dRES*, t/ha/d).

$$SOURCE = PHOTO + dRES \quad (5\text{-}18)$$

Photosynthesis is calculated using a radiation-use efficiency approach (Eqn. (5-19)). Two measures of radiation use efficiency are used: one for above-ground mass only (as this is practical to measure), and another for whole-plant radiation use efficiency (as this allow for mass allocation to roots). The theoretical maximum radiation use efficiency trait parameter (*RUEadm*, g/MJ) is defined for optimal temperature and moisture availability conditions, for above-ground biomass only. Growth and maintenance are already accounted-for in the *RUEadm* value, but this excludes mass allocated to roots. Whole-plant theoretical maximum radiation use efficiency (*RUEo*, g/MJ) is defined as the net photosynthesis rate (after deductions from growth and maintenance respiration), per unit of intercepted photosynthetically-active radiation under optimal temperature, moisture and nutrition conditions. *RUEo* is estimated (Eqn. (5-20)) from *RUEadm* and the average fraction of respiration-net photosynthesis that is allocated to roots over the lifetime of the crop (trait parameter *RTPFavg* (g/g)).

$$PHOTO = FTphoto * FIinter * RUEo * PAR * 0.01 \quad (5\text{-}19)$$

$$RUEo = RUEadm * \left(\frac{1}{1 - RTPFavg}\right) \quad (5\text{-}20)$$

where $PAR = 0.5 * SRAD$, and *SRAD* (MJ/m^2/d) is daily incident global solar radiation (model input).

The basis for the values for *RUEadm* came from previous studies (Jones et al., 2021, 2019), but interactive calibration was required to finalise the *RUEadm* values. The value for *RTPFavg* (19%) was a compromise between values used in the DSSAT-Canegro (≈14% of total dry mass for a crop of 100 t/ha total dry mass) and APSIM-Sugar models (30% of ADM at emergence decreasing asymptotically to 20% at flowering (Keating et al., 1999), i.e. 0.30/(1+0.30) = 23% to 0.20/(1+0.2) = 16%). In practice, this parameter value is calibrated in conjunction with *RUEadm*, but should be between 14 and 23% on the basis of values derived from published parameter values for DSSAT-Canegro and APSIM-Sugar.

Additional source can come from the underground carbohydrate reserve (*RES*, t/ha, see Section 5.2.7, Eq. (5-45)). Source from *RES* is considered supplementary and can only be used to meet structural sink shortfalls; it does not contribute to stalk sugars. Little is known about the rhizome nature of the stool. No more than *RESexfrac* (tentative value = 0.1 g/g) of the reserve can be made available for use by the plant each day. The maximum mass of supplementary source from this reserve that can be supplied each day ($dRES_{max}$, t/ha/d) is calculated as:

$$dRES_{max} = RESexfrac * RES \quad (5\text{-}21)$$

5.2.7 Biomass partitioning to plant components

Root system mass accumulation

The growth of the root system is not simulated in terms of rooting front depth or root length density, because water stress and uptake are not simulated, but the mass of the root system is simulated: source is allocated to the root system to ensure that the source:sink dynamics accommodate this sink.

The daily biomass partitioning to roots (*dRDM*, t/ha/d) follows a similar approach to DSSAT-Canegro, where the daily allocation fraction to roots (*RTPF*, g/g) is a function of total crop dry mass (*TDM*, t/ha), starting at a relatively high proportion of daily net photosynthesis ($RTPF_{c1}$, g/g) and then decreasing exponentially (controlled by parameter $RTPF_{c3}$) to a relatively stable smaller value (defined by the parameter controlling the maximum allocation of assimilate to above-ground biomass, $RTPF_{c2}$, g/g). *dRDM*, t/ha/d) is calculated as:

$$dRDM = RTPF * PHOTO \quad (5\text{-}22)$$

where

$$RTPF = \min\left(RTPF_{c1}, 1 - RTPF_{c2} * \left(1 - e^{-1*RTPF_{c3}*TDM}\right)\right) \quad (5\text{-}23)$$

This is illustrated in Figure 5-9.

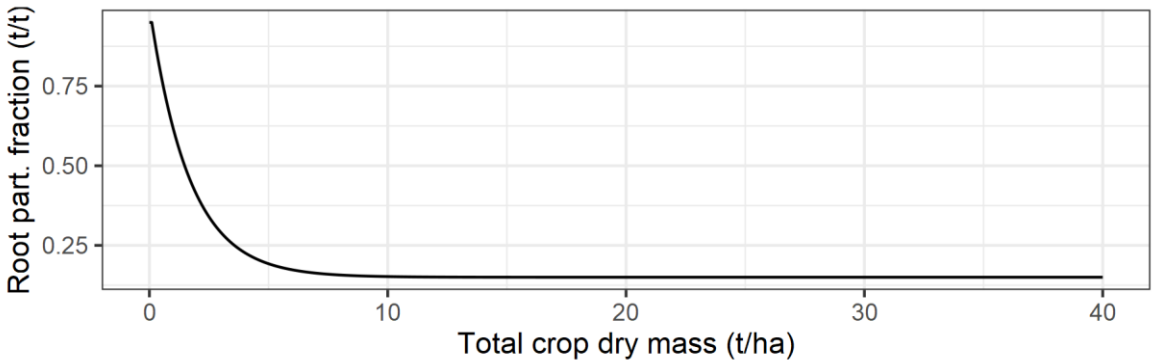

Figure 5-9. Example (using standard parameter values: $RTPF_{c1}$ = 0.95, $RTPF_{c3}$ = 0.6, $RTPF_{c2}$ = 0.88) of root biomass partitioning fraction as a function of total crop dry mass.

Values for $RTPF_{c1}$, $RTPF_{c2}$ and $RTPF_{c3}$ were based on published values for the Canegro model, corresponding with DSSAT-Canegro's *Max_rootpf* (Jones, 2013), $ADMPF_{max}$ and *B* parameters (Singels et al., 2002) respectively.

Canopy mass

The daily allocation of photo-assimilate partitioned to above-ground parts of the stalk (*dADMsource*, t/ha/d) is calculated as the remainder of the daily biomass increment after deducting the daily allocation to root dry mass (*dRDM*, t/ha/d):

$$dADMsource = PHOTO - dRDM \quad (5\text{-}24)$$

The dry mass of the canopy (*LDM*, t/ha) is integrated daily as $LDM = LDM + dLDM$, where *dLDM* (t/ha/d) is the daily increase in canopy dry mass:

$$dLDM = dLDMsource + dLDMres \quad (5\text{-}25)$$

dLDMsource (t/ha/d) is the daily photo-assimilate partitioned to the canopy, and *dLDMres* (t/ha/d) is the daily allocation of underground reserves to supplement canopy expansion.

dLDMsource is calculated allometrically, based on *OSGfrac*:

$$dLDMsource = (1 - OSGfrac * SKPFmax) * dADMsource \quad (5\text{-}26)$$

where *SKPFmax* (g/g; trait parameter) is the maximum fraction of above-ground biomass that can be partitioned to stalk, and *dLDMres* is calculated as:

$$dLDMres = \min(dLDMsf, (1 - OSGfrac * SKPFmax) * dRES_{max}) \quad (5\text{-}27)$$

where the daily leaf source shortfall (d*LDMsf*, t/ha/d) is calculated as

$$dLDMsf = \max(0, dLDMsink_{min} - dLDMsource) \quad (5\text{-}28)$$

Values for *SKPFmax* have been reported in the literature (Jones and Singels, 2018; Keating et al., 1999; Singels and Bezuidenhout, 2002) as this is a common sugarcane model input, can be estimated from observation as the fraction *SDM*/*ADM* (Jones et al., 2019), or calibrated interactively by comparing simulated vs observed *SDM* and *ADM*.

Senesced canopy mass

The increase in dry mass of senesced leaves each day (*dSEN*, t/ha/d) is calculated as

$$dSEN = dLAIsen * \frac{1.0}{SLA} * 100.0 \quad (5\text{-}29)$$

where *SLA* is specific leaf area (cm^2/g) of new leaf area developing on the day of senescence. Ideally, *SLA* on the day the senescing leaves were formed would be used, but as this model does not simulate individual leaves, it was not possible to do this.

The mass of senesced material (*SEN*, t/ha), sometimes termed 'trash', is calculated as:

$$SEN = SEN + dSEN \quad (5\text{-}30)$$

Stalk fibre mass

The objective of calculating stalk expansion is to determine the daily demand for stalk fibre, in order then to estimate stalk sugar accumulation. In principle, a set of growing (well-watered) stalks will elongate (at a rate determined by temperature) and create a

demand for structural stalk fibre as the volume of stalk increases – whether this is rapid elongation for fewer/thinner stalks or less rapid elongation for more/thicker stalks.

dSKF (t/ha/d) is the daily increase in stalk fibre mass. *dSKF* is determined by the lower of sink strength for stalk fibre (*dSKsink*, t/ha/d) and source allocated to stalks (*dSKsource*, t/ha/d), and the allocation of underground reserves to expanding stalks (*dSKres*, t/ha/d):

$$dSKF = min(dSKsource, dSKsink) + dSKres \quad (5\text{-}31)$$

dSKsource is simply any source remaining after meeting leaf expansion demands, and *dSKsink* is based on a daily potential stalk volume expansion rate (d*SKpot*, $cm^3/m^2/d$), which is driven by temperature and crop development (represented by *OSGfrac*).

$$dSKsource = dADMsource - dLDM \quad (5\text{-}32)$$

$$dSKsink = dSKpot * \frac{1}{SSV} * 0.01 \quad (5\text{-}33)$$

$$dSKpot = SKexp_o * FTsk * OSGfrac \quad (5\text{-}34)$$

where $SKexp_o$ is a trait parameter indicating the maximum daily stalk expansion rate ($cm^3/m^2/d$; the 0.01 term in Eqn. (5-33) translates units from cm^3/m^2 to m^3/ha), and *FTsk* is the temperature control factor for stalk expansive growth (see Section 5.2.4).

dSKres is the lower of the stalk fibre biomass allocation shortfall (*dSKsf*, t/ha/d) and available underground reserve source remaining after deductions by leaf expansion demand:

$$dSKres = \min(dSKsf, dRES_{max} - dLDMres) \quad (5\text{-}35)$$

where

$$dSKsf = \max(0.0, dSKsink - dSKsource) \quad (5\text{-}36)$$

Stalk sugars mass

Stalk sugars consist of sucrose and hexoses (glucose and fructose). Stalk sugars mass (*SKS*, t/ha) and the daily change in stalk sugars mass (*dSKS*, t/ha) are given by

$$dSKS = dSKSsource * SKSPF_{pot} \quad (5\text{-}37)$$

$$SKS = SKS + dSKS \quad (5\text{-}38)$$

where *dSKSsource* (t/ha/d) is the daily assimilate available for partitioning to stalk sugars, and $SKSPF_{pot}$ (0-1) represents the relative capacity of the stalk to store sucrose.

dSKSsource is calculated as:

$$dSKSsource = PHOTO + dLDMres + dSKres - dRDM - dLDM - dSKF \quad (5\text{-}39)$$

while $SKSPF_{pot}$ is based on the number of internodes that have appeared since the start of the transition from tillering to stalk growth. $SKSPF_{pot}$ is linearly interpolated (Eq. (5-40), Figure 5-10) between zero (before and until $LFNOsks_{min}$ internodes have appeared), and one (when $LFNOsks_{max}$ internodes have appeared) – at which point the stalks reach their maximum capacity to store sugars.

$$SKSPF_{pot} = f(LFNO) \quad (5\text{-}40)$$

SKSPF$_{pot}$ is limited such that stalk sugar storage capacity cannot exceed 65% of total stalk dry mass.

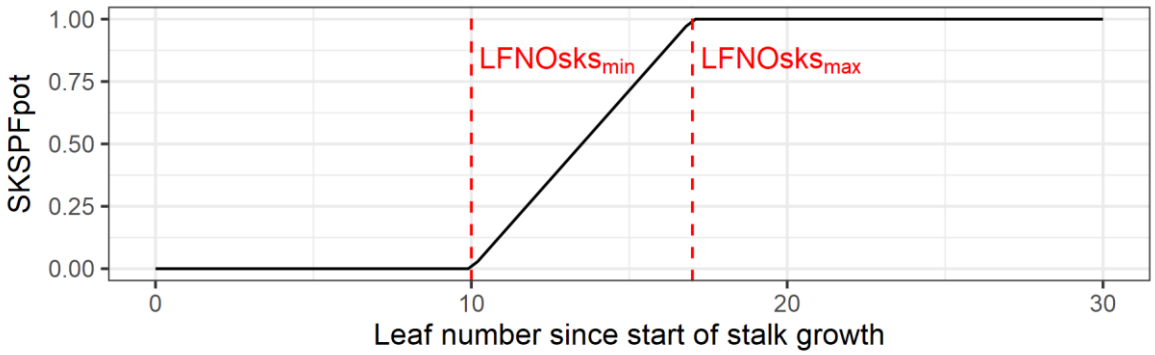

Figure 5-10. Relative stalk sugars partitioning capacity (SKSPF$_{pot}$) vs number of leaves appeared after the start of stalk growth. Parameter values are explained in the text.

The values for trait parameters *LFNOsks$_{min}$* and *LFNOsks$_{max}$* are based on internode sucrose accumulation reported in Singels & Inman-Bamber (Singels and Inman-Bamber, 2011). Considering that this model effectively simulates a single "big stalk" that represents a population of stalks of different ages and stages of development, these parameter values ought to be determined interactively to match simulated and observed mass and dry matter contents of stalk sugars.

Stalk hexose mass

It is assumed that a certain mass of stalk hexose is required for a unit of fibre synthesis in the stalk (stalk fibre synthesis:hexose ratio, *SFHXR*, calibrated as 30.0 g/g) (M. R. Jones et al., 2011). Stalk hexose mass each day (*SKH*, t/ha) is calculated as

$$SKH = SKS * SKHfrac \quad (5\text{-}41)$$

where *SKHfrac* is the hexose fraction of stalk sugars, calculated as:

$$SKHfrac = SKHfrac_{d-1} * 0.95 + min\left(1.0, \begin{cases} \frac{SFHXR * dSKF}{SKS}, SKS > 0.001 \\ 1.0, SKS \leq 0.001 \end{cases}\right) * 0.05, \quad (5\text{-}42)$$

$$0.1 \leq SKHfrac \leq 0.50$$

where *SKHfrac$_{d-1}$* is the previous day's *SKHfrac* value (initialised to 0.50 on the first day of the simulation). The 0.95 and 0.05 terms in Eq. (5-42) ensure that hexose mass fraction cannot change by more than 5% per day; and the hexose fraction is restricted to between 10 and 50% (based on values reported by Singels & Inman-Bamber (2011)).

Eq. (5-42) is illustrated, for two fibre growth rates (dSKF = 0.150 (favourable conditions for expansive growth) and 0.015 (favourable conditions for sucrose accumulation)), in Figure 5-11.

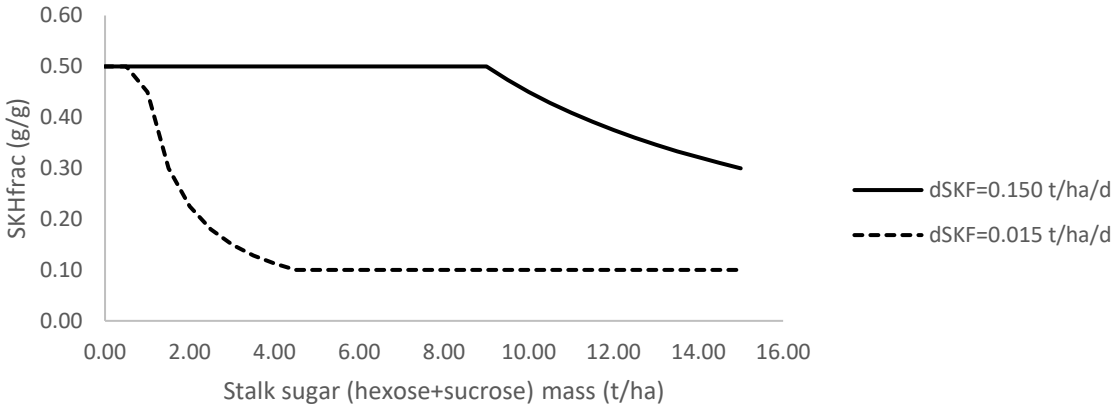

Figure 5-11. Relationship between total stalk sugar mass (SKS, t/ha) and the fraction of sugar that is required to be hexose (SKHfrac, g/g), for two daily fibre synthesis (dSKF, t/ha/d) rates: low (dSKF = 0.015) and high (dSKF = 0.150).

Stalk sucrose mass

The mass of stalk sucrose (*SKSC*, t/ha) is calculated from the complement of hexose fraction:

$$SKSC = (1 - SKHfrac) * SKS \quad (5\text{-}43)$$

As the stalk gets bigger, the mass of fibre synthesised each day varies according to temperature and radiation intensity while the mass of stalk sugars continually increases, resulting in a decreasing overall proportion of hexoses in the stalk (*SKHfrac*, g/g, Eq. (5-42)), and so increasing juice purity. This approach is different to that of Singels & Inman-Bamber (2011), who used a static relationship between structural dry mass fraction of total ADM and hexose fraction.

Underground reserve mass

If the allocation of source to stalk sugars exceeds the capacity of the stalks to store sugars, any remaining source is allocated to the underground reserve.

$$dRES = dSKSsource - dSKS - dLDMres - dSKres \quad (5\text{-}44)$$

The mass of the underground carbohydrate reserve is calculated as

$$RES = RES + dRES \quad (5\text{-}45)$$

5.3 Methodology

5.3.1 Overview and general considerations

The use of simulated soil temperatures to drive germination was assessed for plant and ratoon crops at the ICSM sites. There appeared to be no consistent advantage to using soil temperatures rather than air temperatures, so CaneGEM model development continued using a standard model of germination based on air temperatures. A short description of the exploration of simulated soil temperatures is provided in Appendix A.

CaneGEM model calibration and validation was undertaken using detailed growth analysis datasets (Table 5-3) for cultivar NCo376, as sufficiently detailed datasets for the Gs used for final evaluation of GxE interaction effects were not available.

It was hypothesized that improvements in the simulation of canopy cover would lead to improved simulation of GxE interaction effects for this as well as biomass and yield accumulation. Accordingly, fractional radiation interception (*FIPAR*, *FIPARa*) and aboveground dry biomass accumulation (*ADM*, *ADMh*) were considered important variables both for model calibration and validation. Where it was necessary to estimate date of emergence (DoE, d) from shoot density data, such observations were required. The set of NCo376 datasets with observations that fulfilled these requirements were assigned to the calibration and validation sets.

For the calibration of canopy development, DoE was “forced” to match observations. This was done to ensure the most accurate and realistic calibration of canopy-related parameters. Calibration of germination-related model trait parameters was then conducted separately.

The CaneGEM model was then validated separately with predetermined DoE and simulated DoE.

The calibrated CaneGEM model (with DoE predetermined and simulated, separately) was then assessed for its performance in simulating E, G and GxE interaction effects on seasonal radiation interception and dry biomass accumulation for a multi-environment, multi-genotype trial.

As the objective was to develop an improved sugarcane crop growth simulation model, it was important to compare the performance of the new model with that of an existing well-respected and widely-used crop model. The DSSAT-Canegro v4.5 model (Jones and Singels, 2018) was used to simulate the NCo376 validation experiments (using standard NCo376 trait parameter values) as well as the E, G and GxE interaction effects (using parameter values for N41, R570 and CP88-1762 determined by Jones et al. (2021)). DSSAT-Canegro was chosen for comparison because it presented the strongest challenge against which to benchmark CaneGEM performance:

- DSSAT-Canegro has been calibrated extensively for cultivar NCo376, and the available data for CaneGEM model calibration and validation was also for NCo376.
- DSSAT-Canegro gave the strongest prediction of GxE interaction effects in Jones et al. (2021) (Chapter 4 of this thesis)

The CaneGEM and DSSAT-Canegro models were compared using exactly the same validation and GxE assessment experiments. Only the CaneGEM model was calibrated for variety NCo376, and while both CaneGEM and DSSAT-Canegro were calibrated using the same datasets. It should be noted that datasets used for development of CaneGEM model concepts and model calibration were not explicitly separated, although model validation was undertaken with independent datasets.Datasets for model calibration, validation and assessment of E, G and GxE interaction effects are listed in Table 5-3.

Table 5-3. Detail of experimental datasets used for model calibration (CaneGEM model) and validation (CaneGEM and DSSAT-Canegro models), and for assessment of E, G and GxE interaction effects (CaneGEM and DSSAT-Canegro models). "CC" refers to crop class (P = plant crop, R1 = 1st ratoon, R2 = 2nd ratoon, etc). "Task": C = calibration, V = validation, GE = G / E/ GxE interaction evaluation. "RS" is row-spacing (m).

Country, site and location, reference	CC	Start date	Harvest date	RS	Task	Irrigation type	Soil type
NCo376 crops for calibration and validation							
Pongola, South Africa 27°24'0"S; 31°35'0"E; 308 m a.s.l. "A/Growth/14", Donaldson (2009)	R2	1998-03-03	1999-03-29	1.5	C	Drip	Ultisol Haploxerults
	R2	1998-04-08	1999-05-03	1.5	C		
	R2	1998-05-06	1999-06-07	1.5	V		
	R2	1998-08-06	1999-08-03	1.5	V		
	R2	1998-12-08	1999-11-30	1.5	V		
Mount Edgecombe, South Africa 29°42'S, 31°02'E; 96 m a.s.l. "A/Growth/16", Singels *et al.* (2005c)	R0	1999-12-09	2000-12-07	1.2	C	Drip	Lithocutanic / Mollisol
	R1	2000-06-06	2001-08-13	1.2	V		
	R1	2000-12-07	2002-04-16	1.2	V		
Mount Edgecombe, South Africa 29°42'S, 31°02'E; 96 m a.s.l. "A/Temp", Smit & Singels (2007)	R1	2003-10-13	2004-09-07	1.2	C	Drip	Humic Ferrasols
	R1	2004-04-07	2005-09-06	1.2	V	Drip	
Bruyn's Hill, South Africa 29°25'S, 30°41'E; 990 m a.s.l. "A/Temp", Smit & Singels (2007)	R1	2003-10-16	2004-12-22	1.1	C	Overhead sprinkler	Haplic Phaeozens
	R1	2004-04-07	2005-09-07	1.1	V		
Mount Edgecombe, South Africa 29°42'S, 31°02'E; 96 m a.s.l. "A/Space"[2], Smit & Singels (2009)	R1	2003-08-29	2004-07-20	0.9	V	Drip	Haplic Phaeozens
				1.2	V		
				1.3	C		
				1.6	V		
				1.7	V		
USA, Belle Glade 26°39'02"N; 80°38'08"W; 5 m a.s.l "ICSM IGEP", Jones *et al.* (2019)	P	2013-12-12	2015-01-04	1.5	C	Water table	Dystric Sapric Histosol
	R1	2015-01-23	2016-01-26	1.5	C		
Reunion, La Mare. 20°57'0"S; 55°18'0"E; 70 m a.s.l. "ICSM IGEP", Jones *et al.* (2019)	R1	2016-01-18	2017-01-25	1.5	C	Overhead sprinkler	Hypereutric Nitisols
La Mercy, South Africa 29°50'S, 31°06'E; 50 m a.s.l.	R1	1995-03-08	1996-05-27	1.2	C	Overhead sprinkler	Alfisol Rhodoxeralf

Country, site and location, reference	CC	Start date	Harvest date	RS	Task	Irrigation type	Soil type
"A/Growth/11", Singels *et al.* (1998)							
Pongola, South Africa 27°24'0"S; 31°35'0"E; 308 m a.s.l. "Rostron"	R1	1968-12-17	1970-05-05	1.5	V	Overhead sprinkler	Ultisol Haploxerults
		1969-02-11	1970-06-30		C		
		1969-04-08	1970-08-25		V		
		1969-06-03	1970-10-20		C		
		1969-07-29	1970-12-15		C		
		1969-09-23	1971-02-09		V		
		1969-11-18	1971-04-06		C		
		1970-01-13	1971-05-29		V		
ICSM IGEP[1] dataset (for evaluating E, G and GxE interaction effects) – N41, R570 and CP88-1762							
Reunion, La Mare. 20°57'0"S; 55°18'0"E; 70 m a.s.l. Jones *et al.* (2019)	R1	2016-01-18	2017-01-25	1.5	GE	Overhead sprinkler	Hypereutric Nitisols
South Africa, Pongola. 27°24'0"S; 31°35'0"E; 308 m a.s.l. Jones *et al.* (2019)	P	2014-03-25	2015-03-24	1.5	GE	Overhead sprinkler and drip	Rhodic Cambisol
USA, Belle Glade. 26°39'02"N; 80°38'08"W; 5 m a.s.l Jones *et al.* (2019)	Plant	2013-12-12	2015-01-04	1.5	GE	Water table	Dystric Sapric Histosol
	R1	2015-01-23	2016-01-26	1.5	GE		
Zimbabwe, Chiredzi. 21°02'01.95"S, 31°36'58.52"E, 420 m a.s.l. Jones *et al.* (2019)	R1	2015-06-03	2016-06-03	1.5	GE	Floppy overhead sprinkler	Eutric Luvisol

[1]International Consortium for Sugarcane Modelling, International GxE Project

[2]Destructive (biomass) and non-destructive (radiation interception) measurements were taken in different plots, where, by virtue of the radial row orientation and experimentalist choices, row-spacing was slightly different.

5.3.2 Models, datasets and model assessment scenarios

Two sugarcane crop growth simulation models were assessed: DSSAT-Canegro ("DC") and the newly-developed CaneGEM ("GEM") model. Three datasets (detailed in Table 5-3) were used: two were fully-irrigated growth analysis trial datasets for the South African reference cultivar NCo376, for model calibration ("376C") and validation ("376V"); the third dataset used was the ICSM IGEP dataset previously described. Two sources of germination phase duration (i.e. 50% primary shoot emergence date) values were used: predetermined ("P") values were specified in advance (based on observations), while simulated ("S"), values were calculated internally by the models themselves.

The CaneGEM model does not simulate water relations and so assumed ideal water availability. The DSSAT-Canegro model simulations were set up to irrigate automatically with the objective of minimising water stress. The water stress status of each of the experiments is detailed in Chapter 3 (Jones et al., 2019). Stressed crops, and the last data point where drying-off (imposition of water stress by withholding irrigation prior to final harvest) had been undertaken, were removed from the analyses.

The name, details and purpose for each combination of crop model, assessment dataset and emergence date source is listed in Table 5-4 and used hereon for clarity.

Table 5-4. Combinations of crop growth simulation model, dataset, source of germination phase duration (i.e. date of primary shoot emergence) data, the purpose of the model assessment and corresponding scenario and task codes used in this chapter.

Model assessment scenario code	Crop growth simulation model name	Dataset	Task	Source of duration of germination phase data	Trait parameter values	Purpose
GEM_376C_P	CaneGEM	NCo376 calibration	C	Predetermined	NCo376	Calibrate the CaneGEM model, with a focus on plant processes after primary shoot emergence
GEM_376C_S	CaneGEM	NCo376 calibration	C	Simulated	NCo376	Show the impacts of simulation of duration of germination phase.
GEM_376V_P	CaneGEM	NCo376 validation	V	Predetermined	NCo376	Establish performance of CaneGEM model, with a focus on plant processes after primary shoot emergence
GEM_376V_S	CaneGEM	NCo376 validation	V	Simulated	NCo376	Establish overall performance of CaneGEM model, including simulation of duration of germination phase. Also to permit like-for-like comparison with DSSAT-Canegro
DC_376V_S	DSSAT-Canegro	NCo376 validation	V	Simulated	NCo376	Establish DSSAT-Canegro model performance for simulating E effects, to provide a benchmark against which the CaneGEM model can be compared.
DC_IGEP_S	DSSAT-Canegro	ICSM IGEP	GE	Simulated	N41, R570 and CP88-1762	Establish DSSAT-Canegro model performance for simulating GxE interaction effects, to provide a benchmark against which the CaneGEM model can be compared.
GEM_IGEP_P1	CaneGEM	ICSM IGEP	GE	Predetermined	NCo376	Evaluate performance of CaneGEM model for predicting the impact of correct prediction of duration of germination phase on E, G and GxE interaction effects on canopy development, biomass accumulation and partitioning, without the confounding effects of non germination-related genotypic differences.
GEM_IGEP_P	CaneGEM	ICSM IGEP	GE	Predetermined	N41, R570 and CP88-1762	Evaluate performance of CaneGEM model for predicting E, G and GxE interaction effects on canopy development, biomass accumulation and partitioning, without the confounding effects of possible errors in predicting date of emergence.
GEM_IGEP_S	CaneGEM	ICSM IGEP	GE	Simulated		Evaluate overall performance of CaneGEM model for predicting E, G and GxE interaction effects in terms of canopy development, biomass accumulation and partitioning.

5.3.3 Statistical measures of model performance

Corresponding simulated and observed values for the following set of variables were compared for model calibration and validation:

- Duration of the primary shoot germination phase (*DoGP*, d)
- Green leaf area index (daily, *GLAI*, m^2/m^2)
- Fractional PAR interception (daily, *FIPAR*; seasonal, *FIPARa*),
- Above-ground dry biomass (daily, *ADM*, t/ha; at harvest, *ADMh*, t/ha)
- The duration of the tillering phase, from date of 50% primary shoot emergence to the date of onset of stalk growth (*DoTP*, d)
- Millable stalk dry mass (daily, *SDM*, t/ha; at harvest, *SDMh*, t/ha)
- Sucrose mass (daily, *SUC*, t/ha; at harvest, *SUCh*, t/ha)
- Senesced canopy ("trash") mass (daily, *SEN*, t/ha; at harvest, *SENh*, t/ha)

DoTP was determined for observations by (1) linearly regressing *SDM* against *ADM* and solving for the maximum *ADM* where *SDM* = 0 t/ha (*ADM_OSG*, t/ha); (2) then estimating the thermal time age (*TT_OSG*, °Cd) at which *ADM_OSG* occurred, using linear interpolation between observations of *ADM* against cumulative thermal time (base 10 °C, approximating effective temperature for photosynthesis); and (3) looking up the nearest calendar date to *TT_OSG*, and expressing this in days after date of 50% primary shoot emergence.

Simulated *DoTP* was determined from simulation outputs as the date (days after crop start) when *OSGfrac* was nearest 50%.

Model performance (similarity with observations) was quantified with these statistical measures:

- the coefficient of determination (R^2),
- the slope and intercept ("Y-int") of linear regressions between simulated and observed values
- the Student's t-statistic representing the significance of the slope of linear regression, yielding a p-value
- root mean squared error (RMSE, calculated as the square root of the mean squared difference between observed and simulated values)
- average prediction error (APE, calculated as the average difference between simulated and observed values).

These were calculated using time series and season/at-harvest data, for comparison with published model performance. Data were restricted to observations taken at 12 months' age or less, as sugarcane breeding for irrigated production is aimed at 12-month harvests in most industries, and to avoid confounding from possible RGP effects. Biomass data from the ICSM IGEP experiments were restricted to 9 months' age to minimise confounding from lodging and drying-off.

5.3.4 Determination of dates of emergence

Dates of emergence, expressed as the duration of germination (*DoGP*), were identified for each NCo376 crop in the calibration and validation sets, for two reasons: firstly, to support more accurate calibration of the model's canopy development and biomass partitioning aspects by allowing *DoGP* to be predetermined; and secondly, in order to calculate optimal parameter values for base temperature for germination (*Tb_germ_P*

and *Tb_germ_R*, °C) and thermal times to emergence for plant (*TTem_P*, °Cd) and ratoon crops (*TTem_R*, °Cd).

For the purposes of estimating *DoGP*, the definition of date of emergence was taken as the earliest date that 50% of primary shoots had emerged. 100% primary shoot density was assumed to be the stalk density at harvest (Bezuidenhout et al., 2003; Donaldson, 2009; Keating et al., 1999). Stalk density at harvest was determined as the mean of the last three observations of stalk density, unless these were clearly taken during the shoot senescence phase, in which case fewer observations were used or another similar experiment (at the same location) was used to estimate this value. A 2nd-order polynomial function of thermal time (base 16 °C, Jones & Singels (2018)) was fitted to the shoot density data in order to interpolate/extrapolate from observations, to determine the date at which shoot density was equal to 50% of the final density (Figure 5-12). In cases where the apparent date of emergence was before the date of crop start, date of emergence was set to the date of crop start. Where possible, the polynomials were fitted to data up to peak shoot density, or a very clear earlier inflection point (as described in Jones et al. (2019)). On the basis that shoot density growth and senescence often appear roughly symmetrical with respect to thermal time, where insufficient pre-peak observations had been made, additional observations beyond the date of peak density were included for fitting.

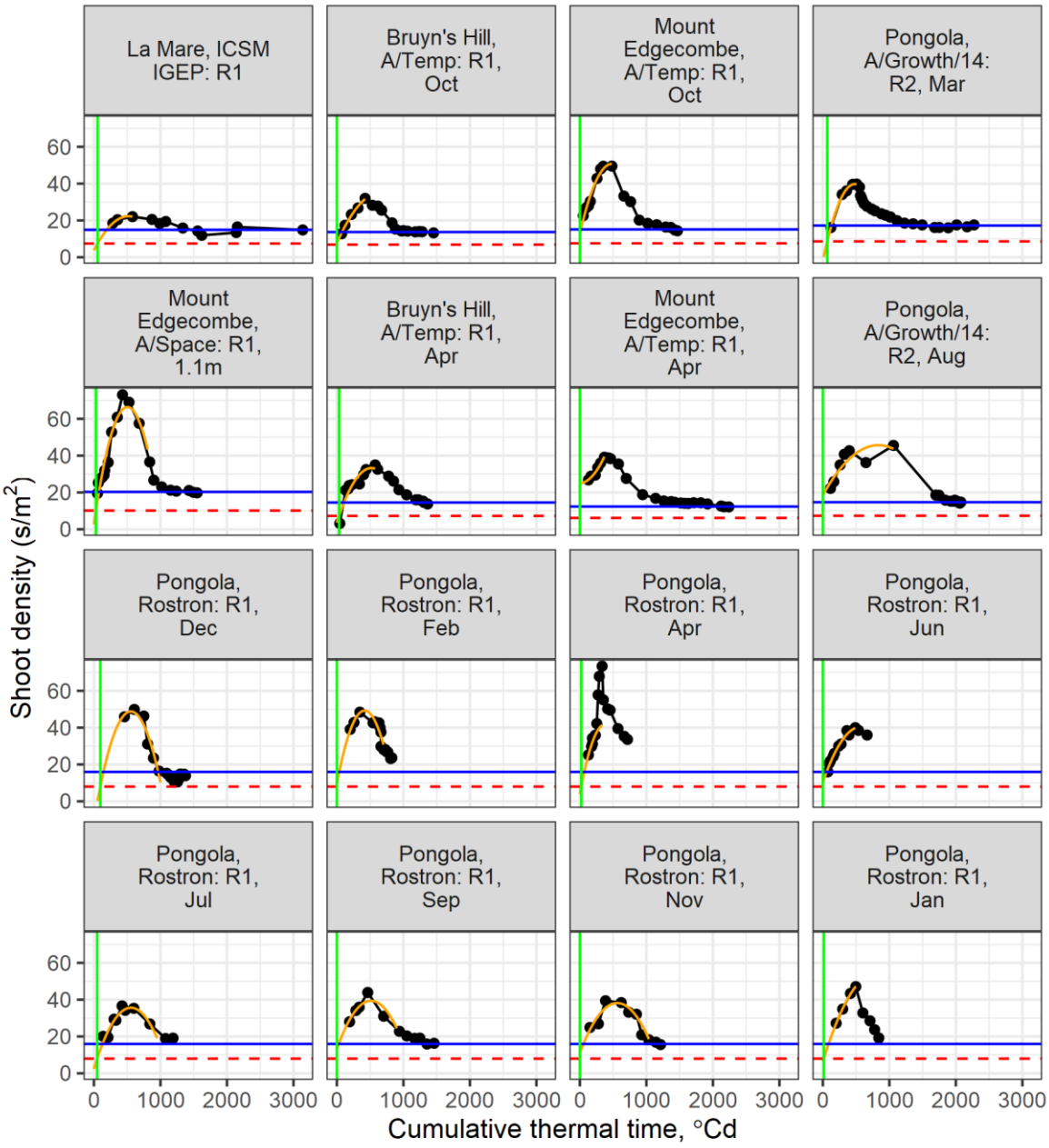

Figure 5-12. Polynomial functions (orange solid lines) fitted to shoot density vs thermal time from crop start data (black lines and points) to estimate dates of emergence for NCo376 calibration and validation datasets. The vertical green lines indicate the date of emergence; blue solid lines the estimated final stalk density; and the dashed red line 50% of the final value.

Observed or estimated dates of emergence from Jones et al. (2019) were used for the ICSM IGEP simulations for assessment of E, G and GxE interaction effects.

5.3.5 CaneGEM trait parameter calibration with NCo376 data

As only two plant crops had been included in the calibration dataset ("376C", Table 5-4), and both were in relatively cold environments, the optimal base temperature for plant crops was not calculated. *Tb_germ* for plant crops was set to 16 °C and *TTem_P* set to 150 °Cd, based on literature (Jones and Singels, 2018).

The thermal time from crop start to emergence was determined for 28 ratoon crops, calculated with *Tb_germ* values 5-20 °C in 0.5 °C increments, with fixed *To_germ* and *Tfin_germ* values of 28 and 41 °C respectively (Jones and Singels, 2018). CV% was calculated and minimised to identify an optimal *Tb_germ* and corresponding *TTem_R* value.

In general, canopy- and biomass-related parameter values were set initially based on knowledge of process implementation and E responses from the DSSAT-Canegro and/or APSIM-Sugar models. These were then refined in an iterative manner, by running the model with a changed parameter and assessing the impacts on a range of development and biomass-related variables in the 376C dataset.

Values for $FI\_OSG_{1.5}$ and $OSGfrac_{c1}$ were derived by fitting Eqn. (5-3) to observations of SDM/ADM (a measurable proxy for *OSGfrac*) and *FIinter*.

The adjustments for row-spacing were derived from analysis of trial data (Singels and Smit, 2009).

Values for *Tb_lai* and *To_lai* were initially based on published values for leaf and tiller development and expansion (Jones and Singels, 2018).

Initial guidance for determining values for $RGRlai_{max}$ and $RGRlai_{min}$ came from analysing well-calibrated DSSAT-Canegro simulations (Jones et al., 2021). Attempts were made to determine values from observations, but this was deemed unreliable due to the difficulty in estimating total leaf area index, requiring the area of senesced leaves. Values were finalised by interactive calibration, comparing (and minimising differences between) simulated and observed *GLAI*.

Ke_{min} and Ke_{max} values were taken from DSSAT-Canegro (Singels et al., 2008).

For senescence-related parameters, *GLAIst* was initially set to 3 m^2/m^2 (a "rule of thumb" value at which a sugarcane canopy is considered closed) and $LAIsen_{c1}$ was set to 0.007 (Keating et al., 1999).

Minimum and maximum SLA values (i.e. for SLA_{min} and SLA_{max} respectively) were calculated by dividing measurements of green leaf area index by green leaf canopy dry mass (i.e. the sum of leaf blade, meristem and leaf sheath dry masses), although these ranges were modified (during calibration only) where necessary to ensure accuracy of simulated canopy dry mass.

Biomass partitioning to sucrose was indirectly calibrated by ensuring that stalk dry mass and canopy mass were accurately predicted, and then modifying the stalk fibre density parameter *SSV* to predict stalk fibre dry mass and/or sucrose mass accurately.

The coupled nature of canopy development and biomass partitioning required interactive, iterative calibration for finalising canopy growth rate, senescence and radiation use efficiency parameters, by evaluating parameter values changes against FIPAR, GLAI, biomass component values, and biomass component fractions.

5.3.6 Model validation with NCo376 data

The CaneGEM model, calibrated with the trait values listed in Table 5-1, was validated against an independent NCo376 dataset ("376V", Table 5-4). The model was validated with duration of germination phase predetermined (scenario "GEM_376V_P", Table 5-4) as well as simulated by the model (scenario "GEM_376V_S", Table 5-4). The same validation experiments were run with the DSSAT-Canegro model (Jones and

Singels, 2018) to establish a model performance benchmark for comparison (scenario "DC_376V_S", Table 5-4).

5.3.7 CaneGEM model calibration for ICSM IGEP cultivars (R570, N41 and CP88-1762)

Very few data were available for calculating *Tb_germ*, *TTem_P* and *Ttem_R* from observations. It was decided to focus on varying only *Tb_germ* and keep *TTem_P* and *TTem_R* the same as NCo376. Errors in *TTem_x* parameter values were expected to manifest as poor fits for G effects, while changes to cardinal temperatures would affect GxE effects, the improvement of which is the objective of this work.

Calibration was based on the following observations from the "IGEP" dataset (Table 5-4):

- At Belle Glade P, duration of germination (*DoGP*) for R570 was 71 days, vs CP88-1762 (56 days), N41 (50 d) and NCo376 (57 d): average daily air temperature ($TAVE_{GP}$, °C) during this period was ≈ 17 °C
- At Belle Glade R1, *DoGP* for R570 = 64 days, vs NCo376 (30 d), N41 (28 days), CP88 (22 d): $TAVE_{GP}$ ≈ 19 °C
- At La Mare R1, *DoGP* for R570 = 24 d, vs N41 (18 d) and CP88 (24 d): $TAVE_{GP}$ = 27 °C
- At Pongola P, *DoGP* for R570 = 23 d, vs N41 (26 d) and CP88 (26 d): $TAVE_{GP}$ = 23 °C

At $TAVE_{GP}$ = 17 °C, the coldest site, R570 took 30% longer than the other Gs; at 19 °C, R570 took 150% longer. At 23 and 27 °C, DoTP was very similar for all three Gs. NCo376 at Belle Glade was faster than R570 but slower than the others, but same as the others at La Mare. In the NCo376 model calibration, NCo376 required 150 °Cd and 43 °Cd at *Tb_germ* = 16 °C for emergence for P and R1 crops respectively. Keeping these *TTem_x* values the same, the following *Tb_germ* parameter values were chosen: 18 °C for R570, 15 °C for N41 and CP88-1762.

Parameters *To_germ*, *Tb_lai*, *To_lai*, *Tb_sk*, *To_sk*, *Tb_photo*, *To1_photo* and *To2_photo* were all offset by the same amounts (+2 °C for R570, -1 °C for N41 and CP88-1762) relative to the standard NCo376 values (Table 5-1). *Tf_xx* values were not modified as temperatures seldom exceeded *To_xx* values.

Attempts at further improvement in prediction of G effects by modifying the cardinal temperatures further, $RGRlai_x$ or *GLAIst* only resulted in poorer prediction of GxE effects when *DoGP* was predetermined.

Satisfactory simulation outcomes were achieved without modifying the reference radiation use efficiency parameter *RUEadm* from the value calibrated for NCo376. This is different to what was done in the work described in Chapter 4, where the maximum radiation use efficiency parameter values were determined for cultivars N41, R570 and CP88-1762 for the DSSAT-Canegro, Mosicas and APSIM-Sugar models. Reference RUE is widely considered a conservative parameter and there is little consensus regarding whether or not it differs between elite sugarcane Gs. For these reasons, no attempts were made to explore calibrating *RUEadm* for more accurate simulation outcomes.

5.3.8 DSSAT-Canegro calibration for ICSM IGEP cultivars

For the simulation of the ICSM experiments (for assessing prediction performance of G, E and GxE interaction effects), the DSSAT-Canegro model was run using the parameters determined for Jones et al. (2021), Chapter 4 of this thesis.

5.3.9 Evaluation of E, G and GxE interaction effects

Simulated vs observed E, G and GxE interaction effects were assessed for the "IGEP" dataset (Table 5-4), consisting of three cultivars (N41, R570 and CP88-1762) grown under full irrigation and experimental conditions at three sites (Belle Glade, Florida, USA; La Mare, Reunion Island, France; and Pongola, South Africa; Chiredzi, Zimbabwe was omitted as *FIPAR* data were not recorded at that site).

A similar methodology for assessing E, G and GxE interaction effects as used in Jones et al. (2021) was followed, but is repeated here for convenience:

G, E and GxE interaction effects were evaluated using the following additive model:

$$X_{ij} = \bar{X} + G_i + E_j + GE_{ij} \quad (1)$$

where X_{ij} is observed or simulated seasonal photosynthetically-active radiation interception fraction (*FIPARa*) or aerial dry biomass at harvest (*ADMh*, t/ha), $\bar{X}$ is the grand mean value calculated over all environments and genotypes, *Gi* is the average value for genotype *i*, and *Ej* is the average value for environment *j*:

$$G_i = \left(\frac{1}{n_E}\sum_{j=1}^{n_E} X_{ij}\right) - \bar{X} \quad (2)$$

$$E_j = \left(\frac{1}{n_G}\sum_{i=1}^{n_G} X_{ij}\right) - \bar{X} \quad (3)$$

where n_E is the number of environments and n_G is the number of genotypes. GE_{ij} was calculated by rearranging eq. (1) and substituting the other terms.

G_i, E_j and GE_{ij} values, termed "effects", were calculated for observations and simulations.

- G_i represents the mean deviation from the grand mean, per genotype *i*, across all environments. This reflects the genotypic effect, where, over all environments considered, one genotype might, for example, intercept less radiation compared to the others (negative effect value, i.e. $G_i < 0$), or produce higher biomass (positive value), than the other genotypes.
- E_j represents the mean deviation from the grand mean, per environment *j*, across all Gs. This reflects the environmental effect, where, over all genotypes tested, one environment might, for example, result in all Gs intercepting less radiation (negative effect vlaue), or producing higher biomass (positive effect value), than the other environments.
- GE_{ii} represents deviation by individual combinations of genotype *i* and environment *j* from the grand mean, with the mean G and E effects removed. GE_{ii} therefore isolates the GxE interaction effects.

Statistical measures of similarity (described in Section 5.3.3) were then calculated for G_i, E_j and *GE* values calculated for simulations and observations. Where the R^2 value and slope (of observed vs simulated effect values) were close to 1.00, and the correlation between simulated and observed effect values was statistically significant ($p<0.05$), the model was deemed to be able to predict that effect accurately (and *vice versa*).

Observed *FIPARa* was calculated by interpolating linearly between observations of *FIPAR*, multiplying these interpolated daily *FIPAR* values by incident photosynthetically-active radiation (PAR, calculated as 50% of recorded global radiation), and expressing as a fraction of total PAR over the same period. Simulated *FIPARa* used simulated daily *FIPAR* values, limited to the periods between first and last FIPAR observations for each experiment. This differs with Jones et al. (2021), who fitted integrated Canesim (Singels and Paraskevopoulos, 2017) canopy model curves to *FIPAR* observations, and used whole-season *FIPAR* to calculate *FIPARa*; this approach was not chosen primarily due to the difficulty of identifying the "real" date of emergence in some cases, and the consequent danger of misrepresenting early observed canopy cover.

The CaneGEM and DSSAT-Canegro models were calibrated for N41, R570 and CP88-1762 using the same datasets as for their assessment for predicting GE effects. Ideally models should be evaluated on independent datasets, but as this is the only study (and dataset) of its kind, this was not possible. However, the models were calibrated to minimise differences between observed and simulated values, independently per genotype. The models were explicitly not calibrated to minimise differences between observed and simulated GxE interaction effects. The assessment of GxE interaction effects represents an independent test of the models' abilities.

5.4 Results

5.4.1 Introduction

The results of this study are divided between those derived from NCo376 datasets for calibration and validation ("376C" and "376V" respectively, Section 5.4.2), and those derived from the ICSM IGEP dataset (Section 5.4.3). The NCo376-related assessments represent traditional evaluation of the crop models, amounting to evaluation of E effects only. The ICSM IGEP assessments seek to evaluate models' abilities to predict G and GxE interaction effects in addition to E effects.

Results are listed by model assessment scenario code; these are defined in Table 5-4.

As the objective of this study was to develop an improved sugarcane model, validation results for the DSSAT-Canegro model ("DC") are presented to establish model performance benchmarks (scenario codes "DC_376V_S" and "DC_IGEP_S").

Predetermined (as recorded by experimentalists, or estimated from shoot population density data) duration of germination phase (*DoGP*, d) values, for each dataset, are listed in Table 5-5. Trait parameter values for the standard cultivar NCo376 are listed in Table 5-1. Trait parameters calibrated for cultivars N41, 570 and CP88-1762 using the ICSM IGEP experiments, that differ from the standard NCo376 trait parameter values, are shown in Table 5-6.

Table 5-5. Duration of germination (days from crop start to 50% primary shoot emergence) for crops used for CaneGEM model calibration, validation and assessment of E/G/GxE effects. "Obs. DoGP" indicates date of emergence values recorded by experimentalists, while "Est. DoGP" indicates DoGP estimated from shoot density observations; "Crop" refers to ratoon number, where R0 is a plant crop, R1 is the first ratoon and R2 the second ratoon.

Country, site and location, reference	Crop	Crop start date	Cultivar	Obs. DoGP	Est. DoGP
NCo376 datasets ("376C" and "376V")					
Pongola, South Africa "A/Growth/14"	R2	1998-03-03	NCo376		7
	R2	1998-04-08		21	
	R2	1998-05-06		40	
	R2	1998-08-06			1
	R2	1998-12-08		9	
Mount Edgecombe, South Africa "A/Growth/16"	R0	1999-12-09	NCo376		49
	R1	2000-06-06			29
	R1	2000-12-07			2
Mount Edgecombe, South Africa "A/Temp"	R1	2003-10-13	NCo376		1
	R1	2004-04-07			1
Bruyn's Hill, South Africa "A/Temp"	R1	2003-10-16	NCo376		5
	R1	2004-04-07			11
Mount Edgecombe, South Africa "A/Space"[1]	R1	2003-08-29	NCo376		12
USA, Belle Glade "ICSM IGEP"	R0	2013-12-12	NCo376	55	
	R1	2015-01-23		28	
Reunion, La Mare "ICSM IGEP"	R1	2016-01-18	NCo376		5
La Mercy, South Africa	R1	1995-03-08	NCo376		30
Pongola, South Africa "Rostron"	R1	1968-12-17	NCo376		10
		1969-02-11			1
		1969-04-08			4
		1969-06-03			1
		1969-07-29			30
		1969-09-23			1
		1969-11-18			1
		1970-01-13			2
ICSM IGEP dataset					
Reunion, La Mare	R1	2016-01-18	N41		18
			R570		24
			CP88-1762		24
South Africa, Pongola	R0	2014-03-25	N41		26
			R570		23
			CP88-1762		26
USA, Belle Glade	R0	2013-12-12	N41	50	
			R570	71	
			CP88-1762	56	
	R1	2015-01-23	N41	28	
			R570	64	

			CP88-1762	22	

[1]All row-spacings had the same DoGP

Table 5-6. Trait parameter values calibrated for R570, N41 and CP88-1762 necessary for capturing improvements in G, E and GxE interaction effects in seasonal radiation interception, with NCo376 values shown for reference.

Parameter name	Units	Description	R570	N41	CP88-1762	NCo376
Tb_germ_P	°C	Base temperature for germination	18	15	15	16
To_germ	°C	Optimal temperature for germination	30	27	27	28
Tb_lai	°C	Base temperature for leaf area index expansion	15	12	12	13
To_lai	°C	Optimal temperature for leaf area index expansion	37	34	34	35
Tb_sk	°C	Base temperature for stalk expansion	18	15	15	16
To_sk	°C	Optimal temperature for stalk expansion	37	34	34	35
Tb_photo	°C	Base temperature for photosynthesis	12	9	9	10
To1_photo	°C	Start of optimal temperature range for photosynthesis	22	19	19	20
To2_photo	°C	End of optimal temperature range for photosynthesis	42	39	39	40

5.4.2 NCo376 experimental data

CaneGEM model calibration results (NCo376 data)

Time-series scatter plots for crop development and biomass values (Figure 5-13), and biomass fractions (Table 5-7), illustrate good performance by the CaneGEM model, for the NCo376 calibration set and duration of germination (DoGP) predetermined (assessment scenario code “GEM_376C_P”). Time series PAR interception values (“PAR Int. Frac” in Table 5-7) were reasonably accurately simulated (slope = 1.26, R^2 = 0.85), partly explaining accurate simulation of ADM accumulation (slope = 0.85, R^2 = 0.78). FIPAR was underestimated for low GLAI values, despite the positive y-intercept statistic for GLAI indicating over-simulation of low values of GLAI. Low ADM values are also over-estimated, which is inconsistent with under-estimated FIPAR. Simulation performance for GLAI was relatively poor (Table 5-7).

While low values of ADM were slightly over-estimated, larger values were slightly under-estimated. This ADM over/under-estimation dynamic was inherited by the biomass pools into which ADM is partitioned. Predictions of biomass components (stalk, stalk fibre, stalk sucrose, and senesced ‘trash’ mass) were generally accurate (slopes 0.77-0.94, R^2 values 0.67-0.85), while their fractions were slightly more accurately simulated (slopes 0.93-1.12) overall. Green canopy mass was the least accurately simulated biomass fraction (slope and R^2 = 0.60).

Seasonal radiation interception (FIPARa, Figure 5-14) was adequately predicted (slope = 1.38, R^2 = 0.65, Table 5-7). DoTP was predicted 15 days short (Table 5-7, Figure 5-14) overall, which accentuated the effect of over-estimation of early ADM values on corresponding SDM values. The impact of this seemingly large error in DoTP was minimal in terms of SDM outcomes, and this is probably because a transition from the tillering phase to onset of stalk growth is simulated, rather than a step change as in DSSAT-Canegro and APSIM-Sugar. Attempts to recalibrate the model to predict DoTP more accurately were unsuccessful, as this resulted in unsatisfactory predictions of canopy cover and biomass values.

For seasonal/at-harvest values, only radiation interception (FIPARa), sucrose mass, and sucrose mass fraction showed statistically-significant relationships between simulated and observed values.

Model performance was very similar when DoGP was simulated (scenario "GEM_376C_S"; results are shown in Appendix B, Figure 11-1, Figure 11-2 and Table 11-1).

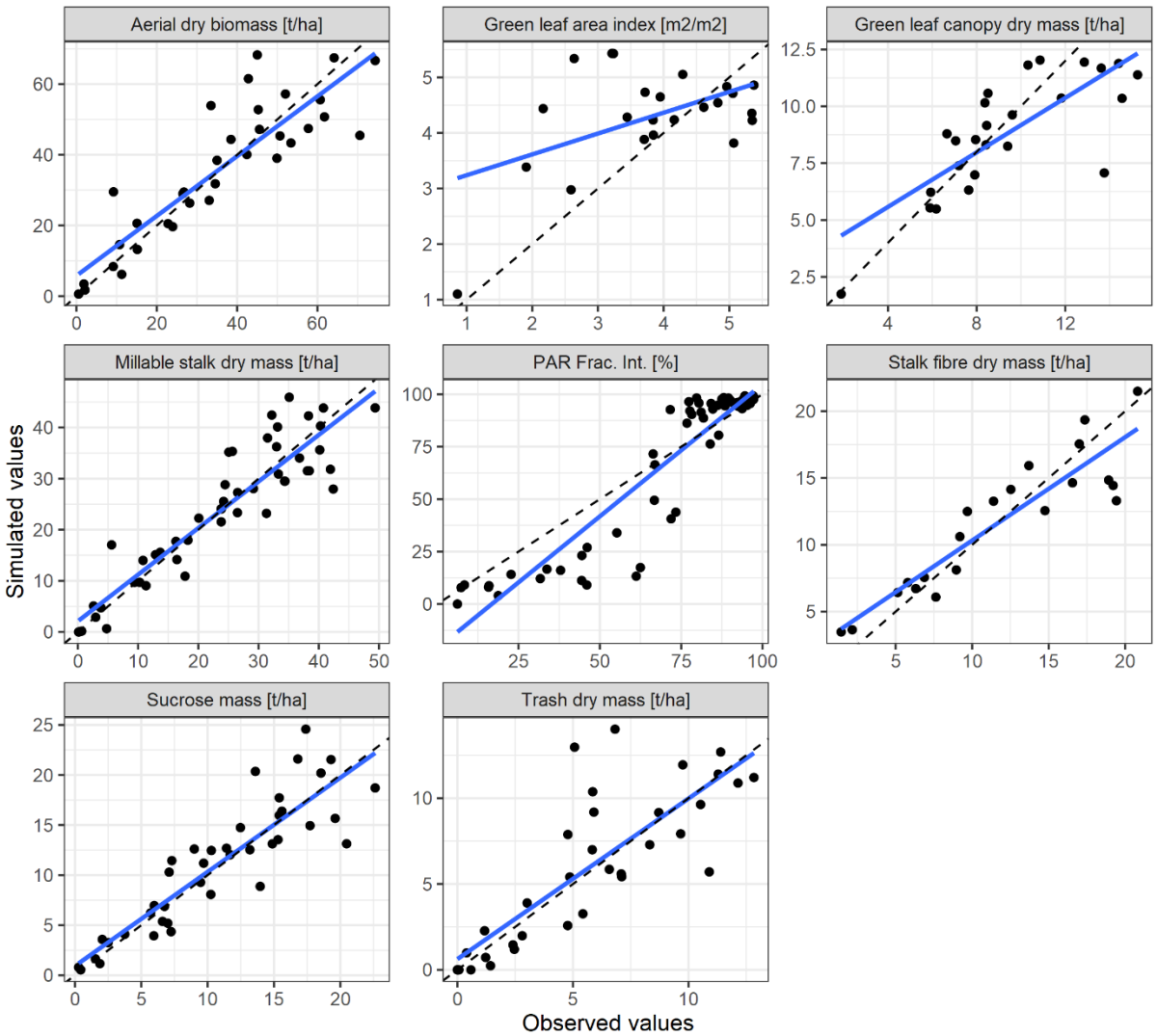

Figure 5-13. Scatter plot of time-series observed and CaneGEM model-simulated data for crop phenology- and biomass-related variables, for the calibration dataset for

NCo376. Durations of germination phases were predetermined (scenario code "GEM_376C_P" in the text).

Table 5-7. CaneGEM model performance statistics, for calibration with NCo376 data and predetermined germination phase duration (scenario code "GEM_376C_P" in the text). N = "number of observations", "Y-int" is the y intercept of linear regression between observed and simulated values, RMSE is the root mean squared error, and APE is the average prediction error (explained in the text).

Variable	N	R^2	p-value	Y-int	Slope	RMSE	APE
Time-series values							
Aerial dry biomass [t/ha]	34	0.78	0.00	5.80	0.85	9.73	0.38
Green leaf area index [m2/m2]	23	0.24	0.02	2.86	0.38	1.18	0.47
Green leaf canopy dry mass [t/ha]	25	0.60	0.00	3.18	0.60	2.13	-0.58
Millable stalk dry mass [t/ha]	51	0.85	0.00	2.10	0.91	5.24	0.07
PAR Frac. Int. [%]	59	0.86	0.00	-21.1	1.26	15.85	-2.85
Stalk fibre dry mass [t/ha]	21	0.83	0.00	2.59	0.77	2.44	-0.05
Sucrose mass [t/ha]	40	0.79	0.00	0.93	0.94	2.92	0.29
Trash dry mass [t/ha]	34	0.67	0.00	0.63	0.94	2.56	0.26
Time-series biomass fractions							
"Trash" / ADM	33	0.64	0.00	-0.02	1.12	0.05	0.00
Green leaf canopy dry mass / ADM	25	0.83	0.00	0.00	1.07	0.10	0.02
Stalk dry mass / ADM	39	0.66	0.00	0.03	0.93	0.11	-0.01
Sucrose / stalk dry mass	43	0.73	0.00	0.01	0.99	0.07	0.00
Values at harvest							
Seasonal PAR interception fraction	8	0.65	0.01	-0.25	1.38	0.09	0.03
Onset of stalk growth (d)	8	0.51	0.05	13.78	0.75	24.38	-15.6
Aerial dry biomass [t/ha]	9	0.17	0.28	35.57	0.37	12.28	1.91
Green leaf area index [m2/m2]	6	0.27	0.29	6.27	-0.42	1.46	0.89
Green leaf canopy dry mass [t/ha]	6	0.44	0.15	3.61	0.49	2.90	-1.72
Millable stalk dry mass [t/ha]	14	0.35	0.03	17.64	0.55	6.87	1.83
PAR Frac. Int. [%]	8	0.01	0.83	95.97	0.02	6.28	5.28
Stalk fibre dry mass [t/ha]	5	0.68	0.09	0.14	0.91	2.98	-1.41
Sucrose mass [t/ha]	13	0.45	0.01	6.97	0.65	3.89	1.66
Trash dry mass [t/ha]	9	0.04	0.60	12.59	-0.18	4.48	2.23
Biomass fractions at harvest							
"Trash" / ADM	9	0.04	0.59	0.24	-0.28	0.06	0.04
Green leaf canopy dry mass / ADM	6	0.00	0.90	0.16	0.02	0.08	-0.03
Stalk dry mass / ADM	10	0.42	0.04	0.50	0.21	0.08	0.00
Sucrose / stalk dry mass	14	0.34	0.03	0.09	0.85	0.07	0.02

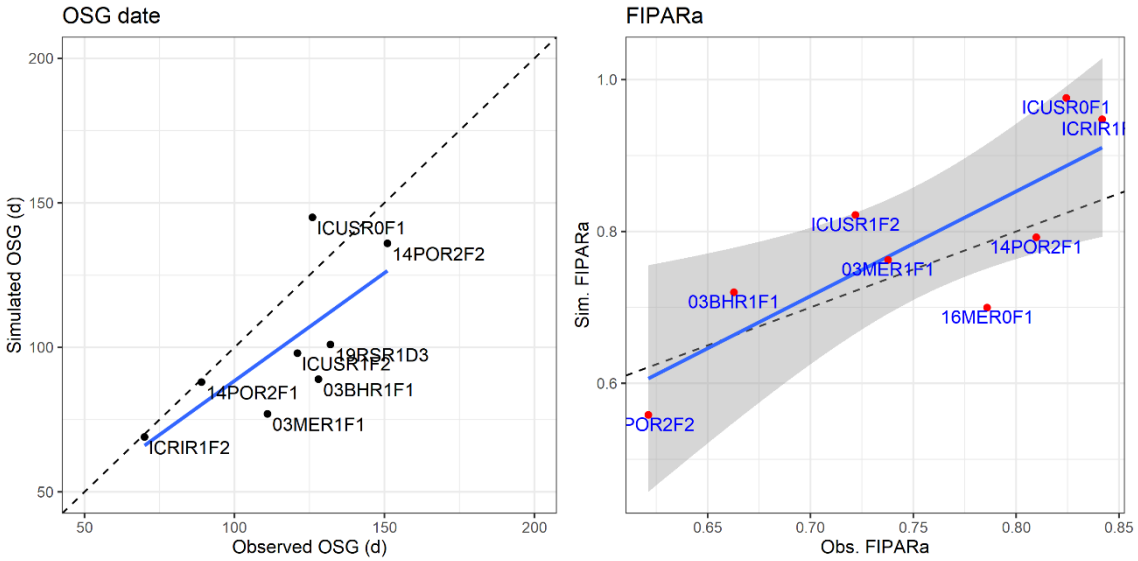

Figure 5-14. Scatter plots of age of onset of stalk growth (days since crop start) and seasonal PAR interception, simulated (CaneGEM model) and observed, for the calibration dataset for NCo376. Durations of germination phases were predetermined (scenario code "GEM_376C_P" in the text).

CaneGEM model validation results (NCo376 data)

Results for model assessment scenario "GEM_376V_P" (CaneGEM model, NCo376 validation dataset, predetermined durations of germination phase and model parameters as listed in Table 5-1) are shown in Figure 5-15 (time-series biomass and crop phenology values), Figure 5-16 (time-series biomass fractions), Figure 5-17 (DoTP and FIPARa), and Table 5-8 (model performance statistics).

Prediction of time-series FIPAR values was good (slope = 1.11, R^2 = 0.88); predictions of time-series ADM and SDM were (partly consequently) also accurate (slopes = 0.91 and 0.84, R^2 = 0.73 and 0.79, respectively). Low observed values of FIPAR (< 0.5) were generally under-estimated, while higher values were over-estimated. Similar trends in GLAI or ADM were not evident, however. RMSEs for ADM, SDM and sucrose mass were 10.5, 6.4 and 4.3 t/ha, including biases (APE) of 4.3, 1.5 and 1.5 t/ha respectively.

Predictions of time series values green leaf canopy mass fraction of ADM and sucrose fraction of stalk dry mass were accurate (slopes = 1.03 respectively, R^2 = 0.90 and 0.76 respectively). Predictions of trash mass and stalk dry mass fractions of ADM were less accurately simulated (slope = 1.41 and 0.70 respectively), although the R^2 value was reassuring high for the stalk dry mass fraction (0.85), suggesting that the variation in SDM fraction of ADM was adequately captured by the model. Predictions of values at harvest were inaccurate overall. DoTP was predicted with an R^2 value of 0.75 ($p < 0.01$), but with slope statistic of 0.49 and average under-estimation of ≈25 days on average. Trash dry mass at harvest was predicted with a slope of 0.70, but with an R^2 of only 0.30.

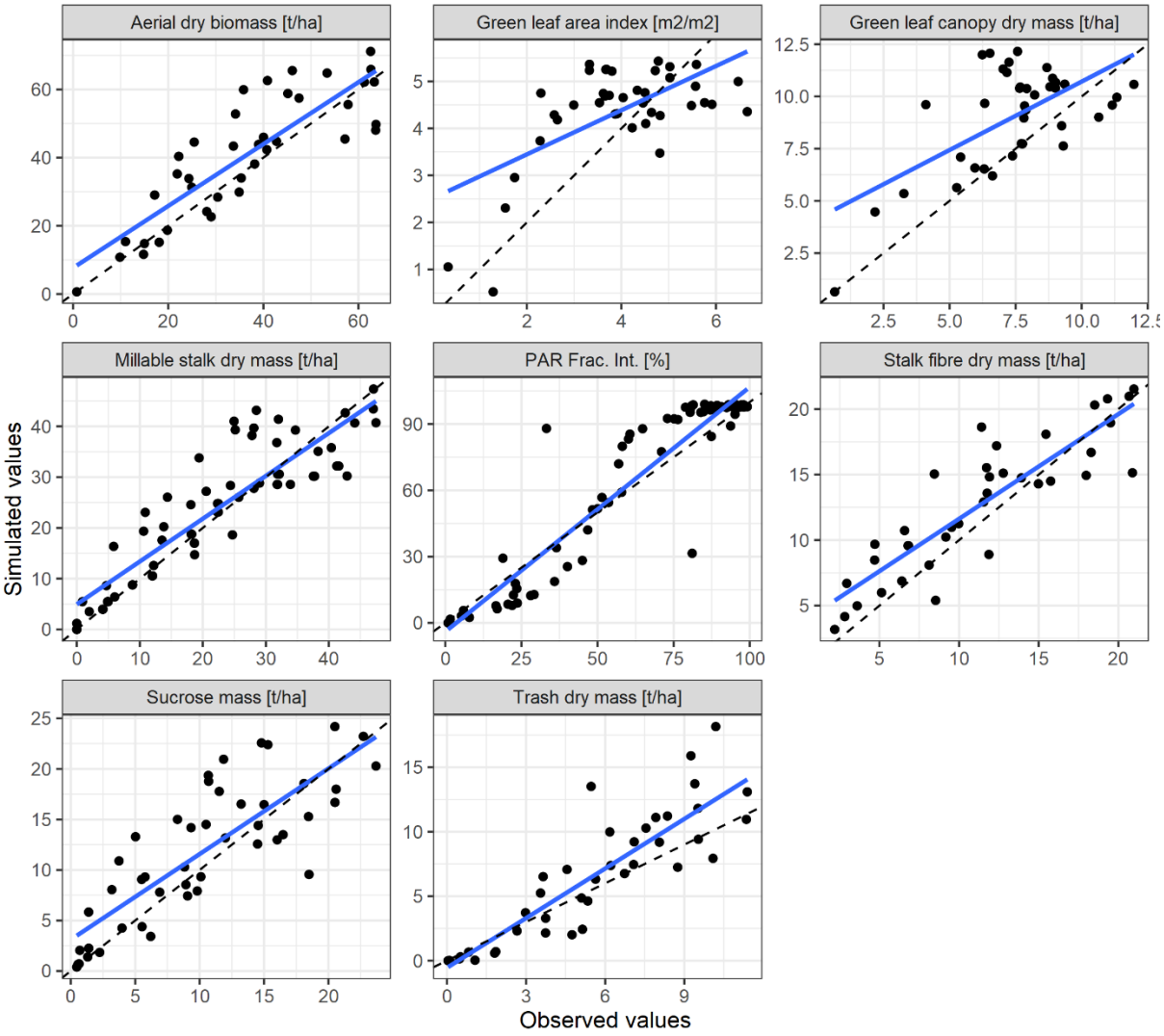

Figure 5-15. Scatter plot of time-series data for crop phenology- and biomass-related variables, for the CaneGEM model and the validation dataset for NCo376. Durations of the germination phase were predetermined (model assessment scenario code "GEM_376V_P").

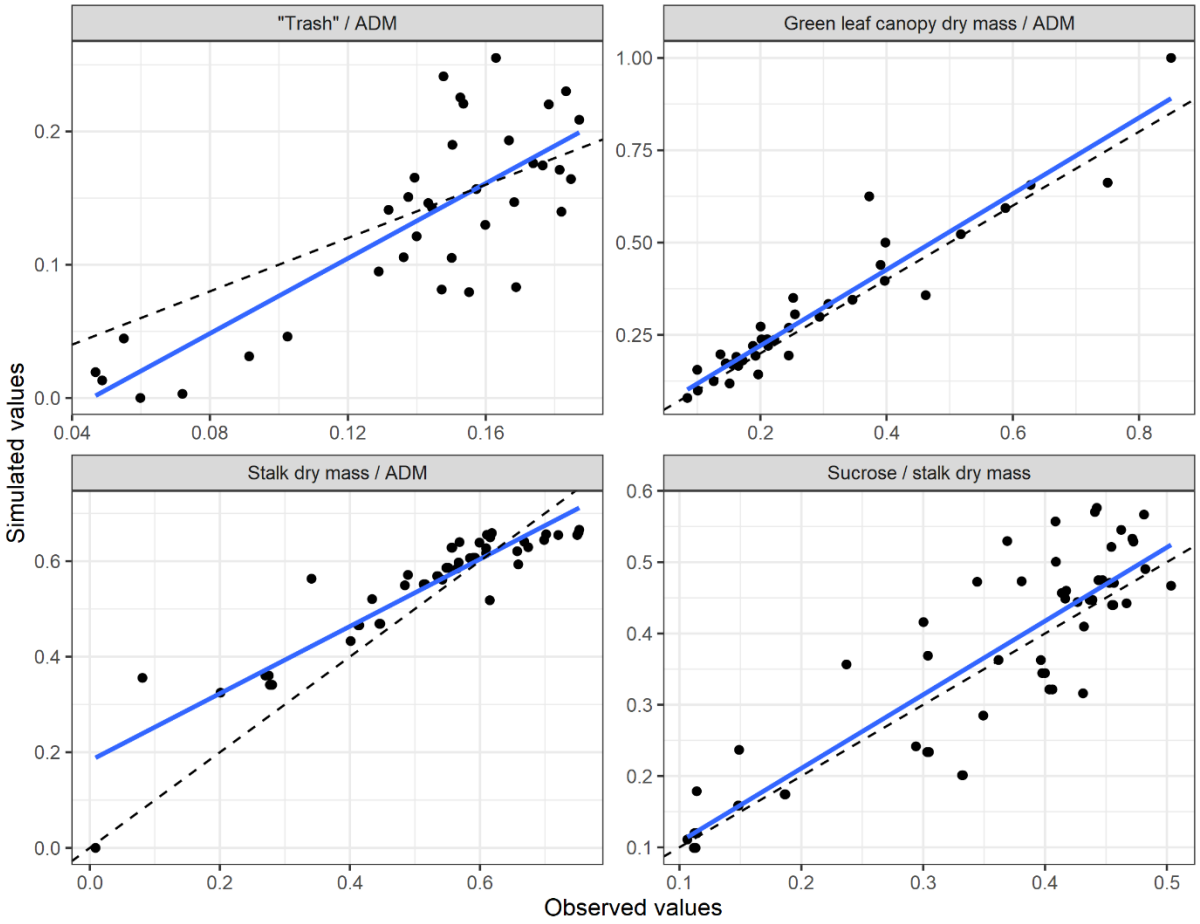

Figure 5-16. Scatter plot of time-series data, showing CaneGEM model-simulated and observed biomass fractions, for the validation dataset for NCo376. Durations of the germination phase were predetermined (model assessment scenario code "GEM_376V_P"). "ADM" is above-ground dry biomass (t/ha), while "Trash" refers to senesced leaf and stalk dry mass.

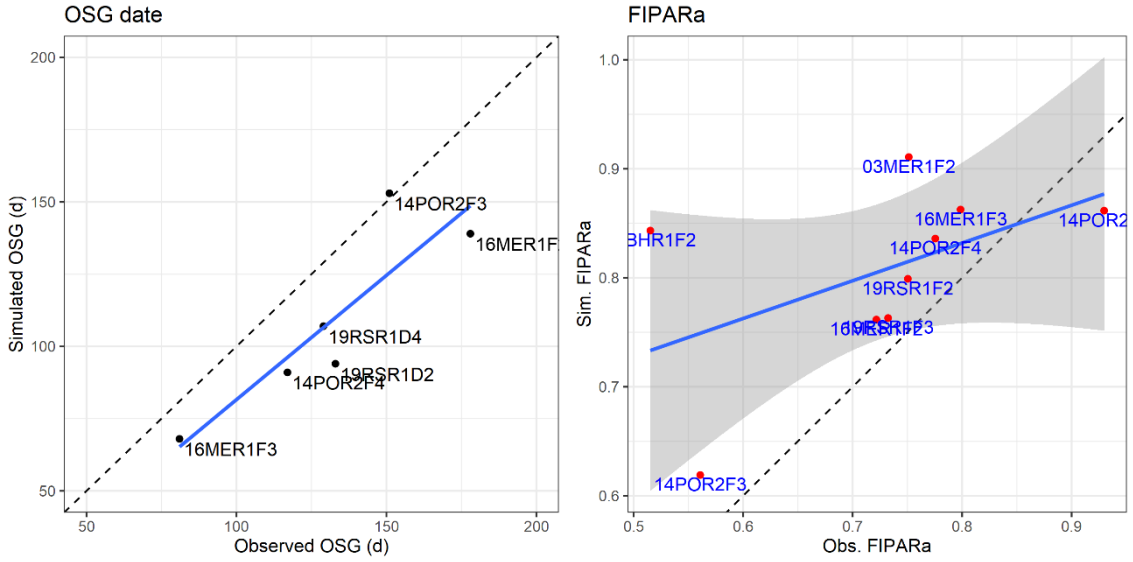

Figure 5-17. CaneGEM model-simulated vs observed duration of tillering phase (labelled as "OSG", i.e. onset of stalk growth, left), and seasonal PAR interception (FIPARa, right), for the NCo376 validation dataset. Durations of the germination

phase were predetermined (model assessment scenario code "GEM_376V_P"). The shaded region shows values within one standard error range.

Model validation performance outcomes for the CaneGEM model for simulated durations of germination phase (scenario "GEM_376V_S") are shown in Table 5-8, with supplementary figures (Figure 11-3-Figure 11-5) shown in Appendix B. Performance was very similar to that of the "GEM_376V_P" scenario (i.e. with predetermined durations of germination phase).

This model assessment scenario presents the most appropriate comparison with the DSSAT-Canegro model.

Table 5-8. CaneGEM and DSSAT-Canegro model performance statistics, for validation with NCo376 data. p = "p-value"; "Y-int", "Slp." = y-intercept and slope respectively of linear regression between observed and simulated values; "p" indicates the p-value (statistical strength of linear regression between observed and simulated values); "RMSE" and "APE" = root mean squared error and average prediction error respectively between observed and simulated values. Model assessment scenario codes "GEM_376V_P", "GEM_376V_S" and "DC_376V_S" are included here for correspondence with the presentation of results and discussion in the text.

	CaneGEM model												DSSAT-Canegro model					
	Predetermined germination phase, "GEM_376V_P"						Simulated germination phase, "GEM_376V_S"						Simulated germination phase, "DC_376V_S"					
Variable	**R^2**	**p**	**Y-int**	**Slp.**	**RMSE**	**APE**	**R^2**	**p**	**Y-int**	**Slp.**	**RMSE**	**APE**	**R^2**	**p**	**Y-int**	**Slp.**	**RMSE**	**APE**
Time-series values																		
Aerial dry biomass [t/ha]	0.73	0.00	7.68	0.91	10.49	4.33	0.73	0.00	6.13	0.90	9.64	2.36	0.75	0.00	8.29	0.84	9.05	2.61
Green leaf area index [m2/m2]	0.39	0.00	2.51	0.47	1.17	0.40	0.39	0.00	2.48	0.47	1.17	0.37	0.37	0.00	1.97	0.52	1.16	0.06
Green leaf canopy dry mass [t/ha]	0.39	0.00	4.16	0.66	2.60	1.59	0.38	0.00	4.14	0.63	2.50	1.41	0.21	0.00	4.14	0.47	2.42	0.16
Millable stalk dry mass [t/ha]	0.79	0.00	4.95	0.84	6.38	1.45	0.81	0.00	3.66	0.84	5.97	0.02	0.77	0.00	6.14	0.8	6.68	1.67
PAR Frac. Int. [%]	0.88	0.00	-4.00	1.11	13.26	3.28	0.87	0.00	-8.18	1.13	14.13	0.83	0.91	0.00	-20.45	1.25	14.70	-3.65
Stalk fibre dry mass [t/ha]	0.77	0.00	3.63	0.80	2.98	1.36	0.75	0.00	3.18	0.80	2.92	0.87						
Sucrose mass [t/ha]	0.67	0.00	3.10	0.85	4.28	1.54	0.69	0.00	2.37	0.84	3.89	0.71	0.75	0.00	4.29	0.78	3.91	2.07
Trash dry mass [t/ha]	0.77	0.00	-0.55	1.28	2.71	0.99	0.77	0.00	-0.62	1.25	2.53	0.72	0.83	0.00	1.69	0.71	1.45	0.11
Time-series biomass fractions																		
"Trash" / ADM	0.63	0.00	-0.06	1.41	0.05	-0.01	0.67	0.00	-0.08	1.51	0.05	-0.01	0.01	0.57	0.16	-0.08	0.05	0.01
Green leaf canopy dry mass / ADM	0.90	0.00	0.02	1.03	0.07	0.02	0.92	0.00	0.01	1.11	0.08	0.04	0.80	0.00	0.03	0.81	0.09	-0.03
Stalk dry mass / ADM	0.85	0.00	0.18	0.70	0.07	0.03	0.86	0.00	0.13	0.77	0.06	0.01	0.69	0.00	0.19	0.74	0.11	0.05
Sucrose / stalk dry mass	0.76	0.00	0.00	1.03	0.07	0.02	0.75	0.00	-0.03	1.09	0.08	0.00	0.89	0.00	0.04	1.02	0.06	0.04
Values at harvest																		
Seasonal PAR interception fraction	0.25	0.17	0.55	0.35	0.13	0.08	0.29	0.13	0.58	0.28	0.11	0.06	0.72	0.00	0.22	0.7	0.06	0.00
Onset of stalk growth (d)	0.75	0.00	45.3	0.49	45.7	-25.7	0.86	0.00	51.1	0.50	39.8	-18.9	0.79	0.00	73.8	0.44	40.7	-8.6
Aerial dry biomass [t/ha]	0.10	0.40	46.50	0.22	14.75	7.32	0.13	0.34	42.68	0.26	13.63	5.35	0.16	0.28	41.65	0.23	11.95	2.72
Green leaf area index [m2/m2]	0.00	0.89	4.56	0.02	1.30	0.84	0.00	0.90	4.57	0.02	1.30	0.85	0.01	0.85	4.45	0.02	1.20	0.73

	CaneGEM model												DSSAT-Canegro model					
	Predetermined germination phase, "GEM_376V_P"						Simulated germination phase, "GEM_376V_S"						Simulated germination phase, "DC_376V_S"					
Variable	**R^2**	**p**	**Y-int**	**Slp.**	**RMSE**	**APE**	**R^2**	**p**	**Y-int**	**Slp.**	**RMSE**	**APE**	**R^2**	**p**	**Y-int**	**Slp.**	**RMSE**	**APE**
Green leaf canopy dry mass [t/ha]	0.10	0.45	5.83	0.47	2.91	2.17	0.07	0.54	6.32	0.38	2.87	2.07	0.01	0.79	9.90	-0.09	2.91	2.37
Millable stalk dry mass [t/ha]	0.21	0.10	26.99	0.30	8.62	3.10	0.26	0.06	23.91	0.35	7.91	1.70	0.16	0.16	28.39	0.21	8.28	1.48
PAR Frac. Int. [%]	0.10	0.42	101.5	-0.03	5.76	3.58	0.11	0.38	101.9	-0.04	5.76	3.55	0.38	0.08	84.4	0.15	5.26	3.65
Stalk fibre dry mass [t/ha]	0.45	0.05	9.55	0.47	3.75	1.39	0.41	0.06	9.09	0.47	3.76	0.95						
Sucrose mass [t/ha]	0.20	0.15	14.09	0.33	6.14	4.11	0.30	0.07	11.77	0.43	5.31	3.23	0.37	0.03	13.36	0.29	4.78	2.77
Trash dry mass [t/ha]	0.30	0.16	5.91	0.72	4.51	3.60	0.33	0.14	5.59	0.72	4.16	3.28	0.28	0.17	5.99	0.25	2.05	-0.22
Biomass fractions at harvest																		
"Trash" / ADM	0.01	0.81	0.16	0.25	0.05	0.03	0.04	0.65	0.12	0.48	0.05	0.03	0.09	0.47	0.12	0.20	0.02	-0.01
Green leaf canopy dry mass / ADM	0.69	0.01	0.02	0.96	0.03	0.02	0.66	0.01	0.03	0.97	0.04	0.02	0.08	0.50	0.16	0.11	0.05	0.03
Stalk dry mass / ADM	0.54	0.01	0.47	0.25	0.06	0.00	0.74	0.00	0.45	0.29	0.05	0.00	0.22	0.15	0.55	0.17	0.07	0.03
Sucrose / stalk dry mass	0.18	0.13	0.26	0.53	0.07	0.05	0.20	0.10	0.18	0.68	0.07	0.04	0.18	0.13	0.39	0.22	0.06	0.04

DSSAT-Canegro model validation results (NCo376 data)

Validation results for model assessment scenario "DC_376V_S", are shown in Figure 5-18 (time-series biomass and crop phenology values), Figure 5-19 (time-series biomass fractions), Figure 5-20 (DoTP and FIPARa), and Table 5-8 (model performance statistics).

Considering time-series (daily) values, DSSAT-Canegro underestimated low values of FIPAR and over-estimated high values (slope = 1.25, intercept = -20.45), although it accounted well for FIPAR variations across experiments (R^2 = 0.91) (Table 5-8). Low values of ADM were over-estimated and high values underestimated, despite the FIPAR simulation error pattern. This over-/under-estimation pattern was (consequently) evident in other biomass-related variables, such as SDM, sucrose and trash dry mass yields. The R^2 for ADM accumulation was poorer than that of FIPAR at 0.75. RMSEs for ADM, SDM and sucrose mass were 9.1 t/ha, 6.7 t/ha, and 3.91 t/ha respectively. R^2 values for biomass component variables were generally in the range 0.75-0.91, with the exception of green leaf canopy mass (R^2 = 0.21; slope = 0.47). GLAI prediction performance was poor (slope = 0.52, R^2 = 0.37).

DSSAT-Canegro model performance was considerably poorer in predicting values at harvest than values through the season. Only two variables, FIPARa and DoTP, had statistically significant slopes (0.70 and 0.44 respectively), with good R^2 values (0.72 and 0.79). Predicted DoTP was 9 days too short on average. Biomass fractions at harvest were poorly predicted, with low R^2 values, and low slope statistics.

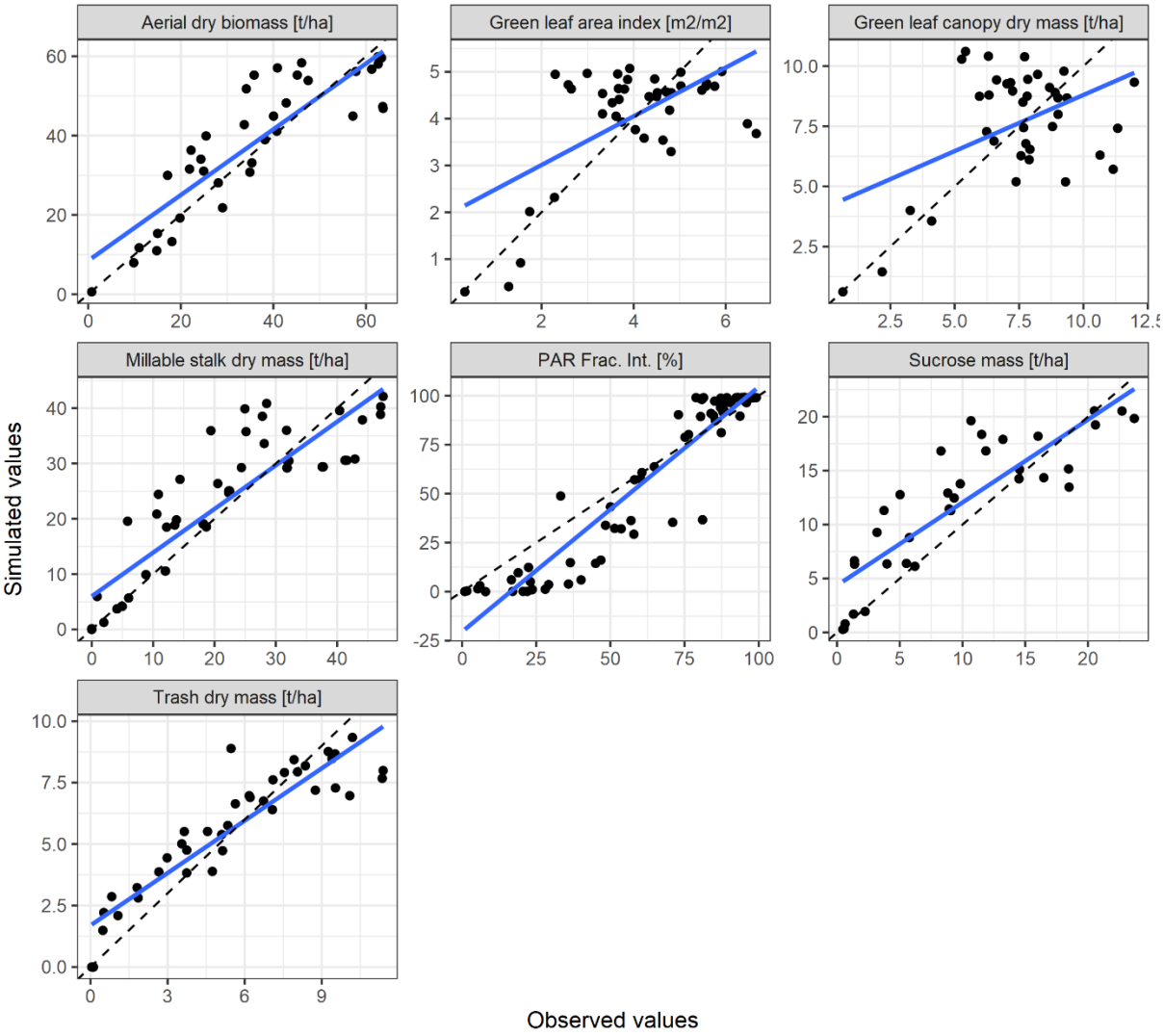

Figure 5-18. Scatter plot of simulated (DSSAT-Canegro) and observed time-series data for crop phenology- and biomass-related variables, for the NCo376 validation dataset. Durations of germination phase were simulated by the model (model assessment scenario code "DC_376V_S"). "PAR Frac. Int [%]" is fractional interception of photosynthetically-active radiation.

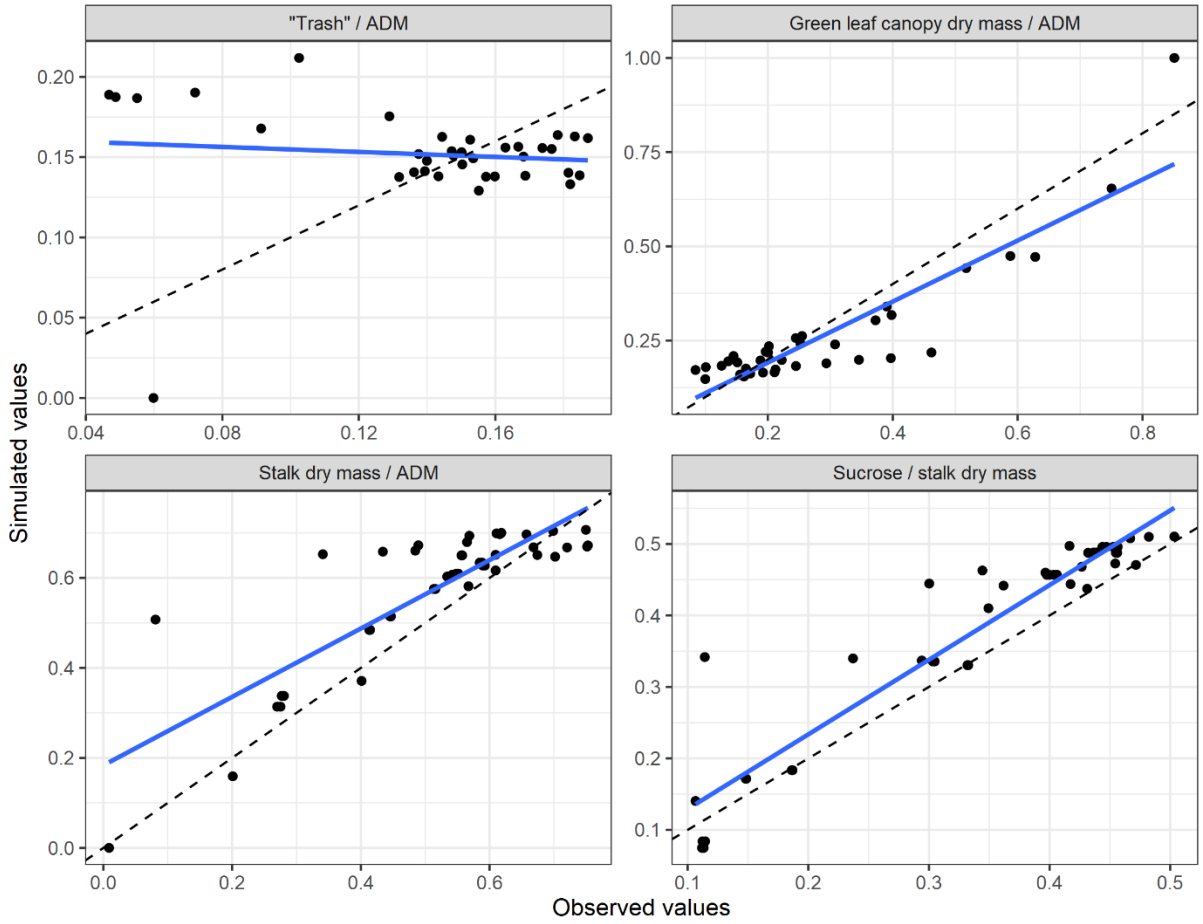

Figure 5-19. Scatter plot of time-series data, showing simulated (DSSAT-Canegro) and observed biomass fractions, for the validation dataset for NCo376. Durations of germination phases were simulated by the model (model assessment scenario code "DC_376V_S"). "ADM" is above-ground dry biomass (t/ha), while "Trash" refers to senesced leaf and stalk dry mass.

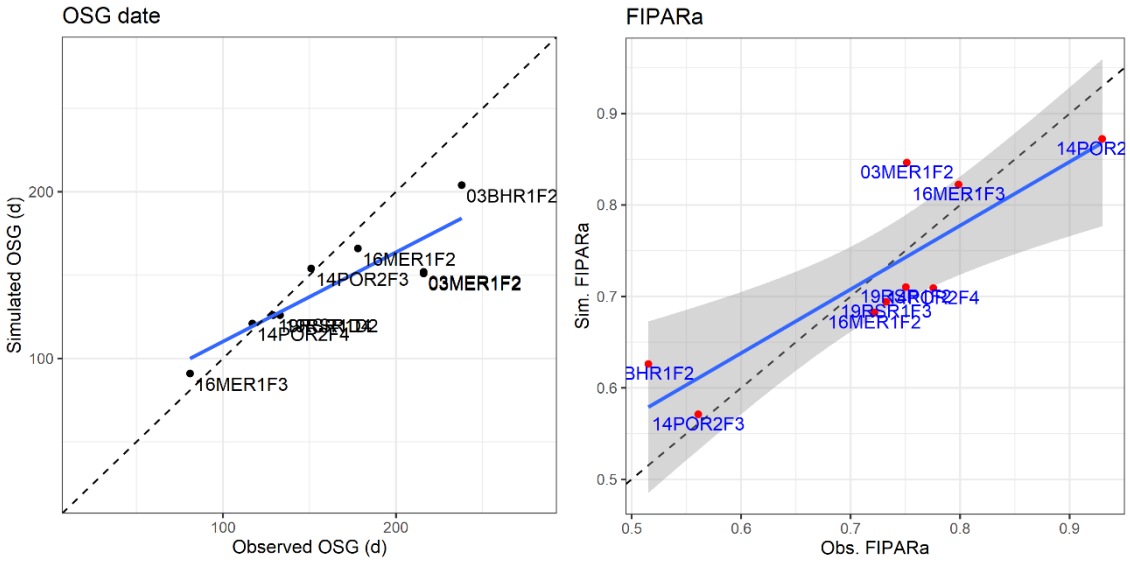

Figure 5-20. Simulated (DSSAT-Canegro) vs observed duration of tillering phase (DoTP, labelled as "OSG date", left), and seasonal PAR interception (FIPARa, right), for the NCo376 validation dataset. Durations of germination phases were simulated

by the model (model assessment scenario code "DC_376V_S"). The shaded region shows values within one standard error range.

Comparison of DSSAT-Canegro and CaneGEM validation performance

The two models gave generally very similar performance (Table 5-8, and Figure 5-21 which graphically compares R^2 values for time-series data) in simulating the NCo376 validation dataset. The CaneGEM model fared slightly better than DSSAT-Canegro for simulating time-series biomass fractions. The DSSAT-Canegro model was able to simulate seasonal radiation interception more accurately than CaneGEM.

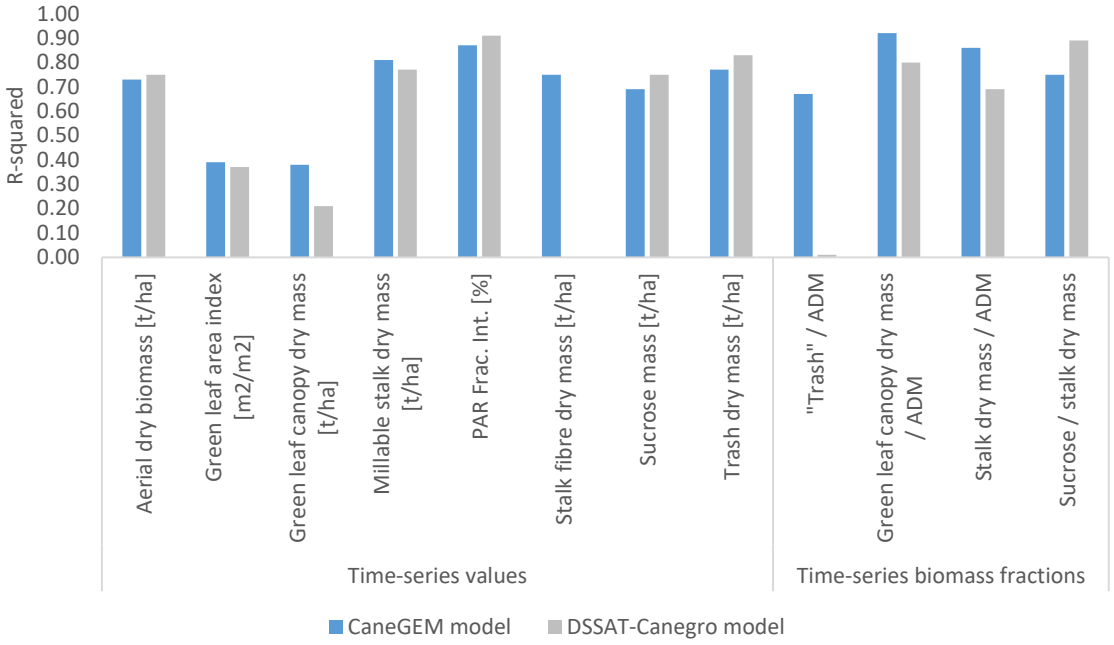

Figure 5-21. Comparison of R^2 values for observed vs simulated time-series data, simulated by CaneGEM and DSSAT-Canegro for the NCo376 validation dataset, with durations of germination phase simulated by the models (model assessment scenario codes "GEM_376V_S" and "DC_376V_S", explained in the text).

5.4.3 ICSM IGEP experimental data

DSSAT-Canegro, ICSM IGEP dataset, simulated date of emergence and calibrated G-specific canopy parameters ("DC_IGEP_S")

Daily simulated (using the DSSAT-Canegro model) and observed FIPAR and ADM values for the ICSM IGEP experiments (model assessment scenario code "DC_IGEP_S") are shown in Figure 5-22 and Figure 5-23 respectively. Simulated and observed E, G and GxE effects are shown in Figure 5-24, while performance statistics for simulating these effects are listed in Table 5-9.

E and GxE interaction effects on FIPARa were poorly captured with analysis yielding negative relationships between simulated and observed values. G effects were simulated very accurately (as was the intention with the calibration procedure described in Chapter 4), however. Despite this, there was a statistically-significant but weak positive relationship between simulated and observed GxE interaction effects on ADMh (slope = 0.07, R^2 = 0.52, $p < 0.05$); E and G effects on ADMh were simulated

more accurately than for FIPARa, although these relationships were not statistically significant.

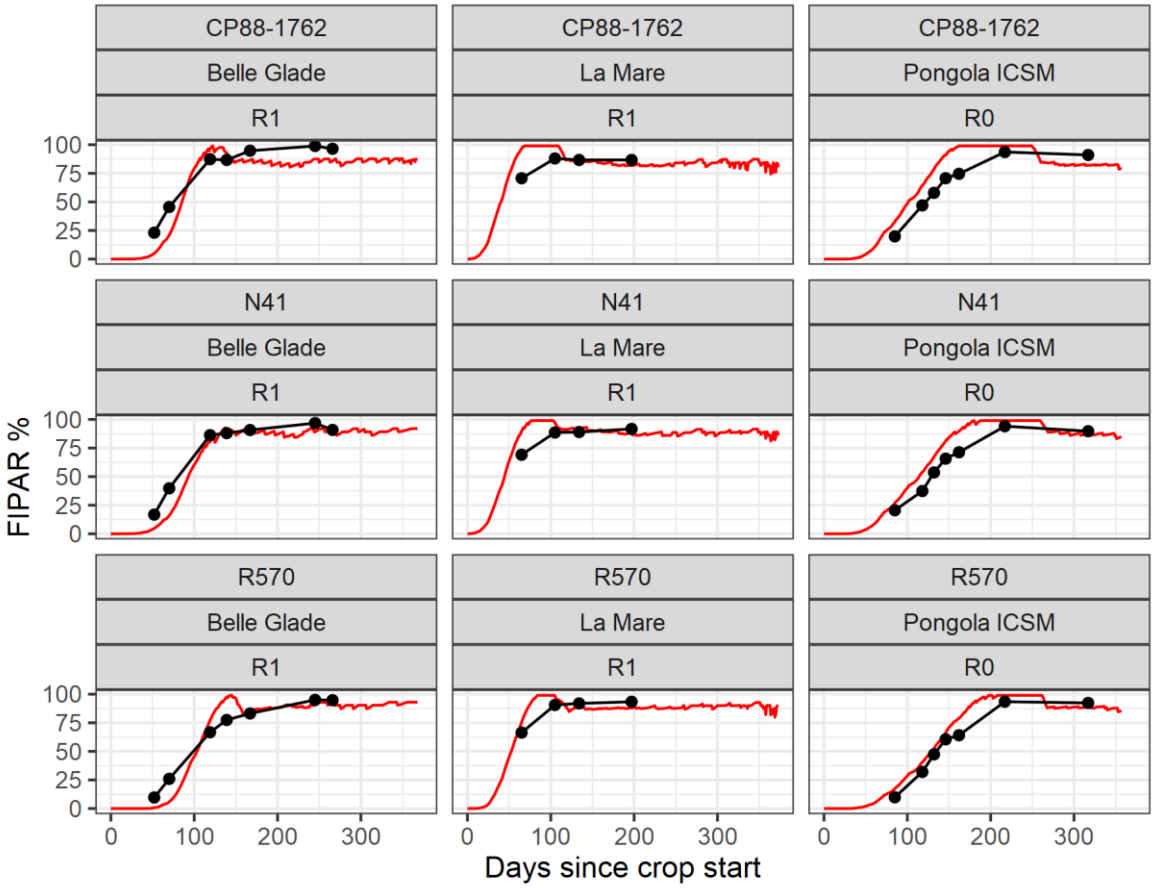

Figure 5-22. Simulated (by DSSAT-Canegro, red line) and observed (black points and linearly-interpolated lines) fractional interception of photosynthetically-active radiation (FIPAR, %), for the non-stressed ICSM IGEP experiments, with simulated duration of germination phase (model assessment scenario code "DC_IGEP_S").

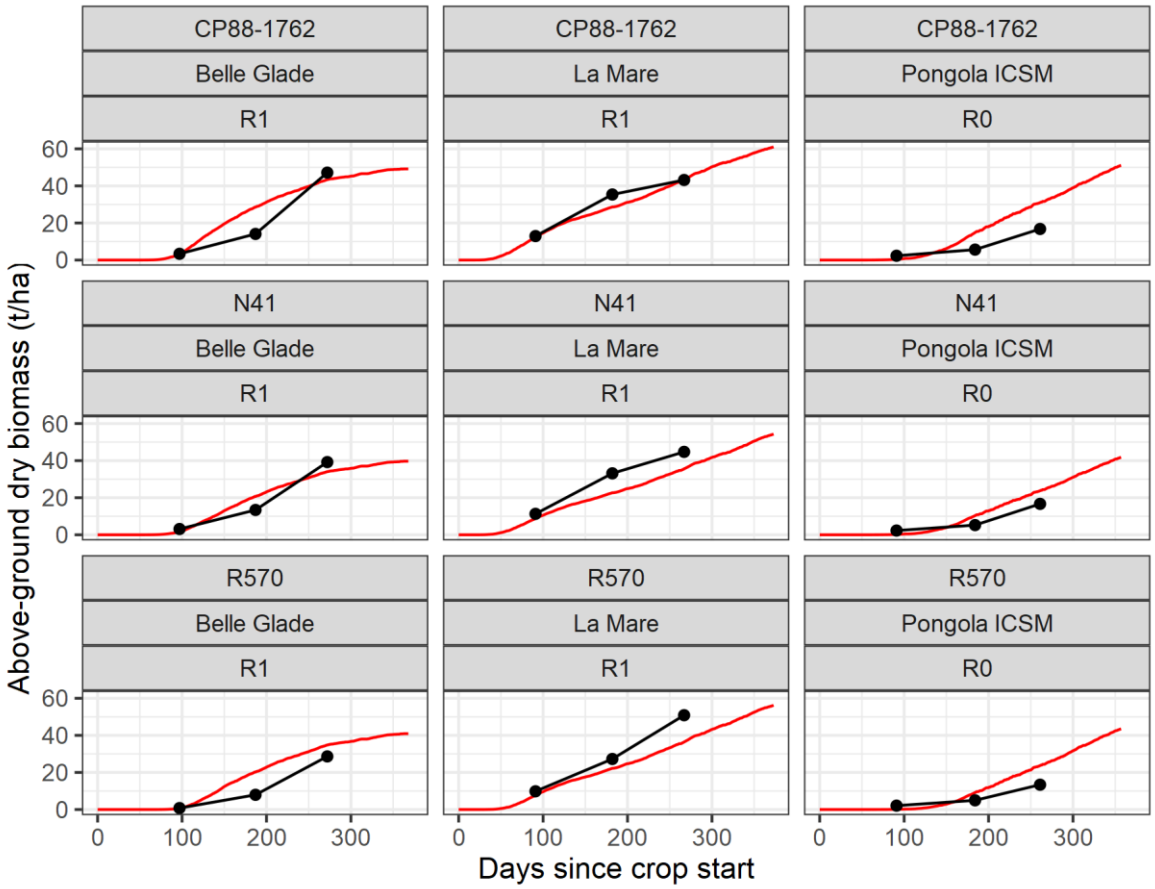

Figure 5-23. Simulated (by DSSAT-Canegro, red line) and observed (black points and linearly-interpolated lines) above-ground dry biomass (t/ha), for the non-stressed ICSM IGEP experiments, with simulated duration of germination phase (model assessment scenario code "DC_IGEP_S").

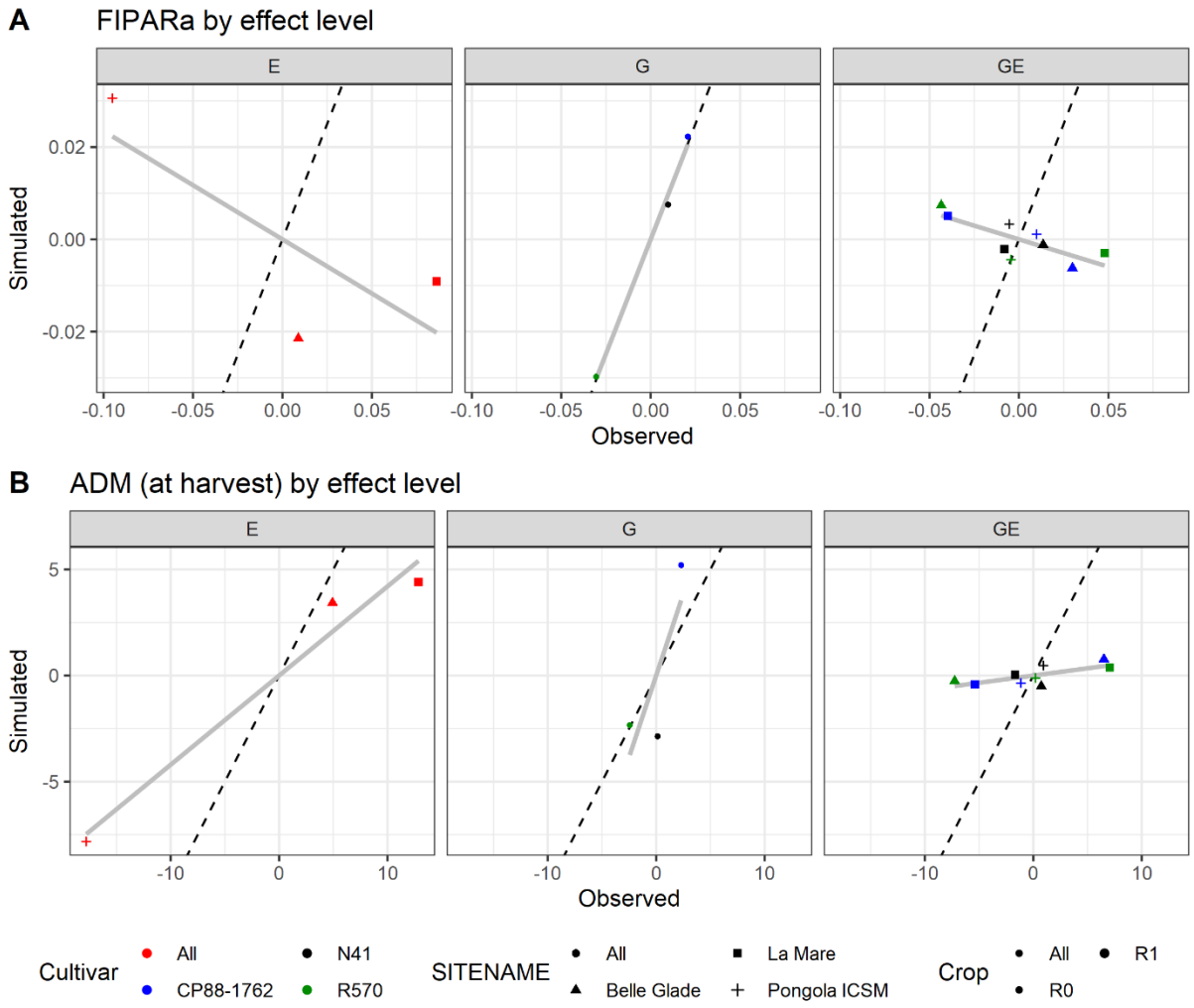

Figure 5-24. E, G and GxE interaction effects on seasonal photosynthetically-active radiation interception (FIPARa) and above-ground dry biomass at harvest (ADM), simulated (DSSAT-Canegro model) vs observed. ICSM IGEP dataset, simulated duration of germination phase (model assessment scenario code "DC_IGEP_S").

Table 5-9. DSSAT-Canegro model accuracy for predicting environmental (E), genotypic (G) and genotype-by-environment interaction (GE) effects on seasonal fractional interception of radiation (FIPARa) and aerial dry biomass at harvest (ADMh, t/ha). ICSM IGEP dataset, simulated duration of germination phase (model assessment scenario code "DC_IGEP_S"). "Y-int" = the y-intercept of the linear regression between observed and simulated values; "N" = the number of data points considered.

Variable	Level	Slope	R^2	p-value	N
FIPARa	E	-0.23	0.62	0.43	3
	G	0.99	0.99	0.05	3
	GE	-0.12*	0.60	0.01	9
ADMh	E	0.42	0.97	0.11	3

	G	1.54	0.65	0.40	3
	GE	0.07*	0.52	0.03	9

*P<0.05

CaneGEM model, NCo376 parameter values and predetermined of durations of germination phase, ICSM IGEP dataset

CaneGEM model FIPAR prediction performance with predetermined DoGP and NCo376 parameter values (model assessment scenario code "GEM_IGEP_P1") was satisfactory (Figure 5-25). Model predictions best-matched observations of N41, which is understandable as this cultivar was selected in the same breeding programme as NCo376. In several cases early FIPAR was underestimated and mid-season FIPAR over-estimated.

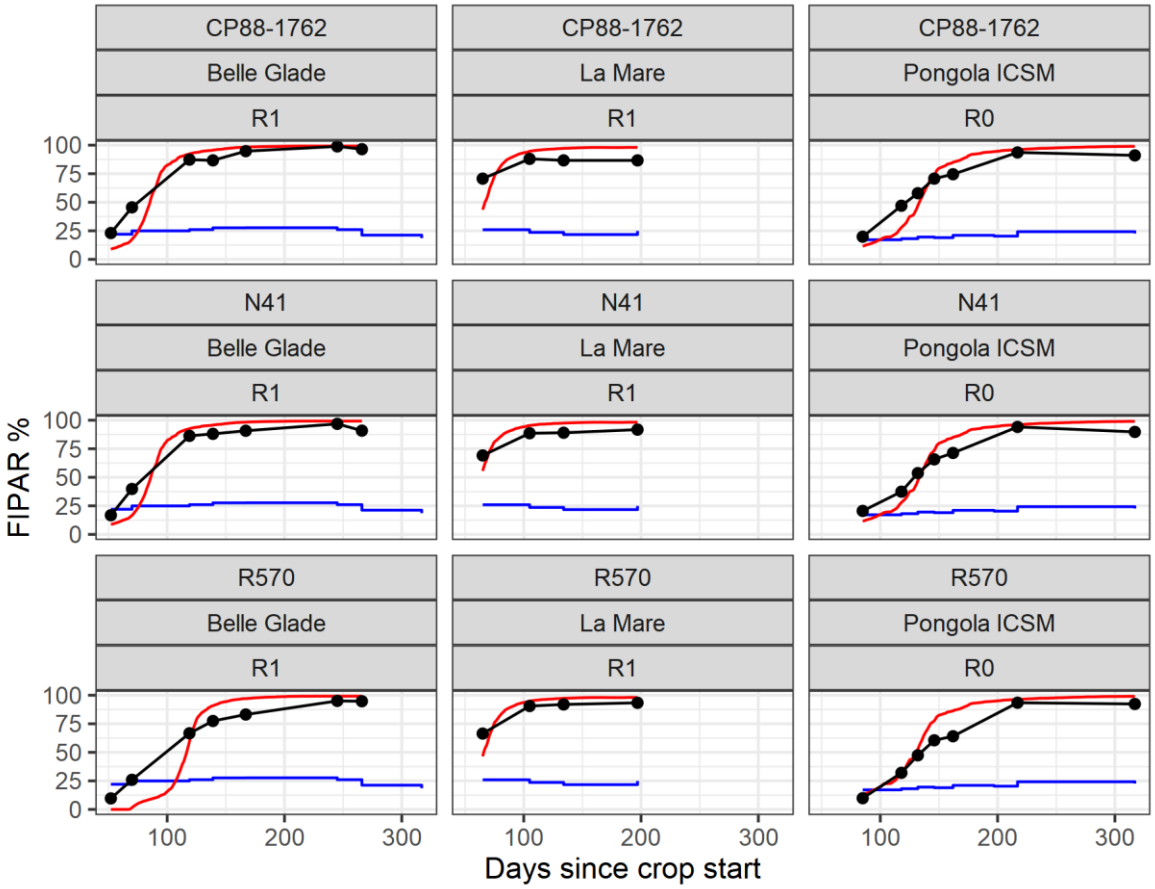

Figure 5-25. CaneGEM-simulated (red line) and observed (black points and linearly-interpolated lines) FIPAR, with predetermined durations of germination phase and NCo376 cultivar parameters for the ICSM IGEP dataset (model assessment scenario code "GEM_IGEP_P1"). Inter-sample period average daily air temperature is also shown (stepped blue line). Differences in simulated values between Gs are due to G-E differences in predetermined durations of germination phase.

FIPARa was under-estimated at Belle Glade and over-estimated at La Mare (Figure 5-26 "E"); and over-estimated for N41. E and G effects on FIPARa and ADMh appeared reasonably accurately simulated, but were not statistically significant (Table 5-10). GxE interaction effects on FIPARa were very accurately simulated and strongly statistically-significant (p<0.01). The GxE interaction effect on ADMh was also

accurately predicted and statistically-significant ($p<0.05$), suggesting that a significant part of the variation in observed ADMh is explained by G-per-E variation in DoGP.

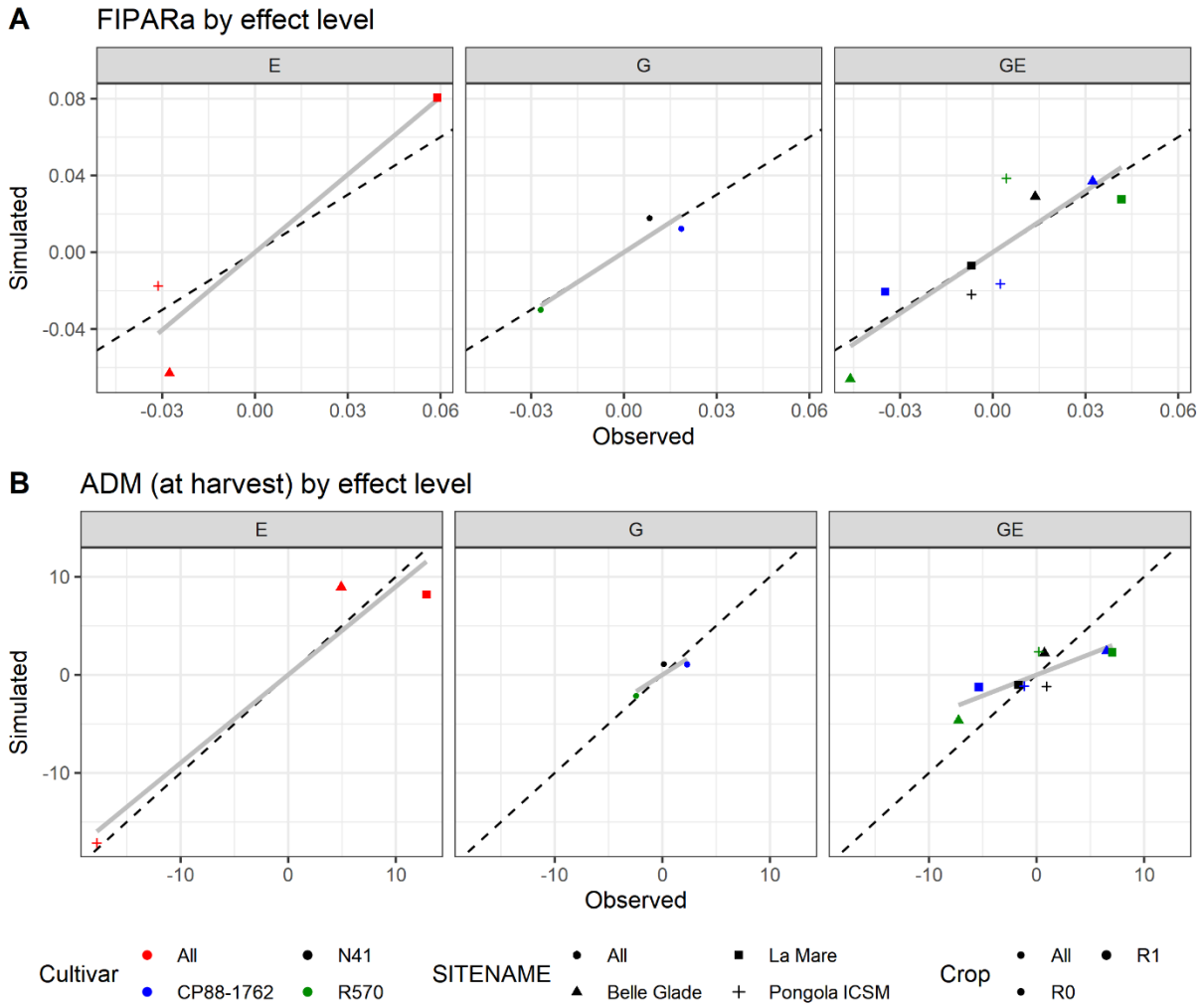

Figure 5-26. E, G and GxE interaction effects, simulated vs observed, with predetermined durations of germination phase and NCo376 cultivar parameters for the ICSM IGEP dataset (model assessment scenario code "GEM_IGEP_P1"). Trait parameter values are otherwise identical (NCo376 parameter values used).

Table 5-10. CaneGEM model accuracy for predicting environmental (E), genotypic (G) and genotype-by-environment interaction (GE) effects on seasonal fractional interception of radiation (FIPARa) and aerial dry biomass at harvest (ADMh, t/ha), for three cultivars (R570, N41 and CP88-1762), for the ICSM IGEP dataset. Durations of germination phase were predetermined (model assessment scenario code "GEM_IGEP_P1"). NCo376 trait parameter values were used.

Variable	Level	Slope	R^2	p-value	N
FIPARa	E	1.35	0.88	0.22	3
	G	1.04	0.90	0.20	3
	GE	1.06**	0.72	0.00	9
ADMh	E	0.90	0.92	0.18	3

	G	0.69	0.78	0.31	3
	GE	0.43*	0.67	0.01	9

**p<0.01, *p<0.05

CaneGEM model, with calibrated cultivars and predetermined date of emergence, ICSM IGEP dataset

CaneGEM model performance, following calibration for R570, N41 and CP88-1762 (model assessment scenario code “GEM_IGEP_P”), for time-series values, is shown in Figure 5-27. A graph of daily FIPAR values is shown in Appendix B (Figure 11-6).

The changes to cardinal temperatures (which were based on changes to Tb_germ) resulted in considerable improvements (reduced scatter) in GxE interaction effects predictions.

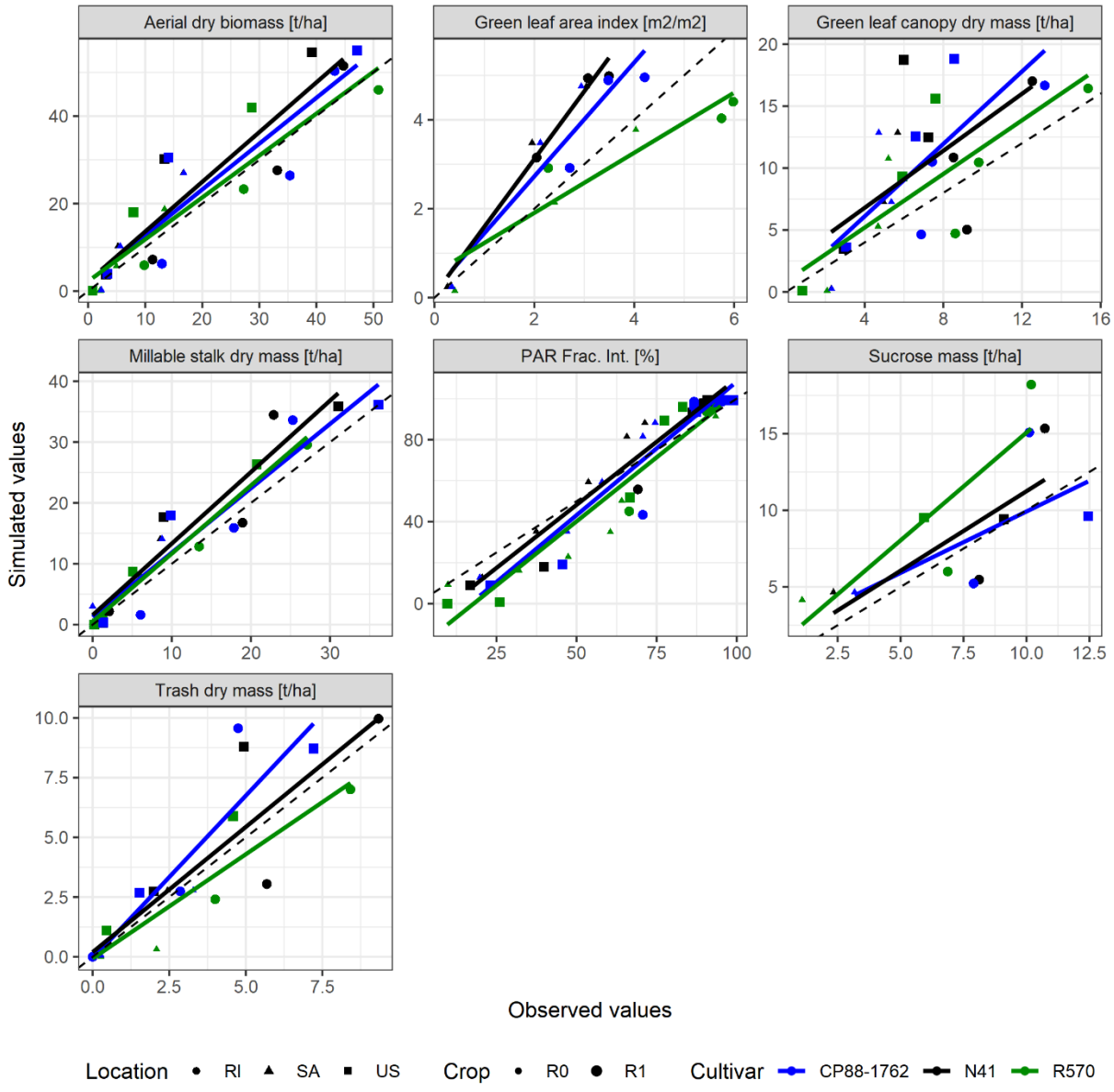

Figure 5-27. Simulated (CaneGEM model, with predetermined date of emergence and G-specific calibration of canopy development parameters, model assessment

scenario code "GEM_IGEP_P") vs observed crop development and biomass components, for the ICSM IGEP dataset.

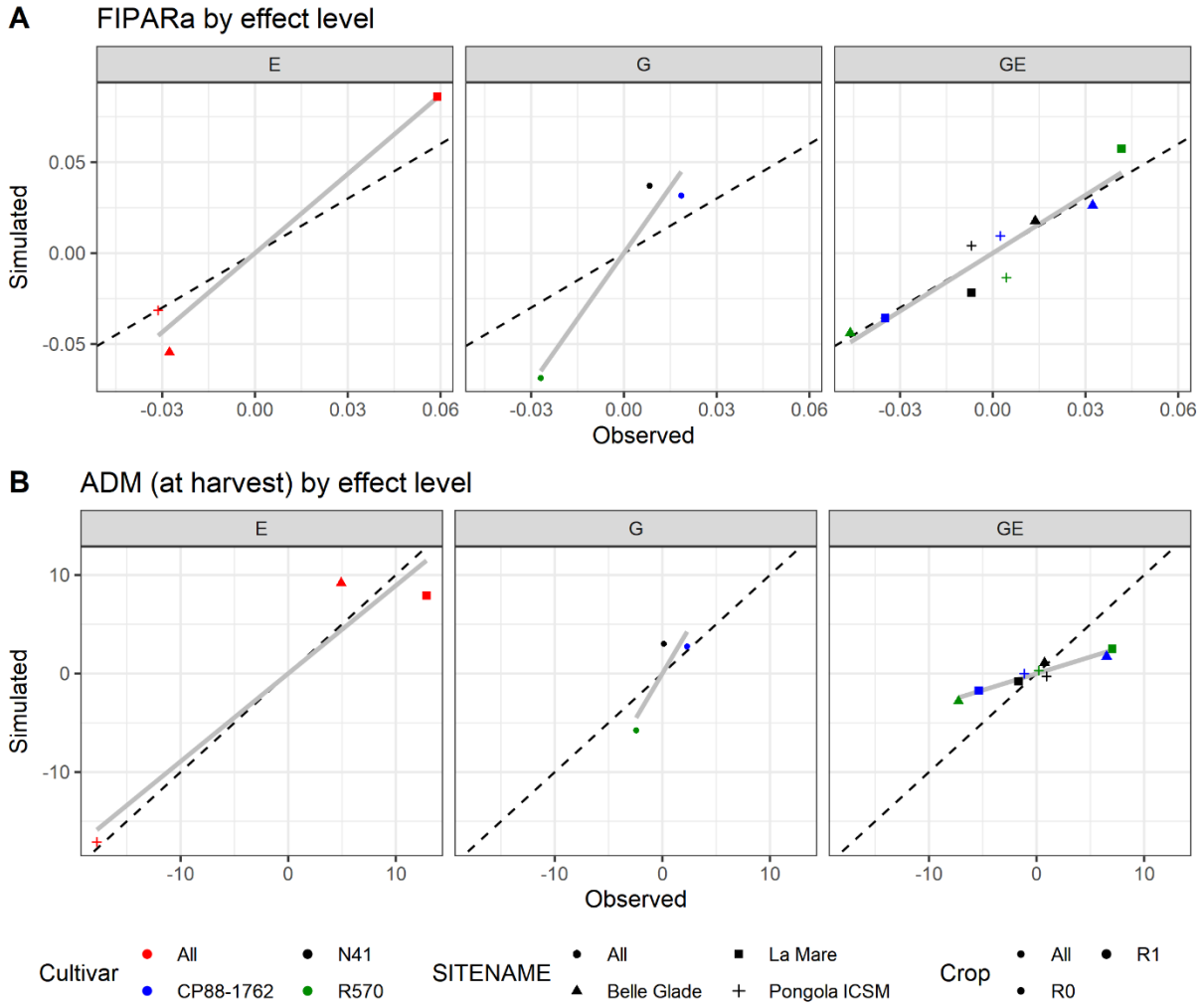

Figure 5-28. E, G and GxE interaction effects, CaneGEM-simulated vs observed, with predetermined duration of germination and G-specific calibration of canopy development parameters, for the ICSM IGEP dataset (model assessment scenario code "GEM_IGEP_P").

Table 5-11. CaneGEM model (with predetermined duration of germination and G-specific calibration of canopy development parameters, model assessment scenario code "GEM_IGEP_P") accuracy for predicting environmental (E), genotypic (G) and genotype-by-environment interaction (GE) effects on seasonal fractional interception of radiation (FIPARa) and aerial dry biomass at harvest (ADMh, t/ha).

Variable	Level	Slope	R^2	p-value	N
FIPARa	E	1.45	0.96	0.12	3
	G	2.41	0.93	0.17	3
	GE	1.06**	0.88	0.00	9
ADMh	E	0.89	0.92	0.19	3
	G	1.85	0.77	0.32	3
	GE	0.34**	0.92	0.00	9

**p<0.01

CaneGEM model, calibrated cultivars with simulated durations of germination phase, ICSM IGEP dataset

CaneGEM model performance results for model assessment scenario "GEM_IGEP_S" are shown Figure 5-29 (E, G and GxE interaction effects on FIPARa and ADMh) and Table 5-12 (statistics summarising performance of simulated E, G and GxE interaction effects). Additional in results are provided in Appendix B, Figure 11-7 (daily FIPAR), Figure 11-8 (crop development and biomass scatter plots).

The CaneGEM model's ability to predict GxE interaction effects in FIPARa and ADMh were greatly reduced when the duration of the germination phase was simulated rather than predetermined. This is further evidence that GxE interaction effects in this dataset was driven primarily by GxE differences in the duration of the germination phase.

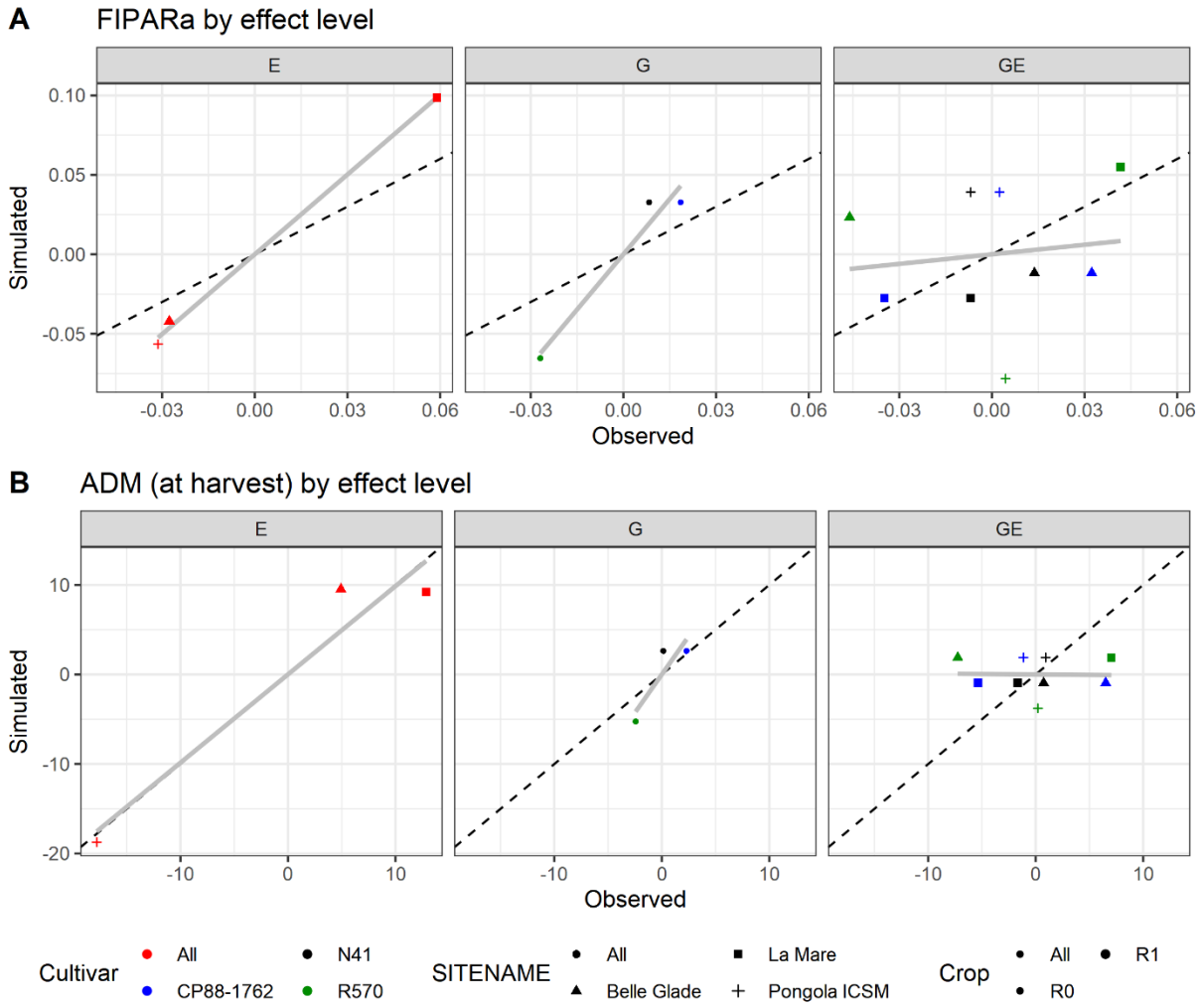

Figure 5-29. E, G and GxE interaction effects, CaneGEM-simulated vs observed, with simulated duration of germination phase and G-specific calibration of cardinal temperatures controlling canopy development and germination rate (model assessment scenario "GEM_IGEP_S").

Table 5-12. CaneGEM model (with simulated duration of germination phase and G-specific calibration of canopy development parameters, model assessment scenario

"GEM_IGEP_S") simulation accuracy for predicting environmental (E), genotypic (G) and genotype-by-environment interaction (GE) effects on seasonal fractional interception of radiation (FIPARa) and aerial dry biomass at harvest (ADMh, t/ha). "N" = number of observations.

Variable	Level	Slope	R^2	p-value	N
FIPARa	E	1.67*	1.00	0.03	3
	G	2.32	0.95	0.14	3
	GE	0.20	0.02	0.73	9
ADMh	E	0.99	0.93	0.17	3
	G	1.71	0.79	0.30	3
	GE	-0.01	0.00	0.95	9

*p<0.05

5.5 Discussion

5.5.1 Novel canopy development and biomass partitioning algorithms are sound, and conventionally-assessed performance of the CaneGEM model is similar to that of DSSAT-Canegro

The DSSAT-Canegro model simulated the NCo376 validation dataset ("376V") very well, which is testament to that model's pedigree. The time-series RMSEs for ADM, SDM and sucrose mass (9.1, 6.7 and 3.9 t/ha respectively are only slightly greater than published values (e.g. 8.4, 5.6 and 3.3 t/ha, Jones et al. (2018)). This appears to be the first time that DSSAT-Canegro has been validated with irrigated experiments only, and this an excellent result because the overriding impact of water stress, combined with well-studied water relations reflected in the model algorithms, makes rainfed/water-stressed experiments generally easier to simulate accurately. Model performance for values and biomass fractions at harvest was generally poor, with the exceptions of FIPARa and DoTP.

For the NCo376 calibration set ("376C"), the CaneGEM model's prediction of FIPAR and FIPARa were reasonably good, and the simulation of time-series ADM, SDM and sucrose mass values was also satisfactory. CaneGEM was unable to capture the dynamics of high values of green leaf area index, although this evidently had little impact on the simulation of radiation interception; this may however have affected the accuracy of green canopy mass simulations, which were considerably less accurately predicted than other biomass components. The poor prediction of values at harvest is of some concern, although it is acknowledged that capturing subtle differences in final fully-irrigated yields is difficult, especially given the possible presence of lodging and unrecorded end-of-season "drying-off" (intentional water stress to enhance sucrose content, Dias et al. (2018)). Similar performance outcomes for predetermined and simulated duration of germination (DoGP) approaches was reassuring.

The performance of the CaneGEM model for the NCo376 validation dataset ("376V") was similar to that of the calibration dataset. RMSEs for ADM, SDM and sucrose mass (10.5, 6.4 and 4.3 t/ha respectively for predetermined DoGP, and 9.6, 6.0 and 3.9 t/ha for predicted DoGP) compare favourably with the DSSAT-Canegro validation results presented here, published DSSAT-Canegro results, and published model performance for APSIM-Sugar (summarised in Jones & Singels (2018)).

Although the predictions of seasonal/at harvest values were poor, they were generally no worse than those of DSSAT-Canegro. DSSAT-Canegro however simulated DoTP (based on thermal time) more accurately than the FIPAR-based approach in the

CaneGEM model, and also simulated FIPARa more accurately. Jones et al. (2019) reported that TT-based DoTP prediction was unreliable for the ICSM IGEP experiments. This also unfortunately calls into question one of the key features of the CaneGEM model, where DoTP is linked to canopy development, which is also connected to the carbon balance (and was intended to capture the reported effect (Jones et al., 2019) of DoTP being earlier under high radiation relative to temperature conditions, as well as conclusions from Singels & Smit (2009)). Despite the DoTP prediction error, time-series values of SDM accumulation and SDM/ADM fractions appear to have been accurately simulated in CaneGEM.

The source:sink-based biomass partitioning between leaves and stems, and between stalk fibre, sucrose and hexoses, in CaneGEM appeared to work well, producing mostly accurate results.

Three possible weaknesses are acknowledged: firstly, the majority of calibration experiments and all the validation experiments were from two relatively temperate locations in South Africa (Pongola and Mount Edgecombe); secondly, NCo376 is no longer widely grown – calibration and validation with a more popular cultivar might be desirable, although NCo376 remains a 'reference' cultivar, both in scientific research and breeding in the South African industry; and thirdly, in some cases low radiation interception values were under-estimated and intermediate values over-estimated. Green leaf area index and ADM accumulation did not show similar trends, however.

The CaneGEM model appears to have generally equivalent performance to DSSAT-Canegro, but with a considerably simpler model structure, clearer inter-process coupling (which can prevent inappropriate/unrealistic parameter combinations) and linkages to genotypic traits, and a greater emphasis on emergent simulation of complex traits. Given the long pedigree of the DSSAT-Canegro model simulating NCo376, the performance outcomes for the CaneGEM model are considered satisfactory.

5.5.2 Simulation of GxE interaction effects on canopy development and biomass accumulation is improved in the CaneGEM model

Results for DSSAT-Canegro's prediction of GxE interaction effects yielded a statistically significant ($p < 0.05$) slope of 0.07 and an R^2 of 0.52 for ADMh. Although the relationship is statistically significant, the very low value of the slope statistic means that this has very little practical utility in model application.

With DoGP predetermined, the CaneGEM model was able to predict GxE interaction effects on FIPARa and ADMh considerably more accurately than DSSAT-Canegro in this study, and in comparison with the evaluation presented in Chapter 4.

With DoGP predicted, however, the CaneGEM model was unable to predict GxE interactions in ADMh as accurately as DSSAT-Canegro, although the prediction of FIPARa GxE effects was arguably better (having a statistically-insignificant positive slope of regression between simulated and observed FIPARa GxE interaction effects, compared to a significant negative slope for DSSAT-Canegro).

At Belle Glade R1, R570 intercepted less radiation than CP88-1762 had lower ADMh yields; and at La Mare R1, R570 intercepted more radiation than CP88-1762 and had higher yields. In GxE analysis terms, these are classic crossover effects (Figure 5-30).

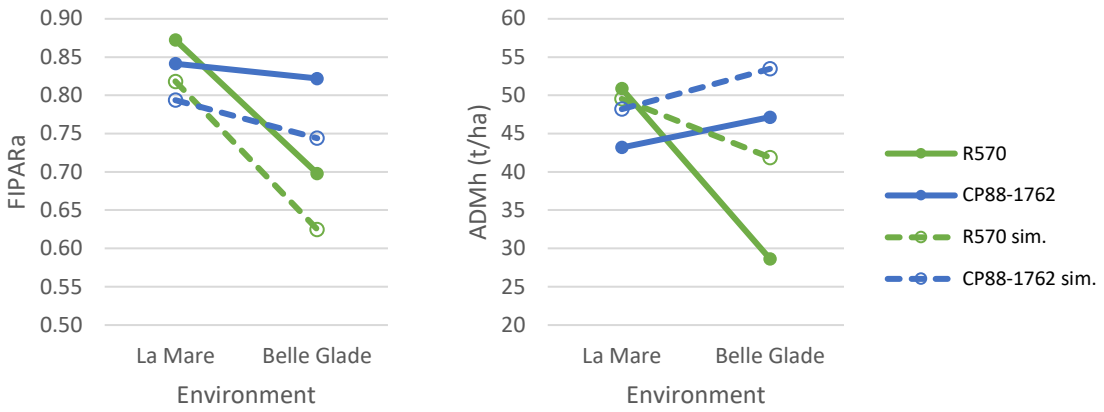

Figure 5-30. Observed (solid lines) and CaneGEM-simulated (dashed lines) genotype-by-environment (GxE) crossover effects for seasonal intercepted photosynthetically-active radiation fraction (FIPARa, left) and above-ground dry biomass yields at harvest (ADMh, right), for two Gs (R570 and CP88-1762) and two Es (La Mare, Reunion Island; and Belle Glade, Florida, USA).

These crossover effects were predominantly explained by the difference in duration of the germination phase for R570:

- at La Mare, R570 and CP88-1762 were deemed to have emerged on the same date – 24 days after crop start. The higher canopy growth rate for R570 drove it to intercept slightly more radiation than CP88-1762, driving higher seasonal radiation interception. Conditions at crop start at La Mare were warm.
- At Belle Glade, R570 took considerably longer than CP88-1762 to complete the germination phase and emerge – 71 vs 56 days. It was relatively cool at Belle Glade at crop start.
- Simulations with CaneGEM accounted for the differences in canopy development rate between the Gs (via differences in trait parameter values for relative canopy development rate and cardinal temperatures for canopy development), but could not adequately predict the GxE differences in duration of germination phase on the basis of different cardinal temperatures for thermal time driving germination rate. Hence, it was necessary to specify the duration of the germination phase in order to predict the seasonal radiation interception differences sufficiently accurately to drive an accurate prediction of GxE effects in ADMh yields.

With the assistance of the CaneGEM model and the ICSM IGEP dataset, GxE differences in ADMh yields were therefore dissected into G-specific temperature response and fundamental canopy development rate traits, and unresolved GxE influences on duration of germination phase.

Dias et al. (2020) determined G-specific canopy-related model input parameter values (green leaf number, tillering factor, thermal time from date of primary shoot emergence to date of onset of stalk growth and radiation extinction coefficient) for APSIM-Sugar, for 27 varieties and two Es in Brazil; their conclusion was that simulation outcomes matched qualitative information about the varieties considered, and that the calibrated model could be used for exploring GxExM interactions. Duration of germination was not explored in this study, and relied on the standard APSIM-Sugar thermal time-based germination model. It is noted that 21 of the 27 crops used in that study were plant

crops, for which this standard germination model appears to function more robustly than for ratoon crops.

5.5.3 GxE interactions for date of emergence drive GxE in canopy cover and biomass yields, but prediction of germination rate is inadequate

Running the CaneGEM model with predetermined DoGP per G and E, and NCo376 trait parameters (model assessment scenario code “GEM_IGEP_P1”), resulted in accurate and statistically-significant prediction of GxE interaction effects in FIPARa and ADMh.

Perhaps the most important finding of this work is then that GxE interaction effects in irrigated sugarcane appear to be mostly explained by GxE interactions in duration of germination phase. Germination phase duration is therefore a key trait determining environmental fitness for irrigated sugarcane, and could possibly be used as a screening trait in selection in a breeding programme.

Implementing germination rate screening would require that duration of germination is phenotyped, which could be achieved by: germinating individual buds in temperature-controlled germination rooms or growth chambers and recording time to emerge (Poser et al., 2019; Smit, 2010); or by counting shoot density per unit area in a field trial scenario (Bezuidenhout et al., 2003; Donaldson, 2009); it may be possible to assess emergence status via remote sensing (Gers, 2003; Som-ard et al., 2021). This could only practically be included from the second selection stage onwards, as phenotyping germination rate from true seed, as undertaken in the first stage, may not correspond with germination rate from setts or stools.

The CaneGEM model was unable to predict DoGP well enough to capture the GxE interactions in FIPARa and ADMh that the DoGP GxE differences appeared to have caused. The model uses a standard thermal time-based approach to predicting germination rate, based on air temperatures. Attempts to use simulated soil temperatures were inconclusive, although plant crop germination rate appeared to be somewhat more accurately predicted by simulated soil temperature than air temperature (results presented in Appendix A). The *Tb_germ* value of 16 °C chosen for NCo376 aligns with reported findings. Smit (2010) reported *Tb_germ* values for South African cultivars (NCo376=18.1, N16=16.8, N27=17.8 °C), while Poser et al. (2019) reported *Tb_germ* values of 11-13 °C for Reunion cultivars (including R570), for plant crops germinated in growth chambers. Their *Tb_germ* value for R570 was considerably lower than the value calibrated in the present study, but this is attributed to dissimilarity between field and growth chamber conditions; more importantly, R570 had a higher Tb than the other Gs in that study, consistent with the present study. *Tb_germ* of 16 °C was reported for NCo376 by Jones et al. (2018), although 10 °C was recommended by Singels et al. (2008).

Efforts to optimise *Tb_germ* based on observed or best-estimated date of emergence resulted in inconsistent outcomes between plant and ratoon crops, and also between air and simulated soil temperatures (results presented in Appendix A, Figure 10-1). The small number of data points, combined with coefficient of variation values of ≈ 35% for R570 plant crops across a wide range of air temperatures, indicates poor statistical strength for these outcomes. Owing to this, a reasoned approach (described in Section 5.3.7) to determining values for cardinal temperatures was followed. It is noted however that the cardinal temperature values determined and used in the evaluation of model GxE prediction abilities are not clearly consistent with the CV%-minimising

Tb_germ temperatures shown in Figure 10-1.The real challenge does however seem to lie with the prediction (and indeed, measurement) of DoGP of ratoon crops; this appears less well-studied; Keating et al. (1999) assumed *Tb_germ* = 9 °C and *TTem_R* = 100 °Cd (with additional time required for coleoptile elongation to the soil surface, 0.8 mm/°Cd), while Bezuidenhout et al. (2003) used similar values for their dynamic model of tillering. The *Tb_germ* of 9.5 °C, where *TTem_R* variation was minimized in the present study, is very close to the APSIM-Sugar base temperature value of 9 °C. The high degree of variation (≈ 100%) is due to many ratoon crops effectively having emerged on the first day of the crop. Singels & Bezuidenhout (2002) adjusted dates of emergence to match simulated stalk mass dynamics for 18 NCo376 crops; in seven of these crops, the date of emergence was calculated to be between 15 and 82 days *earlier* than the date of crop start. It is clear that conditions prior to the start of ratoon crops need to be considered in the prediction of DoGP; such an algorithm is not included in any sugarcane crop growth models.

5.5.4 Differential temperature responses appear to drive remaining GxE interactions, and can be inferred from germination rate

Changes to cardinal temperatures based on insights from DoGP resulted in improved prediction of GxE interaction effects. Such changes were also necessary to generate GxE interaction effects, whereas changing phase durations or activity rates per unit thermal time would have resulted only in G effects.

Smit (2010) and Poser et al. (2019) reported G differences in base temperatures for germination rate in sugarcane. Donaldson (2009) calculated base temperatures for biomass component mass accumulation rate in three South African cultivars. Significant differences were reported for base temperatures for canopy dry mass (12-14 °C), stalk fibre mass (16-19 °C) and sucrose mass (16-19 °C). By contrast, Smit and Singels (2006) did not find stable differences in base temperatures for stalk elongation rate for three South African sugarcane cultivars, suggesting that genotypes differed more reliably in their elongation rates per unit thermal time. There was however considerable variation in base temperatures between sites and start dates, however, suggesting unresolved E effects.

In other species, Tirfessa et al. (2020) found G differences in base and optimal temperatures for plant phenological development in sorghum, which (like sugarcane) is a (sub-)tropical C4 grass. A range of base, optimal and ceiling temperatures for germination rates of genotypes of groundnut and pearl millet have been reported (Mohamed et al., 1988a), and the same authors found close correlations between germination rate and leaf growth rates in pearl millet (Mohamed et al., 1988b), suggesting a possible cardinal temperature coupling across plant processes. Ortiz et al. (2022) reported genotypic differences in photosynthetic performance in response to temperature in soybean. Parent et al. (2010) found evidence for coordination of temperature responses across processes in many species, including sugarcane.

Overall, it appears that there is evidence for G differences in cardinal temperatures controlling phenological development. Indeed, the results of the present study underline the importance of differences in the duration of the germination phase in explaining GxE interaction effects in biomass yields. Although it was not possible to capture the germination rate dynamics with enough accuracy to simulate these effects, there is no doubt that temperature has a powerful influence on germination rate. Consistent cardinal temperature differences for biomass accumulation and expansive growth rates have not been demonstrated for sugarcane.

The present study however revealed improved prediction of GxE interaction effects in biomass yields when cardinal temperature differences identified in germination rates were carried through to expansive processes.

Phenotyping temperature sensitivity for sugarcane field crops can be practically challenging, requiring paired sites at different altitudes (Smit and Singels, 2007); even then, differences in environmental variables and crop development makes comparison challenging. If temperature responses are coordinated across plant processes, however, then using germination rate response to temperature as a proxy for relative temperature sensitivity for other processes could be feasible and practical within an operational breeding programme. Setts or buds can be germinated in temperature-controlled germination rooms relatively easily and cheaply, at a fraction of the spatial scale and complexity of a field trial. By germinating buds of known reference Gs and unphenotyped Gs in two germination rooms set to different temperatures, absolute and/or relative temperature sensitivity for new Gs could be inferred (Poser et al., 2019).

For the predetermined DoGP simulations, E and G effects on ADMh were reasonably well-predicted (not statistically significant, noting that achieving a significant relationship with 3-4 data points is difficult) without changing the maximum radiation use efficiency parameter *RUEadm*. This tends to confirm the conservative nature of maximum RUE, consistent with the APSIM-Sugar model, in which this parameter value cannot be set per-cultivar; and disagrees with the findings of Jones et al. (2019) (Chapter 3) who found different RUE values for each G, and Jones et al. (2021) (Chapter 4) who asserted that models needed to permit G-specific maximum RUE parameter values to be useful for plant breeding applications, as well as Hoffman et al. (2018) who found G-specific RUE values for 14 cultivars in a pot trial. It is possible that maximum RUE can vary in sugarcane genotypes; Basnayake et al. (2015) found a relationship between stomatal conductance and yield in a population of genotypes with wide genetic range. However, the intense selection pressure in breeding programmes means that elite varieties are selected close to the genetic ceiling for RUE (Dias et al., 2019).

Hoffman et al. (2018) also reported significant G differences for DoTP. DSSAT-Canegro parameters based on these G-specific RUE and DoTP parameters were used in simulations of a field trial (plant and ratoons 1-3) at Pongola, South Africa. The model was able to predict yield G effects and rankings accurately across all Es (i.e. all four crops combined), but predictions were less accurate for individual Es (i.e. individual crops), suggesting that these G parameters/traits do not have a strong bearing on GxE interactions in cane yield. Sexton et al. (2017) found that APSIM-simulated biomass and sucrose yields, for two sites in Australia, were insensitive to thermal time to OSG, with more influential parameters including maximum RUE, maximum green leaf number and radiation extinction coefficient; germination/emergence-related parameters were not evaluated.

5.5.5 Model structure and concepts

Many assumptions were made regarding the role of onset of stalk growth (OSG) as the key mediator of several important plant processes. The assumptions made in the development of the model reflect interpretation across a wide body of evidence, however, and the calibration, validation and E/G/GxE evaluation results lend confidence to these assumptions. The model asserts that the OSG transition drives the increase in phyllocron interval, reduces canopy growth rates, stops tillering and triggers senescence; it is quite possible that OSG is actually the outcome of these

correlated phenomena. Even if this is the case, it does not matter much in terms of practical simulation outcomes, and the model could provide a framework for exploring such causality in future research work. That said, the duration of the tillering phase (which ends with OSG, DoTP) was predicted more accurately by DSSAT-Canegro, with its thermal time-based algorithm, than the CaneGEM model's FIPAR-based DoTP prediction, for the NCo376 calibration and validation sets. Future work could assess model performance where the DoTP algorithm is replaced with a thermal time approach, but where OSG status still interfaces with canopy development and biomass partitioning as described for the CaneGEM model. This would however lose the connection between radiation intensity and date of OSG identified in Chapter 3. While APSIM-Sugar also uses a thermal time OSG threshold, the Mosicas model relies on an ADM threshold to trigger OSG. Mosicas shares with the CaneGEM model the concept of a gradual transition to OSG, although in Mosicas this is made a function of ADM, while in this CaneGEM model it is a function of radiation interception. The findings of Chapter 3 were that the ADM threshold was less reliable than the thermal time method.

The focus areas for CaneGEM model development were, for simplicity, in improving seasonal canopy development and biomass accumulation. It was however necessary to account for stalk sugar mass for the source:sink-based approaches to biomass partitioning, and consequently structural growth, to work properly. Daily hexose requirements are estimated dynamically as a proportion of the daily mass of structural fibre synthesised, and sucrose mass determined as the remainder of stalk sugars. This approach to predicting sucrose mass appears to be accurate and is a strength of the CaneGEM model. An additional strength is that the partitioning to stalk sugars is entirely an emergent consequence of source and structural sink strengths, so sucrose accumulation is effectively controlled by trait parameters regulating structural stalk and leaf growth and sink strengths.

DSSAT-Canegro and APSIM-Sugar make use of a leaf-aging concept, modelled as a maximum number of green leaves per stalk. This has been shown to be unstable with respect to row-spacing (Singels and Smit, 2009). The literature review (Chapter 2) revealed a remarkable dearth of information about leaf senescence in sugarcane – and the experience with this model suggests that senescence has an important bearing on biomass partitioning dynamics in the plant. Canopy dynamics following OSG are poorly understood and this warrants further research; the CaneGEM model provides a suitable framework for further exploration in this regard.

The focus in this study was on fully-irrigated, well-fertilised sugarcane. It is on this basis that the absence of water stress effects in the CaneGEM model is justified. Additionally, a multi-G, multi-E water stress response trial dataset is not available, so it would not have been possible to evaluate such a model. Basnayake et al. (2012) found that genotype yield rankings were very similar under unstressed and mild- to moderately-stressed conditions. This suggests that a model capable of simulating only unstressed sugarcane growth is still useful for assisting breeding in industries where mild/moderate water stress may be encountered. Nevertheless, implementing the CaneGEM algorithms in an established crop modelling system (see Recommendations, Section 5.5.8 below) will permit linkages with existing water stress routines and will equip the model for exploring trait modelling more explicitly under stressed environments.

5.5.6 The CaneGEM model is well-suited for supporting plant breeding

Fulfilment of characteristics for breeding-ready models

Biomass partitioning is highly "emergent" (Hammer et al., 2019b; Hammer and Jordan, 2007), with component masses determined as consequences of the interaction between simple plant processes, regulated by G parameters, in response to temperature and radiation.

The CaneGEM model simulates complex traits, such as canopy development and accumulation of biomass, stalk and stalk sugar yields, accurately in an emergent manner.

The model is fundamentally simple and parsimonious, considerably more so than DSSAT-Canegro and APSIM-Sugar. CaneGEM has 48 input parameters, while DSSAT-Canegro (for example) has 88 input parameters across its cultivar, ecotype and species files for simulating unstressed growth. It is acknowledged that some of the complexity of DSSAT-Canegro and APSIM-Sugar relates to their simulation of water, energy and (in ASPIM anyway) nitrogen balances. The key model inputs driving accurate simulation of GxE effects in canopy cover, $RGRlai_{max}$ and *Tb_lai*, can be phenotyped relatively easily with standard instrumentation, and could in principle be inferred from estimates of *GLAI* and *FIPAR* from multi-spectral remotely-sensed imagery. In these ways, the model is suitable for application in sugarcane breeding.

Possible model applications to support breeding

The parameters $RGRlai_{min}$ and $RGRlai_{max}$ appear to be stable per G with respect to temperature and radiation, and permit the emergent accurate simulation of canopy development (a complex phenotype). This model could be used to explore the suggestion by Sinclair et al. (2004) that high initial leaf area development could be used as a marker for high yield in sugarcane breeding. Analysis of GxE interaction effects for R570 and CP88-1762 at Belle Glade R1 and La Mare R1 suggests that the combination of duration of the germination phase (unresolved GxE effects), G-specific cardinal temperatures and *RGRlai* parameter values explain much of the observed GxE variability in seasonal radiation interception and biomass yields at harvest; greater initial leaf area is consistent with any/all of these, and so it would appear that the model outcomes support this hypothesis.

The model could be used for phenotypic prediction across a wide range of Es, if it is calibrated appropriately for other cultivars. In principle, it could be used for genetic trait sensitivity analyses to understand the relative impact of changes in values of genetic trait parameters in different Es, in order to narrow down the domain of target genetic regions for increasing yield within particular TPEs. For example, in warm Es decreasing *Tb_lai* would have minimal impacts, while in a cool E this could dramatically increase canopy development and biomass accumulation rates.

The model lends itself to G trait explorations where the model self-imposes tradeoffs between parameter values. Increasing *RUEadm*, for example, would always enhance yields and would not generate any new insights. If, however, if it can be assumed that $RGRlai_{max}$ and $RGRlai_{min}$ are correlated, faster initial canopy development would imply a greater rate of leaf area production in a mature crop. Unless such genotypes were also to have a higher senescence threshold (parameter *GLAIst*), a greater leaf mass would senesce due to earlier and more profuse senescence. The light interception and biomass accumulation gains of faster canopy development therefore would – in this

situation – need to be balanced against greater unproductive losses of biomass in senesced leaf material.

Finally, it may be possible to identify QTL that are associated with *Tb_lai*, *To_lai*, $RGRlai_{max}$ and/or $RGRlai_{min}$. This would permit indicative predictions of canopy development rate and biomass yields from genetic sequence analysis, bearing in mind that highly significant predictions of GxE interaction effects in *FIPARa* and *ADMh* were predicted for three diverse Gs by changing only five trait parameters.

5.5.7 Additional possible model applications

The novel approach to simulating biomass partitioning may make the CaneGEM model (or future derivatives thereof) uniquely suitable for studies that go beyond the capabilities of existing models. For example, the ability to simulate sink strength limitations might provide further insights into understanding the reduced growth phenomenon (Park et al., 2005; Van Heerden et al., 2010). Additional applications could include investigating fibre content traits (e.g. for biomass energy applications), modelling chemical ripener effects, and assessing climate change impacts and adaptations.

5.5.8 Recommendations

Implementation of novel algorithms in established modelling systems

The CaneGEM model in its current form does not have a water balance and has not been tested for water-stressed conditions. Work is underway to make the model available in the DSSAT v4 system (Jones et al., 2003) and it is recommended that this work is completed, including the necessary linkages to water stress indices.

Implementation of this model in the Mosicas codebase would be relatively easy, as it already operates on an integrated canopy basis. The APSIM-Sugar and DSSAT-Canegro models could be adapted to change the basis for predicting date of OSG from thermal time to the described transition based on radiation interception. APSIM-Sugar already includes the *SLA* and senescence algorithms used in CaneGEM. The advantage of implementing these algorithms in existing modelling systems is reduced complexity for existing users of these models, and compatibility with existing simulation datasets. Additionally, modelling systems such as APSIM and DSSAT allow linkages with existing modules that would potentially expand the functionality of the model, such as those that simulate the soil water and nitrogen balances.

DSSAT-Canegro could be modified to limit leaf expansion to *SLA*-allowable ranges, and to link tillering rate with carbohydrate availability. This could address a potential weakness with the integrated canopy approach, whereby tillering and leaf expansion are reported to have different cardinal temperatures for leaf appearance, expansion and tillering (Bezuidenhout et al., 2003; Bonnett, 1998; Jones and Singels, 2018; Smit and Singels, 2007).

Germination modelling

Duration of germination phase (DoGP) appears to be a strong determinant of GxE interaction effects in canopy development and biomass yield for irrigated sugarcane, and it is recommended that research into improving the prediction of DoGP, particularly for ratoon crops, is undertaken.

Germination in ratoon crops may start sooner than the date of harvest if conditions are suitable – for example, leaf/tiller senescence following water stress, or lodging, might allow more light closer to the ground; damage, from frost or lodging, might break apical dominance. The development of a more accurate model of ratoon crop germination may require data recorded from the last few months of previous crops, as well as data for the new ratoon crops. Data should include: daily weather variables, soil temperature at different depths, date of start of drying-off, lodging dates and severity ratings, and frequent (every 1-2 weeks initially, and then at regular thermal time rather than calendar time intervals) observations of shoot number, leaf number, and green leaf and total radiation interception. Advances in the prediction of lodging events (van Heerden et al., 2015b) might prove valuable in predicting germination timing in following ratoon crops.

The germination of plant crops is limited by soil water content (Keating et al., 1999), and can be completely suppressed in a very dry soil (Smit, 2010). It is recommended that for field trials, soil water content is measured at crop start, and that careful records of early irrigation applications are kept. This will permit crop models such as DSSAT-Canegro and Mosicas to be started on the date of first significant irrigation application (or rainfall event), or alternatively (for models that consider soil water content effect on germination rate, e.g. APSIM-Sugar) to predict date of emergence as accurately as possible on the basis of soil water content.

5.6 Conclusion

A new sugarcane model, CaneGEM, has been developed, which (1) simulates canopy development based on temperature and carbon availability, and is therefore responsive to radiation intensity via photosynthesis; and (2) predicts the onset of stalk growth in response to canopy cover, linking this process to radiation intensity. The model was calibrated and validated on 22 (11 each) experimental crops of the reference cultivar NCo376. The CaneGEM model was calibrated for three cultivars against data from an international multi-environment trial, primarily by setting G-specific temperature sensitivity thresholds. Thereafter, the CaneGEM model's abilities to predict E, G and GxE interaction effects were assessed using the multi-environment trial dataset. The DSSAT-Canegro model was also validated and assessed for G-E effects predictions with the same datasets and methods to provide a performance benchmark for evaluating model improvement.

Validation outcomes for the CaneGEM model indicated generally similar performance to that of DSSAT-Canegro for the NCo376 experiments, but with simpler underlying concepts and a greater degree of inter-process coupling, and with a smaller number of genetic control parameters (48 vs 88). The CaneGEM model outperformed DSSAT-Canegro for predicting GxE interaction effects on seasonal radiation interception and biomass yields at harvest, where growth phase duration was predetermined. With duration of growth phase simulated, however, the CaneGEM model was not capable of predicting GxE interaction effects in radiation interception or biomass yields at harvest. Attempts to modify the germination algorithm to use simulated soil temperatures were inconclusive, so the CaneGEM model uses a standard model of germination that makes use of air temperatures.

All specific objectives were addressed, and the overarching objective of developing an improved sugarcane model, capable of accurately simulating GxE interaction effects

on seasonal radiation interception and (by implication) biomass yields, was partially achieved.

The CaneGEM model represents a balance between conceptual simplicity and biological realism. To a large extent, the model relies on emergent consequences to simulate complex phenotypes such as the transition from tillering to stalk growth phases, accumulation of biomass components and yields. The CaneGEM model fulfils the requirements of suitability as a tool for supporting crop improvement.

The novel approach to simulating biomass partitioning may also make it uniquely suitable for studies that go beyond the capabilities of existing models, such as exploring the reduced growth phenomenon, stalk fibre management, and modelling chemical ripener effects.

Two key recommendations are made for future work. Firstly, in order to become more useful operationally, the CaneGEM model algorithms ought to be implemented in existing crop modelling software to take advantage of existing expertise and databases. The second recommendation is that further research effort is invested in understanding the dynamics of germination, particularly for ratoon crops, with a focus on conditions prior to harvest of the preceding crop. This work revealed the overwhelming importance of germination phase duration in explaining GxE interaction effects on radiation interception and biomass yields; existing germination algorithms are inadequate for predicting date of emergence with sufficient accuracy, particularly for ratoon crops, where in many cases germination appeared to have started before harvest of the previous crop.

6. MODEL APPLICATION CASE STUDY: ASSESSING THE IMPACTS OF CANOPY DEVELOPMENT TRAITS FOR FOUR ENVIRONMENTS

TABLE OF CONTENTS

6.1 Introduction 111
6.2 Methodology 114
6.2.1 Simulations 114
6.2.2 Crop management 117
6.2.3 Data analysis 117
6.2.4 Weather data 118
6.3 Results 119
6.3.1 Overview 119
6.3.2 Belle Glade, Florida, USA 119
6.3.3 Chiredzi, Zimbabwe 123
6.3.4 La Mare, Reunion Island 126
6.3.5 Pongola, South Africa 129
6.4 Discussion 133
6.4.1 Temperature impacts 133
6.4.2 Faster canopy development and higher senescence threshold 135
6.4.3 Significance and provisos for sugarcane breeding 135
6.4.4 Recommendations for model improvement 136
6.4.5 Additional model applications 137
6.5 Conclusion 137

6.1 Introduction

Rising demand for food and biofuels demands rapid acceleration of genetic gain (Diepenbrock et al., 2021). Climate change will affect future crop production, in many cases resulting in decreases in yield. Recent IPCC findings (Engelbrecht and Monteiro, 2021) report that future rainfall is expected to decrease in sub-Saharan Africa, in contrast to previous projections which suggested that future rainfall may be slightly higher than the reference historical periods. The combination of decreased rainfall and increased evaporation might mean that future sugarcane yields will be negatively impacted by climate change in South Africa, Zimbabwe, Mozambique, Reunion island, and Mauritius, rather than the positive impacts generally reported (Jones et al., 2015; Jones and Singels, 2014; Knox et al., 2010; Marin et al., 2013; Singels et al., 2014). While irrigated sugarcane yields may have the potential to

increase under climate change, sucrose content and yield may decrease (Jones et al., 2015); and meeting increased evaporative demand implies greater electricity costs for irrigation. Increased evaporative demand, decreased rainfall and increased demand for water from other crops and economic sectors in future also represent threats to irrigation sugarcane production. Climate is also expected to become more variable, with increased incidence of extremes, and this also represents a threat to sugarcane production (Christina et al., 2021). There is a need to develop capacity to accelerate the production of high-yielding sugarcane cultivars adapted to (changing) growing environments (Es) in order to meet future demands for sugar, bioethanol and fibre.

Sugarcane worldwide is bred using a traditional approach of crossing and selection, which is resource intensive and takes a long time (typically 10-14 years per cultivar). Resource constraints limit breeders to selecting broadly-adapted varieties that perform well across large target populations of environments (TPEs). In doing so, opportunities for exploiting genetic adaptations to more specific Es are lost.

Crop improvement for many species is routinely undertaken with the assistance of advanced modelling techniques (Messina et al., 2018). Whole-genome prediction (WGP) methods use carefully articulated field trial data in multi-environment trials (METs) to train statistical models to predict complex phenotypes (e.g. yield) from genetic sequence data for the TPE. WGP models combined with crop growth models (CGMs) leverage the crop physiology knowledge and dynamic nature of these CGMs to account for genotype-by-environment (GxE) interaction effects and non-linear genetic effects. This can make breeding more efficient, for example by reducing the size of the training datasets for WGP models and to support exploration of niche breeding for sub-environments within the TPE.

CGMs can support plant breeding in simpler ways, such as by characterising environments – e.g. assigning Es to water stress progression categories (Chenu et al., 2009a; Hammer et al., 2019a; Ramburan, 2012). Additionally, given that yield is a complex trait that is the outcome of lower-level plant process interacting with the growing environment throughout the duration of the crop, CGMs allow the roles of these sub-traits in yield formation to be explored, for the complex traits to be 'dissected' (Hammer and Jordan, 2007). The identification of optimal combinations of model G trait input values for specific TPEs amounts to developing 'ideotypes' (Donald, 1968; Hammer and Jordan, 2007; Tao et al., 2017), which can be treated as targets for plant breeding. The application of CGMs to support sugarcane breeding has been very limited, partly due to the genetic complexity of sugarcane (Balsalobre et al., 2017) and partly due to uncertainty regarding the accuracy and applicability of existing general-purpose sugarcane CGMs to support this purpose.

Several widely-used sugarcane CGMs are available. These include the DSSAT-Canegro, APSIM-Sugar and Mosicas models. The use of these models to explore G effects has been limited. Sexton et al. (2017) evaluated sensitivity of trait parameter values for the APSIM-Sugar model for two Es in Australia. Inman-Bamber et al. (2016, 2012) explored water stress adaptation traits in sugarcane, also using APSIM-Sugar. Singels et al (2016) used the DSSAT-Canegro model to explore water uptake traits as a climate change adaptation, as well as rooting characteristics for different stress environments. Hoffman (2018) used a pot trial phenotyping study to estimate trait parameter values for the DSSAT-Canegro model, which was then able correctly to predict cultivar rankings in a field trial.

Owing to the absence of a suitable multi-G, multi-E growth analysis dataset for sugarcane, it was not until the present study (Jones et al., 2021) that these CGMs were formally assessed for their abilities to predict GxE interaction effects, a step recommended by Boote et al. (2021). This study was limited to four fully-irrigated environments: Belle Glade, Florida, USA; Chiredzi, Zimbabwe; La Mare, Reunion island, France; and Pongola, South Africa (Jones et al., 2019). A new model, called CaneGEM, described Chapter 5, was developed to address model shortcomings that limited sensitivity to key sources of GxE interaction effects. These included carbon-linked canopy development, phenology linked to canopy development, source:sink-based biomass partitioning, and G-specific threshold temperatures for plant processes. The CaneGEM model provided some improvement in the prediction of GxE interactions, but also better fulfils the requirements of credibility and appropriateness recommended by Hammer et al. (Hammer et al., 2019b; Hammer and Jordan, 2007) for CGMs intended for supporting breeding. These include emulating growth in biologically-realistic manner, and predicting complex traits as emergent consequences of lower-level processes.

The CaneGEM model uses cardinal temperatures to control the process rate responses to temperature for duration of the germination phase, canopy expansion rate, stalk expansion rate and photosynthesis rate. Process rates increase linearly as air temperature increases from the base temperature to the optimal temperature, and then decrease linearly between the optimal temperature and the ceiling temperature. Lowering the base temperature threshold (*Tb_x* for process *x*) could be expected to increase process rates at low temperatures, but might make the process less efficient at higher temperatures if they exceed a similarly-lowered optimal temperature (*To_x*) or near the ceiling temperature (*Tf_x*).

CaneGEM uses relative growth rate parameters (*RGRlaimin* and *RGRlaimax* parameters; Eqn.(5-13) in Section 5.2.5) to control the maximum extent by which the canopy can expand as a fraction of its current size each day (where the actual extent is limited by temperature and source availability). Two parameters are defined: a high value (*RGRlaimax*) for when the crop is very young, and a smaller rate (*RGRlaimin*) defined for the when the crop has fully transitioned to the stalk growth phase. Increasing these parameter values can be expected to result in faster canopy development, as long as biomass accumulation rates are sufficient to meet the greater sink strengths from a faster-expanding canopy, and so drive greater seasonal radiation interception and consequently greater final biomass yields.

CaneGEM also considers a canopy senescence threshold (*GLAIst* parameter (Eqn. (5-15) in Section 5.2.5): this is the leaf area index value at which leaf senescence starts. In principle, increasing this value should result in a larger canopy being maintained with smaller loss of biomass to leaf senescence, which could be expected to increase yields.

The value of this model to support plant breeding needed to be demonstrated. The study was funded by four cooperating institutions (SASRI, SGC Florida, ZSAES, CIRAD), members of the International Consortium for Sugarcane Modelling (ICSM), with the trial dataset collected by them. For this reason, it was considered important to demonstrate value from the new model at each of the sites where data were collected (Belle Glade, Chiredzi, La Mare and Pongola).

Furthermore, it was recognised that for well-managed fully-irrigated sites, temperature and radiation responses are the key environmental determinants of yield. Temperature

responses in the CaneGEM model are determined by threshold temperatures – base, optimal and ceiling values, for plant processes for which there is evidence of genotype-specificity: controlling germination, canopy development and stalk elongation (Bonnett, 2014; Campbell et al., 1998; Inman-Bamber, 1994; Ngobese et al., 2018; Singels et al., 2005c; Smit, 2010). There is also evidence of G-specific tillering and leaf development traits in sugarcane (Bonnett, 1998; Sinclair et al., 2004; Singels et al., 2005c; Zhou et al., 2003), which regulate radiation interception in the crop. Conversely, there is little evidence for G-specific cardinal temperatures controlling leaf appearance (Bonnett, 1998) or photosynthesis. As crops harvested early-, mid- and late-season experience differing climatic conditions at different developmental stages, it is necessary to account for this in simulation studies (e.g. Bezuidenhout and Singels (2007); Jones and Singels (2015)).

With these considerations in mind, the broad objective was to evaluate the ability of the CaneGEM model to inform crop improvement by exploring the impact of genetically-controlled canopy development traits and temperature sensitivities on crop performance in different environments.

The specific objectives were to use the CaneGEM model to assess and explain the impacts of changes to trait parameter values for:

- cardinal temperatures controlling rates of germination, canopy development and stalk elongation;
- relative canopy growth rate; and
- canopy senescence threshold,

on biomass, stalk and sucrose yields, for early-, mid-, and late-season harvested crops, at each of the four ICSM sites.

6.2 Methodology

6.2.1 Simulations

Simulations were conducted using the new source-sink sugarcane model described in the previous chapter. At each site, the model was set up to simulate crops to be harvested at the typical start, end and middle of the milling season, for the last 20-30 years, depending on the availability of weather data. For each of these simulated cropping seasons, 13 genotypes were assessed, consisting of:

- **A baseline G (NCo376)**. The CaneGEM model was calibrated for NCo376, N41, R570 and CP88-1762; but only validated using NCo376 data. For this reason, NCo376 was chosen as the 'baseline' G on which to test perturbations to G parameter values, at all sites.
- **Four temperature-adapted Gs** with base, optimal and ceiling cardinal temperature parameters for germination, canopy expansion, and stalk elongation (discussed in Section 5.2.4; trait parameters *Tb_germ*, *To_germ*, *Tf_germ*; *Tb_lai*, *To_lai*, *Tf_lai*; and *Tb_sk*, *To_sk*, *Tf_sk*; see Table 5-1) altered from the baseline cultivar by -2 °C, -1 °C, +1 °C, +2 °C respectively.
  In the absence of clear information showing otherwise, and recognising that temperature responses have been demonstrated to be coordinated across plant processes for sugarcane and other crops (Parent and Tardieu, 2012), it was assumed that temperature thresholds across all affected plant processes would increase or decrease by the same extent – i.e. the same genetic region controls

temperature responses for germination, leaf expansion, tillering and stalk elongation.
The range of temperature changes imposed was chosen on the basis of (1) using a set of values sufficiently broad to demonstrate clearly the direction and nature of responses to G variation in temperature thresholds; (2) this roughly representing the perceived variation in temperature across the set of sites considered, and so would reflect temperature response variation for Gs adapted to these regions, and (3) climate change projections indicating warming of 1-2 °C over the next 30 years compared to the 1980s, providing additional insights into consequences of genetic adaptation (or non-adaptation) to climate change likely to be experienced at each of the sites. The temperature increments also needed to be large enough for the CaneGEM model to distinguish clearly between treatments, and the experience with operating the CaneGEM model indicated that 1 °C increments would be appropriate;

- **Four Gs with adapted relative canopy growth rate**, represented with *RGRlaimin* and *RGRlaimax* parameters (Eqn. (5-13) in Section 5.2.5; see Table 5-1 for details) altered from the baseline G by -10%, -5%, +5%, +10% respectively.
  The magnitudes of changes imposed was based on the range of leaf development rates reported by Bonnett (1998), which varied by ≈ 25% between the fastest and slowest genotypes assessed.
- **Four Gs with adapted senescence thresholds**, represented with the *GLAIst* parameter (Eqn. (5-15) in Section 5.2.5; see Table 5-1 for details) altered from the baseline G by -10%, -5%, +5%, +10% respectively.
  The magnitudes of the changes imposed were chosen to reflect the changes in relative canopy development rate.

The changed trait parameter values are listed in Table 6-1. Unchanged trait parameters are listed in Table 5-1 (Chapter 5).

Table 6-1. Trait parameter values used for adapted genotypes in the case study.

	Cardinal temperatures				Relative canopy growth rate				Senescence threshold			
Model param	**-2 °C**	**-1 °C**	**+1 °C**	**+2 °C**	**-10%**	**-5%**	**+5%**	**+10%**	**-10%**	**-5%**	**+5%**	**+10%**
Tb_germ	14	15	17	18	16	16	16	16	16	16	16	16
To_germ	26	27	29	30	28	28	28	28	28	28	28	28
Tf_germ	39	40	42	43	41	41	41	41	41	41	41	41
Tb_lai	11	12	14	15	13	13	13	13	13	13	13	13
To_lai	33	34	36	37	35	35	35	35	35	35	35	35
Tf_lai	43	44	46	47	45	45	45	45	45	45	45	45
RGRlaimax	0.01	0.01	0.01	0.01	0.009	0.0095	0.0105	0.011	0.01	0.01	0.01	0.01
RGRlaimin	0.25	0.25	0.25	0.25	0.225	0.2375	0.2625	0.275	0.25	0.25	0.25	0.25
Tb_sk	13	14	16	17	15	15	15	15	15	15	15	15
To_sk	33	34	36	37	35	35	35	35	35	35	35	35
Tf_sk	46	47	49	50	48	48	48	48	48	48	48	48
GLAIst	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.7	2.430	2.565	2.835	2.970

6.2.2 Crop management

All crops were set up as 12-month ratoon crops, harvested on the 15th of first, last and middle months of the typical milling season at each site. Full irrigation and adequate nutrition were assumed for all sites. Harvest months and year ranges are listed in Table 6-2. In each case, crops are assumed to have started in the harvest month of the preceding year.

Table 6-2. Harvest months and year ranges simulated at each site.

Site	Early season harvest month	Mid-season harvest month	Late-season harvest month	Start year	End year	Reference
Belle Glade, Florida, USA	October	January	April	2003	2020	Baucum and Rice (2009)
Chiredzi, Zimbabwe	March	July	November	2000	2021	M. Zhou, pers. comm.
La Mare, Reunion Island	June	September	December	1994	2020	Lejars and Siegmund (2004)
Pongola, South Africa	March	August	December	1998	2020	

6.2.3 Data analysis

The following variables were extracted or calculated from model outputs:

- **Phenology**:
  - Duration from crop start to 50% primary shoot emergence (DoGP, days since crop start);
  - date of 50% onset of stalk growth (DoOSG, days since crop start)
- **Physiology**:
  - Daily photosynthetically-active radiation (PAR) interception fraction (FIPAR)
  - Daily yield values of above-ground dry biomass, stalk dry mass and sucrose mass
  - seasonal radiation interception fraction (FIPARa), calculated as the ratio of the sum (over the whole cropping season) of daily intercepted PAR divided by the sum of daily incident PAR;
  - average apparent radiation use efficiency (RUEa, g/MJ), calculated as above-ground dry biomass at harvest divided by total intercepted radiation and expressed in g/MJ.
- **Biomass components**:
  - above-ground dry biomass at harvest (ADMh, t/ha);
  - stalk mass at harvest (SDMh, t/ha);
  - sucrose mass at harvest (SUCh, t/ha)

The mean, minimum, maximum and coefficient of variance of these variables were calculated across all cropping seasons, for all harvest times (early/mid/late) together and separately. In some cases, the mean and standard deviation of daily values (per day after crop start) across simulation years, per harvest time, were calculated.

Results for the adapted Gs were expressed relative to that for the baseline G in the form of a percentage change ("delta"), calculated as:

$$delta = \left(\frac{value_{Gx}}{value_{BL}} - 1\right) * 100 \quad (6\text{-}1)$$

where $value_{Gx}$ is the value for adapted genotype Gx, and $value_{BL}$ is the value for the baseline cultivar. All analyses were conducted using R (R Core Team, 2016).

6.2.4 Weather data

A summary of weather data for the four sites is provided in Figure 6-1.

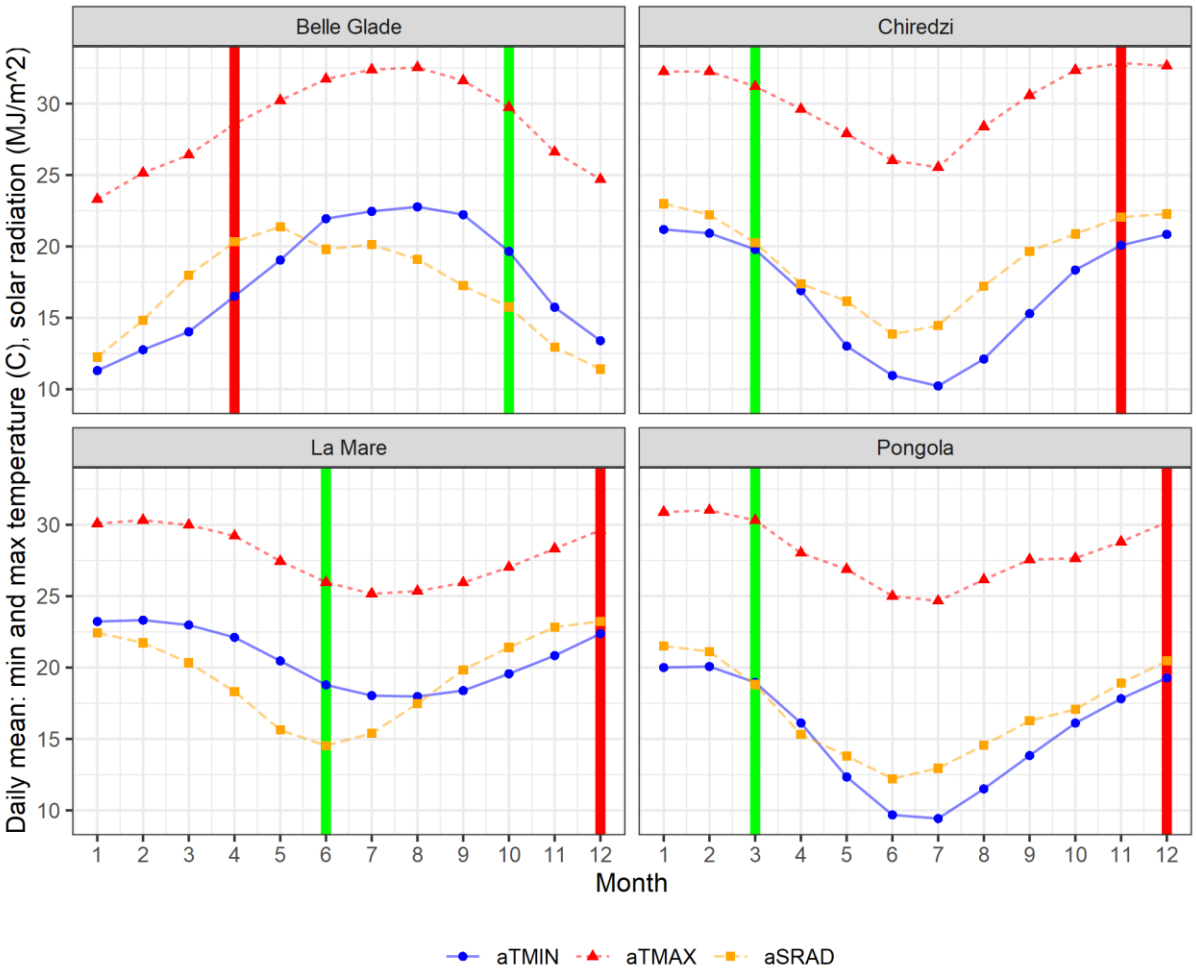

Figure 6-1. Summary of monthly long-term weather data for the four study sites. 'aTMIN' and 'aTMAX' are average daily minimum and maximum air temperature (°C) repectively; 'aSRAD' is average daily global solar radiation (MJ/m²/d). The vertical green and red bars indicate the start and end months respectively of typical milling seasons at each site.

Belle Glade, Florida, USA

Weather data (TMIN, TMAX and SRAD) for Belle Glade were downloaded from the University of Florida web site[3]. Data were available from 2002-11-18 onwards. Missing or suspected incorrect data (TMIN < -10, SRAD > 40 MJ/m²/d, SRAD < 0.2 MJ/m²/d)

[3] https://fawn.ifas.ufl.edu/data/fawnpub/daily_summaries/BY_YEAR/

were replaced with NASA MERRA (Ruane et al., 2015) data downloaded from the NASA Power Portal[4] for location 26.6871 °N -80.6567 °E.

Chiredzi, Zimbabwe

Daily weather data (TMIN, TMAX, SRAD and sunshine hours (SHRS, h/d)) were provided by the Zimbabwe Sugar Association Experiment Station. The SRAD dataset had considerable gaps. The Angstrom-Prescott method in the R 'sirad' package (Bojanowski, 2016) was used to estimate SRAD from SHRS, based on a calibration from the periods where both SRAD and SHRS had been observed. Additionally, NASA Power MERRA data were downloaded for Chiredzi (-21.009°N, 31.6769°E) from the NASA Power Portal. The SRAD data were assessed visually for accuracy (Figure 11-10, Appendix B). Observed SRAD data were used where available for 2000-2010, and MERRA data used otherwise. Missing or suspected incorrect temperature values were replaced with MERRA values, bias-adjusted using linear regressions fitted to observed vs MERRA TMAX and TMIN separately (Figure 11-11, Appendix B).

La Mare, Reunion Island

Daily weather data for La Mare (TMIN, TMAX, and SRAD) were provided by CIRAD; these contained no gaps or obvious outliers for the period 1994-2020.

Pongola, South Africa

Daily weather data for Pongola, South Africa (Figure 11-13, Appendix B) were downloaded from the SASRI WeatherWeb[5]. No gaps or outliers for the period 1998-2000 were noted.

6.3 Results

6.3.1 Overview

Detailed results for each site are shown the following sections. Results are described and explained in the following order for each site:

1. Baseline G, across all times of harvest
2. Baseline G, per time of harvest
3. Adapted Gs, across all times of harvest
4. Adapted Gs, per time of harvest

Given the need to summarise the dynamics of multi-year simulations, in many cases the data are visualised with boxplots. These can be interpreted as follows: the thick horizontal line is the median value; the upper and lower bounds ("hinges") of the box indicate the 25th and 75th percentiles; the lines indicate 1.5 x the interquartile range above and below the hinges, and any data points outside of this range are plotted as points (Wickham, 2016).

A full summary of results for the four sites is shown in Table 11-2 in Appendix B.

6.3.2 Belle Glade, Florida, USA

At Belle Glade, germination for the baseline cultivar required between 32 (late-season) and 70 (mid-season) days. Onset of stalk growth occurred on average 127 days after crop start, ranging from 87 d for late-season crops to 160 d for late-season crops.

[4] https://power.larc.nasa.gov/data-access-viewer/
[5] https://sasri.sasa.org.za/pls/sasri/f?p=129:LOGIN_DESKTOP:9612752017621:::::

Average biomass (ADMh), stalk dry mass (SDMh) and sucrose mass (SUCh) yields were 53, 38 and 12 t/ha respectively (Table 11-2).

Late season crops intercepted the most radiation (FIPARa = 0.67, Table 11-2) and consequently accumulated the most biomass (Figure 6-2). Late-season crops experienced canopy cover losses due to frost nearer to harvest in some seasons, driving inter-seasonal FIPAR variation in the last 3 months of the seasons, while early season crops encountered frost close to crop start, driving FIPAR variation in the partial canopy period (Figure 6-2, Figure 11-14 in Appendix B). Mid-season crops mostly avoided frost damage, with inter-seasonal variation in FIPAR driven by temperature and SRAD fluctuations and not frost per se. Mid-season harvested crops had the lowest biomass and stalk yields, but ranked in the middle for sucrose yields. Early- and late-season crops had similar biomass yields but late-season crops had higher stalk yields. Late-season crops produced substantially higher sucrose yields than early- or mid-season crops (15.1 vs 10.3 and 11.6 t/ha), due to rapid canopy development, high radiation interception and biomass growth, and cool conditions in the 4-5 months before harvest favouring sucrose accumulation in addition to structural stalk growth.

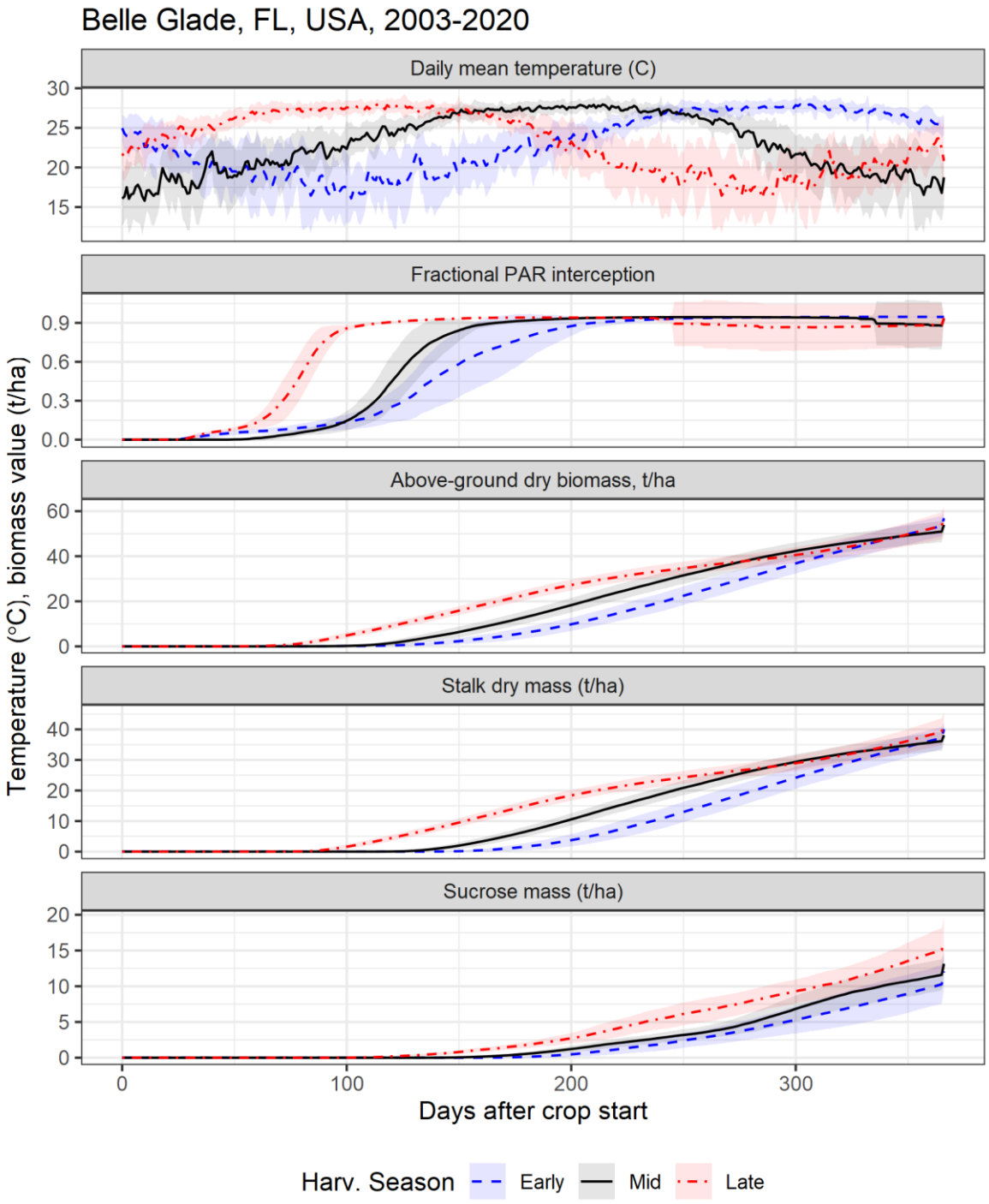

Figure 6-2. Daily average (±1 standard deviation, shaded) mean daily air temperature (C°), photosynthetically-active radiation interception fraction, above-ground dry biomass (t/ha), stalk dry mass (t/ha), and sucrose mass (t/ha) and for Belle Glade, 2002-2020, baseline cultivar (NCo376).

Adaptations to temperature sensitivity (Figure 6-3) had the greatest impact on yield-related outcomes. A clear trade-off between ADMh and SUCh was evident: lower temperature thresholds favoured larger biomass accumulation, but penalized sucrose yields, and *vice versa*. Increasing the temperature thresholds by 1 °C resulted in a 10% increase in SUCh and a 5% decrease in ADMh compared to the default cultivar, over all seasons and times of harvest. The +1 °C adaptation also resulted in the most

consistent changes (across time of harvest) in SUCh compared to the baseline cultivar overall. The +2 °C adaptation resulted in sucrose yields 18% higher than the baseline, but with a nearly 10% reduction in biomass yields (Table 11-2).

Simulated outputs were much less sensitive to changes in relative canopy development rate and canopy senescence threshold. The +10% canopy development rate adaption led to 1.6% increased seasonal radiation interception (FIPARa), but only a 0.2 % increase in biomass accumulation. Slowing canopy growth by 10% did however reduce sucrose yields by 0.8 %. Responses of a similar magnitude were noted for changing the senescence threshold.

ADM and SDM accumulation were least sensitive to temperature adaptations for late-season crops, and most sensitive for mid-season crops. This is because the magnitude of the temperature change was smallest relative to ambient temperatures at crop start for late-season crops, so the impact on radiation interception and biomass accumulation was relatively low. For sucrose accumulation, however, late-season crops showed the greatest sensitivity – increasing more than early- and late-season crops for Gs adapted to higher temperatures and decreasing more than the other times of harvest for adaptations to lower temperatures (Figure 6-3). This dynamic with sucrose accumulation is due to the magnitude of the changes being largest relative to ambient temperatures for the last 100 days of each season.

Sucrose yield deltas were similar (7-11%) for the +1 °C and +2 °C adaptations for early-season crops; for +2 °C Gs grown under early-season cycles, the biomass partitioning (sucrose) advantages were in some years offset by compromised ADM accumulation under the relatively cool season start conditions for these early season crops. This led to much greater variation in the SUCh delta, compared to the +1 °C adaptation. The +2 °C adaptation favoured sucrose yields for mid- and late-season harvesting, with increases in sucrose yield of 13 and 26% respectively. The -2 °C adaptation resulted in larger SUCh decreases: -23 and -29% respectively for mid- and late-season harvests.

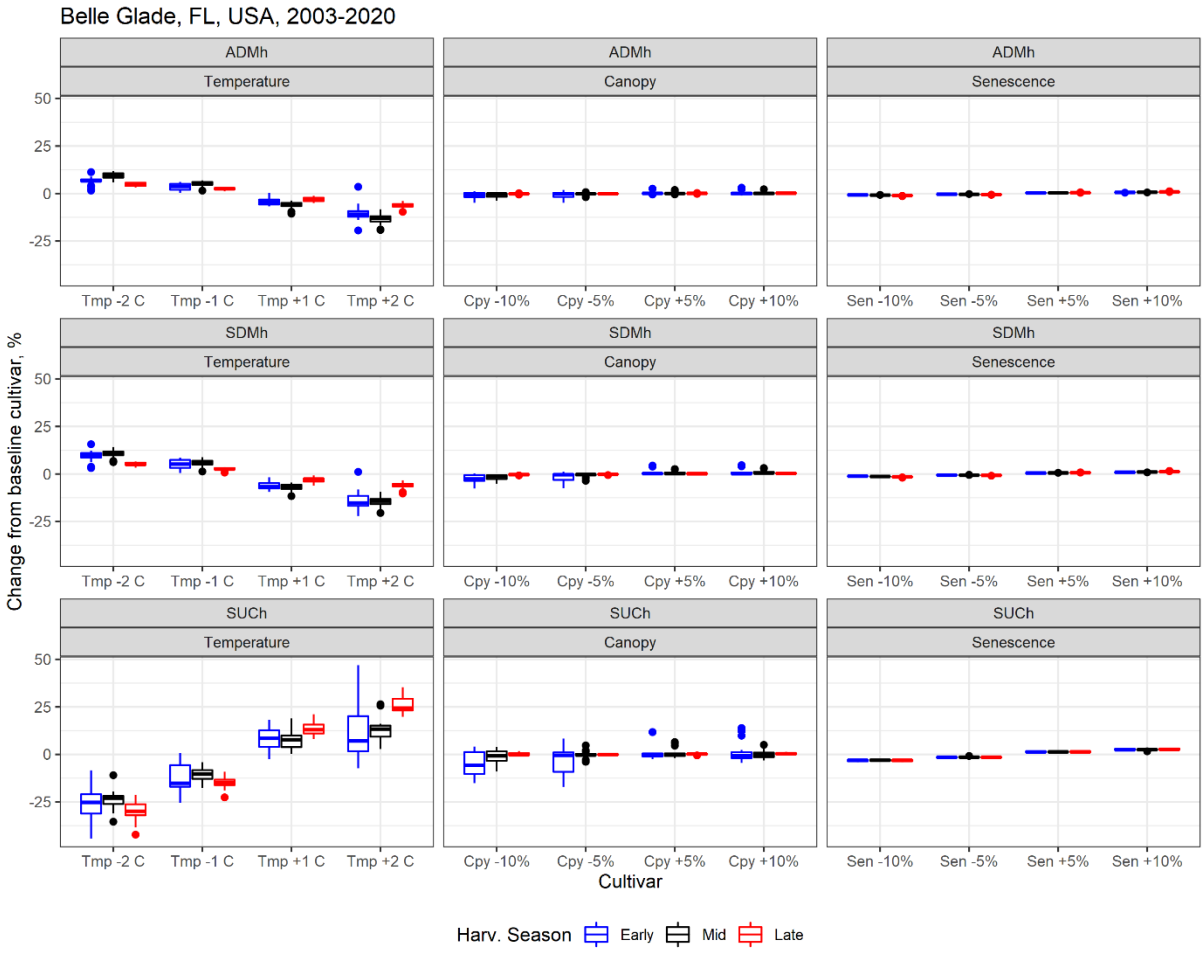

Figure 6-3. Percentage change in above-ground dry mass (ADMh), stalk dry mass (SDMh) and sucrose mass (SUCh) at harvest due to trait differences in temperature sensitivity (Tmp), relative canopy expansion rate (Cpy) and senescence thresholds (Sen) for Belle Glade, 2003-2020.

6.3.3 Chiredzi, Zimbabwe

At Chiredzi, early- and late-season crops emerged after ≈ 26 days (Table 11-2), while mid-season crops took more than twice as long to emerge (59 d) due to cool starting conditions. Onset of stalk growth (OSG) took longest for early-season crops (131 d after crop start), as a result of slow canopy development due to low temperatures for the first few months of growth, while mid- and late-season crops required 117 and 77 days. The time from emergence to OSG (i.e. DoTP) was however similar for mid- and late-season crops, ≈ 55 days. Average simulated ADMh, SDMh and SUCh yields were approximately 67, 48 and 21 t/ha respectively over the period 2000-2021.

ADMh for early and late-season harvests were almost identical (68.3 vs 68.1 t/ha), while mid-season harvests produced slightly lower biomass yields (63.2 t/ha). Early-, mid- and late-season crops were distinguished more clearly with FIPARa (0.71, 0.67 and 0.73 respectively); RUEa was similar (1.95 g/MJ) for early- and late-season crops, and slightly lower for mid-season crops (1.80 g/MJ). Late-season crops produced the highest stalk dry mass yields (49.6 t/ha), followed closely by early-season crops (48.4 t/ha); mid-season crops averaged lower SDMh yields (45.7 t/ha). Sucrose yields were identical for early- and mid-season crops (20.5 t/ha) and greater for late-season crops (23.3 t/ha).

Increasing the cardinal temperature thresholds resulted in decreases in ADMh and *vice versa*. Impacts were smaller at Chiredzi (-8 to +4% change overall) than at Belle Glade and Pongola (≈ -11 to +6% change), due to cooler periods at these sites, and more similar to La Mare (-7 to 4% change). Stalk dry mass followed suit, decreasing with increase temperature thresholds. Sucrose mass showed an inverse response, as at other sites, decreasing by 18.7% in the -2 °C adaptation and increasing by 9.8% overall with the +2 °C adaptation. Increasing the relative canopy growth rate did not increase ADMh yields, but decreasing the rate reduced biomass yields by 0.3%. ADMh yields were slightly more sensitive to the senescence threshold (-0.9 to +0.8%).

Late-season crops showed the smallest response to temperature adaptation in terms of ADMh (-3.2 to 1.6% change) and SDMh (-3.8 to 2.6%), because high temperatures at crop start meant that 1-2 °C changes had little impact on canopy development rate, ensuring that FIPAR and ADM accumulation were not much affected. Late-season crops however showed the greatest response in terms of sucrose yields (-23 to +21% change in for -2 °C to +2 °C): for the +2 °C adaptation, warm conditions early on in the crop ensured a rapidly-developing canopy, and moderate-cool conditions from mid-way through the crop cycle drove a favourable combination of stalk structural growth and sucrose accumulation.

Canopy growth rate adaptations had little impact. Early-season sucrose yields became more variable compared to the baseline cultivar, but benefited slightly overall from faster canopy development. Late-season crops on the other hand benefited (very slightly) from slower canopy development.

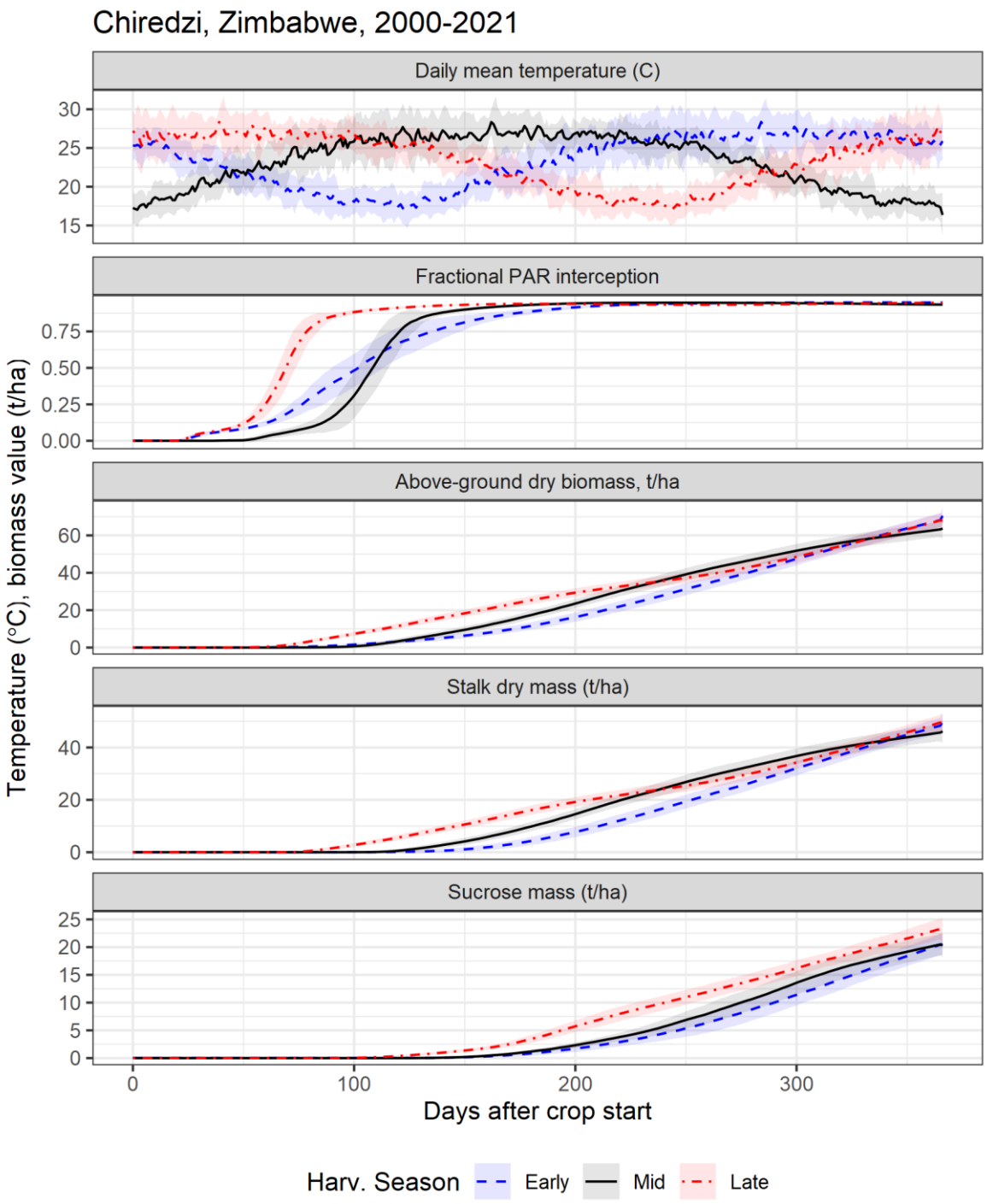

Figure 6-4. Daily average (±1 standard deviation, shaded) mean daily air temperature (C°), photosynthetically-active radiation interception fraction, above-ground dry biomass (t/ha), stalk dry mass (t/ha), and sucrose mass (t/ha) and for Chiredzi, 2000-2021, baseline cultivar (NCo376).

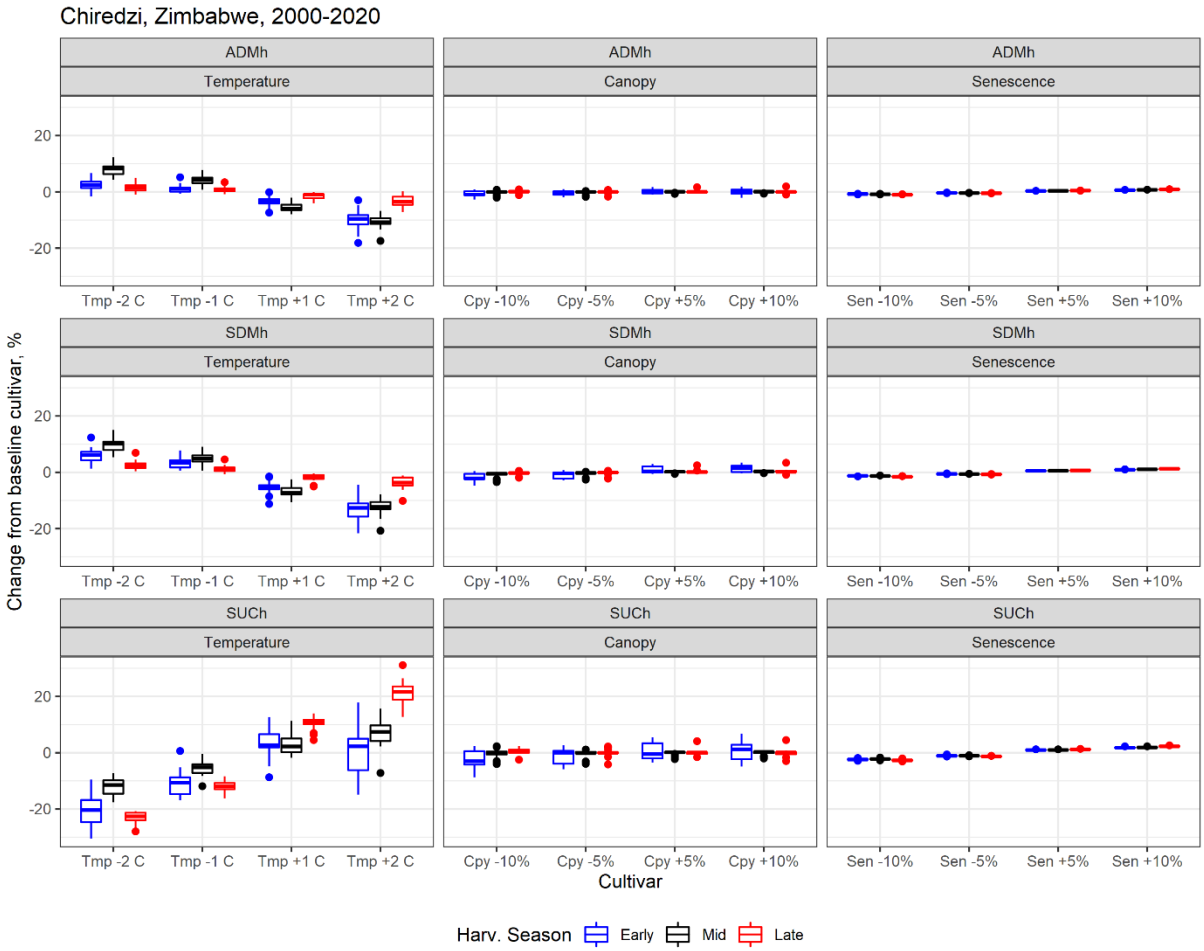

Figure 6-5. Boxplots of biomass variables for Chiredzi, 2000-2020, expressed as percentage changes from baseline cultivar, for early- (blue), mid- (black) and late-season (red) harvested crops. ADMh = 'Above-ground dry biomass at harvest' (t/ha), SDMh = 'Stalk dry mass at harvest' (t/ha), SUCh = 'Sucrose dry mass at harvest' (t/ha). Genetic adaptations to temperature sensitivity, canopy growth rate and senescence threshold are illustrated.

6.3.4 La Mare, Reunion Island

At La Mare, germination for the baseline cultivar required between 23 (late-season) and 40 (mid-season) days. Onset of stalk growth was occurred on average 93 days after crop start, ranging from 75 d for late-season crops to 114 d for early-season crops. Average simulated ADMh, SDMh and SUCh yields were 72.3, 52.5 and 22.9 t/ha respectively over the period 1994-2020 (Table 11-2).

Late season crops intercepted the most radiation (FIPARa = 0.74, Table 11-2) and consequently accumulated the most biomass (Figure 6-2), although the differences between harvest times were smaller than at other sites (Figure 6-6), due to the narrow temperature range at La Mare (20-28 °C, compared with 15-30 °C at Belle Glade and Chiredzi). Biomass accumulation closely followed FIPARa, being greatest (75.2 t/ha) for late-season starts and smallest (70.5 t/ha) for mid-season crops. Stalk dry mass yields were smallest for early-season crops, however, due to delayed onset of stalk growth compared to the other harvest cycles (DoOSG = 114 for early-season crops vs 89 and 75 for mid- and late-season crops).

Late-season crops had considerably higher sucrose yields (26 t/ha) than early- (20.7 t/ha) and mid-season (21.9 t/ha) crops. Late season crops developed canopy cover faster and grew stalk structure to store sucrose in the warm start to the season, and then accumulated sucrose (along with stalk dry mass) from 200 days after crop start onwards; warming in the last month of the season was not sufficient greatly to slow sucrose accumulation rate up to harvest (Figure 6-6).

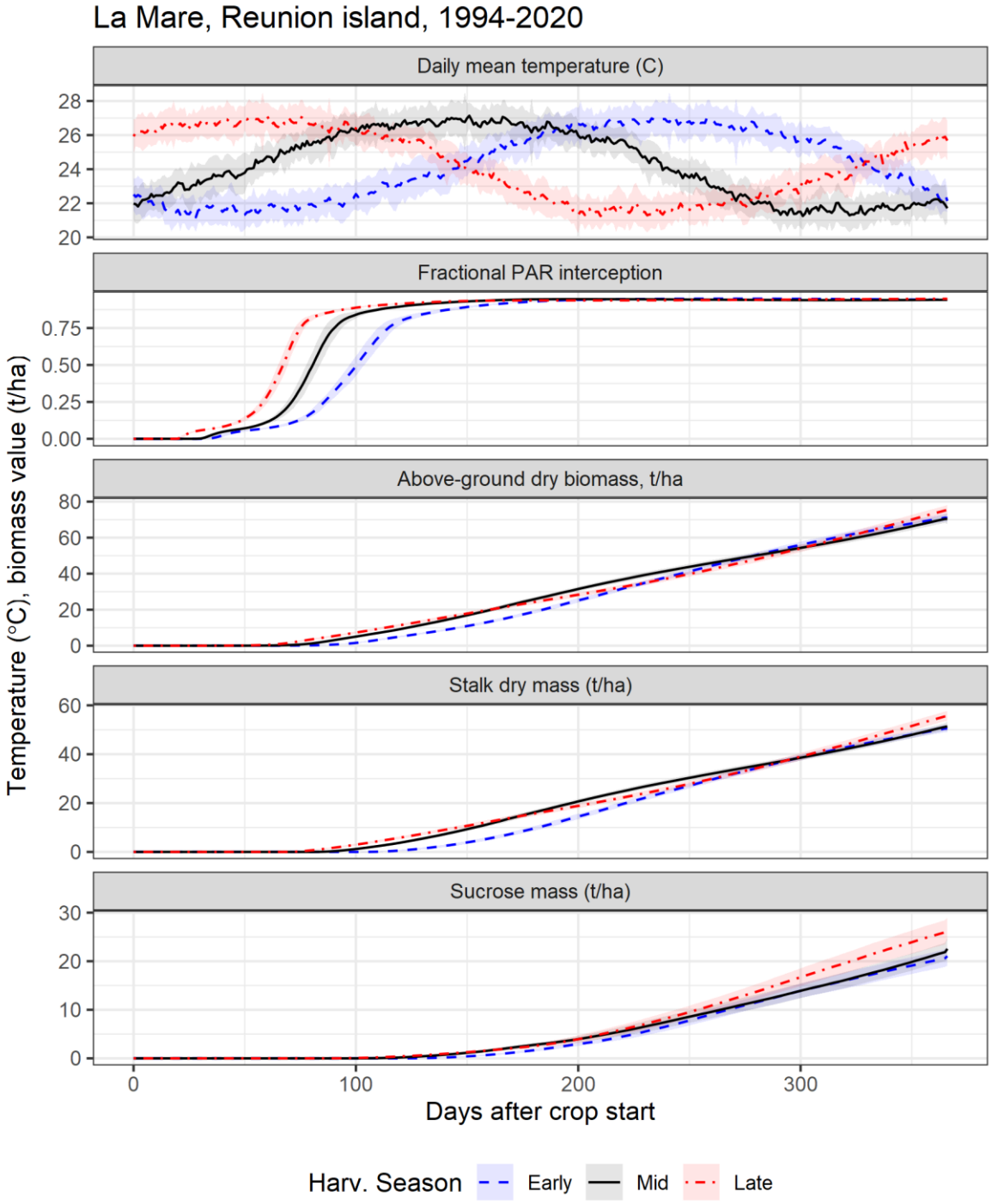

Figure 6-6. Daily average (±1 standard deviation, shaded) mean daily air temperature (C°), photosynthetically-active radiation interception fraction, above-ground dry biomass (t/ha), stalk dry mass (t/ha), and sucrose mass (t/ha) and for La Mare, 1994-2020, baseline cultivar (NCo376).

Gs adapted to higher temperatures produced lower biomass and stalk dry mass yields, but increased sucrose yields (Table 11-2). The -1 °C and -2 °C adaptations reduced sucrose yields by a greater extent (-10 and -20%) than the +1 °C and +2 °C adaptations increased them (+7 and +12%). Adaptations to relative canopy development rate had little impact, although slight reductions in stalk and sucrose yields are noted for a 10% decrease in this trait parameter value. Increasing the senescence threshold also resulted in very subtle impacts, with ADMh and SDMh increasing by a tiny margin, and a more substantial increase (≈ 2%) in sucrose yields with a 10% increase in senescence threshold.

Increasing temperature thresholds reduced ADMh slightly (2-5%) for mid- and late-season crops and vice versa, with a larger decrease for early-season crops (10%). SDMh followed a similar pattern, with slightly greater reductions in response to increasing temperature thresholds compared with ADMh. Mid- and late-season crops SUCh yields increased with the +1 and +2 °C Gs, responding more sensitively than early-season crops. This is because early-season crops had the coolest start, and the decrease in FIPARa and ADMh was greater than the increase in favourability for sucrose accumulation. SUCh yields always exceeded baseline SUCh yields for mid- and late-season crops. It is notable however that for early-season crops there were several seasons where the adapted Gs produced lower SUCh yields, although in most cases and on average, these Gs outperformed the baseline G.

Early-season crops were more sensitive to changes in relative canopy development rate than mid- or late-season crops, suggesting that only with these crops was canopy development rate sink- rather than source-limited, due to low temperatures.

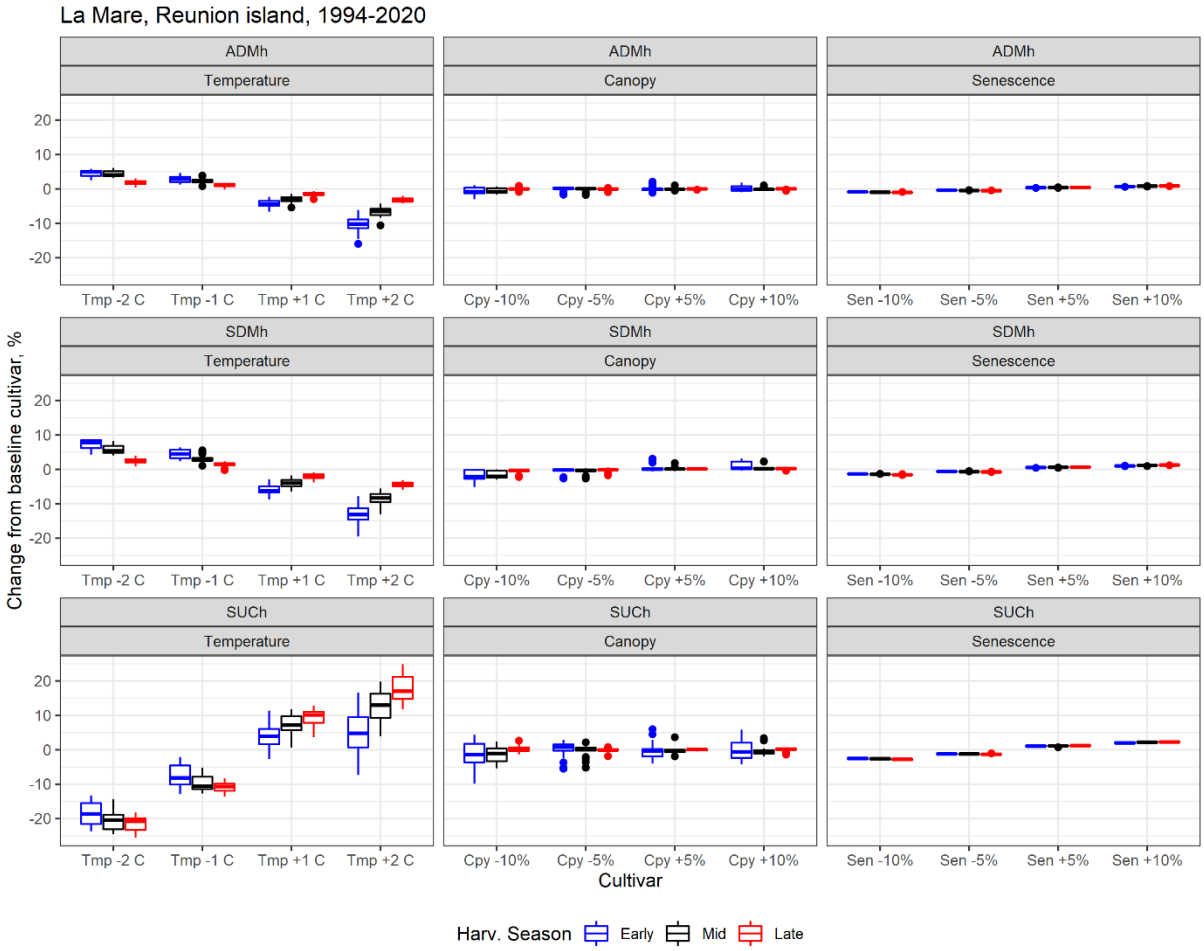

Figure 6-7. Boxplots of biomass variables for La Mare, 1994-2020, expressed as percentage changes from baseline cultivar, for early- (blue), mid- (black) and late-season (red) harvested crops. ADM = 'Above-ground dry biomass at harvest' (t/ha), SDM = 'Stalk dry mass at harvest' (t/ha), SUC = 'Sucrose dry mass at harvest' (t/ha). Genetic adaptations to temperature sensitivity, canopy growth rate and senescence threshold are illustrated.

6.3.5 Pongola, South Africa

At Pongola, for the baseline cultivar NCo376, germination was fastest (26 days) for late season crops starting in summer, and slowest (53 days) for mid-season crops starting in late winter (Table 11-2). Date of OSG occurred considerably sooner (83 days after crop start) for late-season crops with fast canopy development (Table 11-2, Figure 6-9), compared with mid- (156 days) and late-season (121 days) crops. Average ADMh, SDMh and SUCh yields were 53.3, 37.5 and 17.3 t/ha respectively over the period 1998-2020 (Table 11-2).

Early- and late-season crops produced higher biomass and stalk dry mass yields than mid-season crops (Table 11-2, Figure 6-8), due to greater radiation interception and higher radiation use efficiency. Sucrose yields were greatest for late-season crops, however. The dynamics of this, illustrated in Figure 6-8, warrants some explanation: (i) early-season crops, starting in late autumn, had very slow canopy development rate, leading to low biomass accumulation, and high temperatures leading up to harvest

favoured structural growth rather than sucrose accumulation; (ii) mid-season crops experienced faster canopy growth and biomass accumulation than early-season crops, and decreasing temperatures around 250 days after crop start drove up sucrose yields, but thereafter also reduced the rate of biomass accumulation towards harvest, reducing sucrose accumulation rates despite favourable conditions for partitioning to sucrose; (iii) late-season crops developed canopy the fastest and accumulated biomass in a near-linear fashion throughout the season; temperatures in the range 20-25 °C in the 3 months before harvest resulted in a “Goldilocks” situation of a favourable balance of sustained biomass accumulation and stalk partitioning to sucrose.

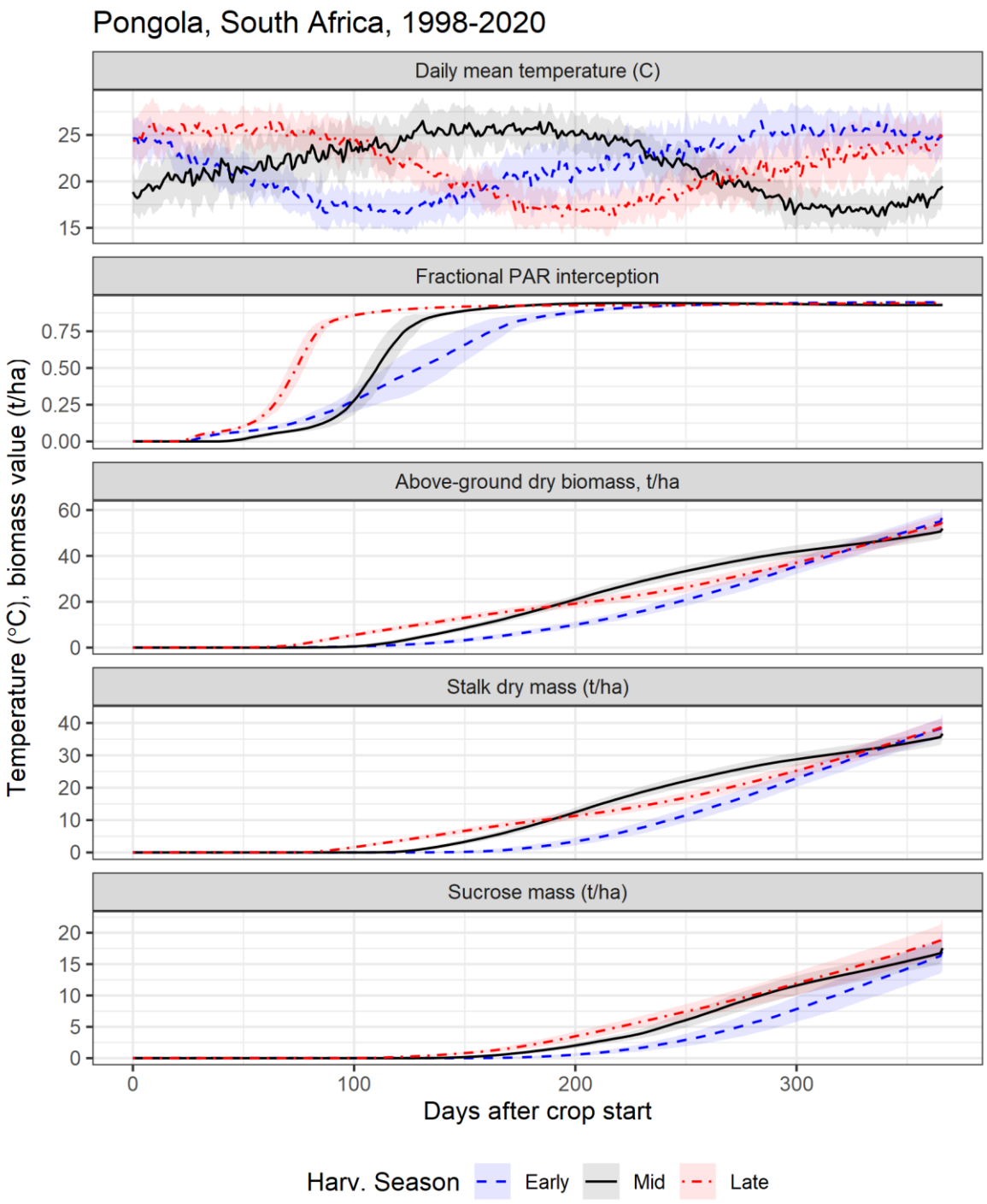

Figure 6-8. Daily average (±1 standard deviation, shaded) mean daily air temperature (C°), photosynthetically-active radiation interception fraction, above-ground dry biomass (t/ha), stalk dry mass (t/ha), and sucrose mass (t/ha) and for Pongola, 1998-2020, baseline cultivar (NCo376).

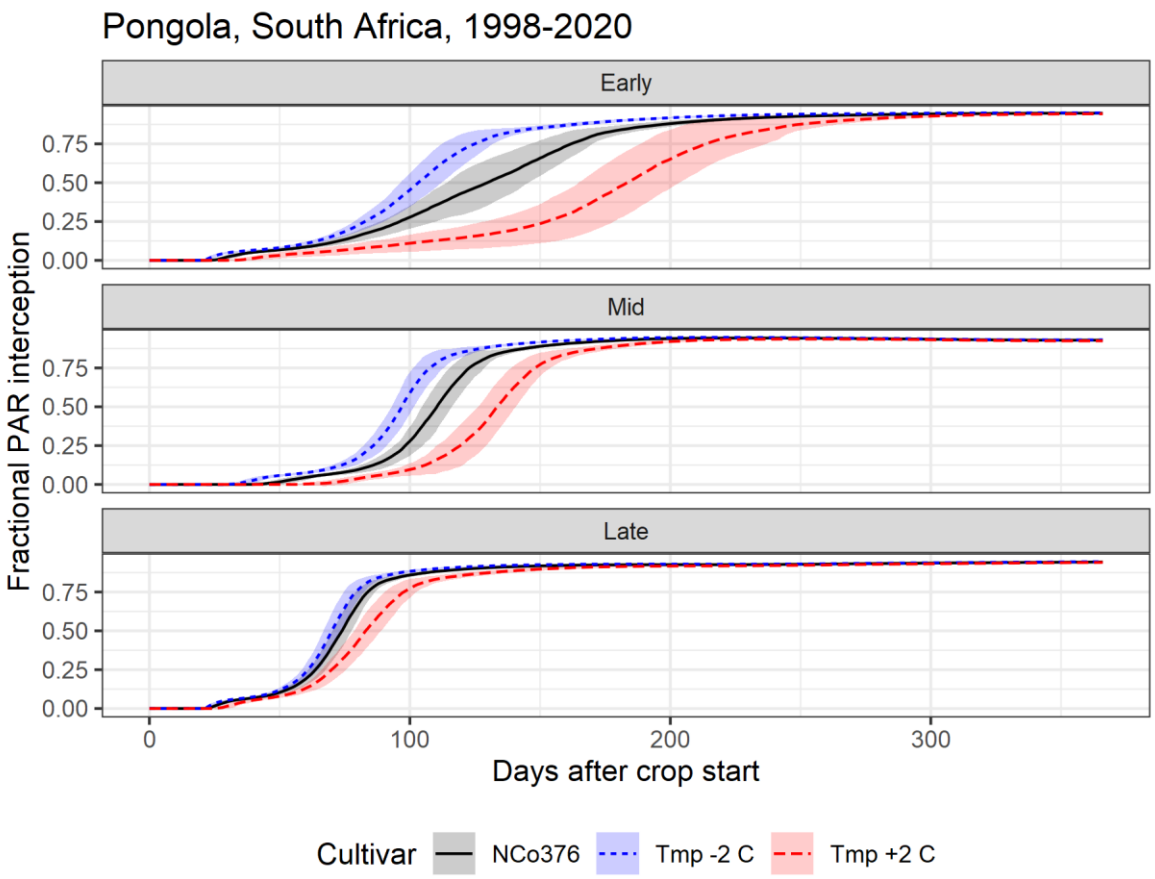

Figure 6-9. Daily mean photosynthetically-active radiation interception for the baseline cultivar (NCo376) and genotypes adapted to lower ("Tmp -2 C", blue, short-dashed series) and higher ("Tmp +2 C", red, long-dashed series) temperatures, at Pongola, South Africa, 1998-2020. Shading indicates 1 standard deviation above and below the mean.

A broadly similar pattern of ADMh temperature adaptation responses was evident at Pongola (Figure 6-10) as at the other sites: increasing the temperature thresholds penalised ADMh, due to reduced PAR interception (Figure 6-9). The impacts on SUCh were less clear-cut, however. SUCh decreased with decreasing temperature thresholds. The +1 °C and +2 °C adaptations also had lower SUCh yields than the baseline for early-season crops, suggesting that the baseline cultivar is near-optimal at Pongola for such crops. For mid- and late-season crops, however, the Gs adapted to higher temperatures reliably produced greater SUCh yields relative to the baseline G. It should be noted that the baseline cultivar NCo376 was bred and selected in South Africa, albeit for rainfed production in the 1950s when temperatures were approximately 0.75 °C lower than the last 20 years (Singels et al., 2005a).

Changes to the relative canopy development rate did not have much effect; ADMh, SDMh and SUCh responded positively to increased canopy development rate only for the early-season crops. As at the other sites, increasing the senescence threshold was (very weakly) beneficial for yields.

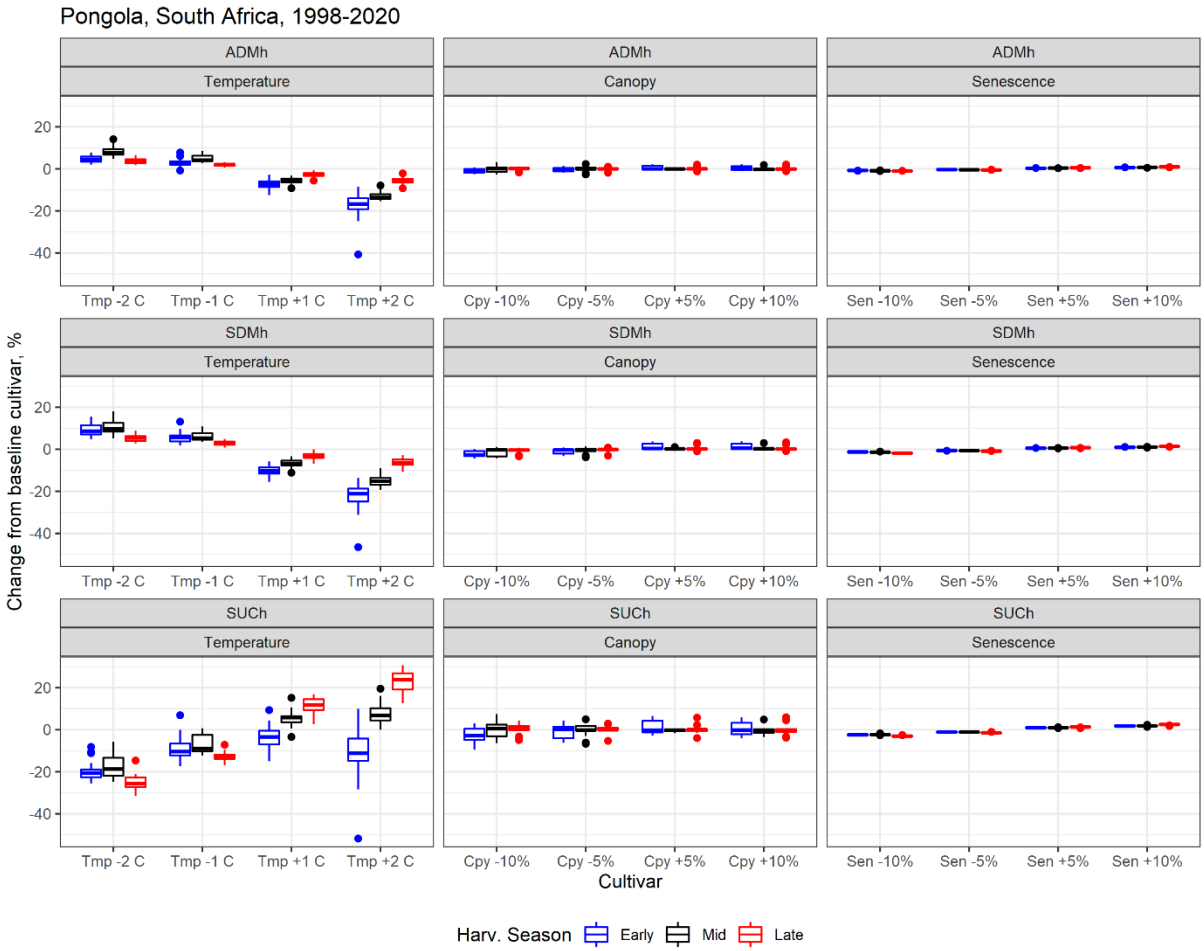

Figure 6-10. Boxplots of biomass variables for Pongola, 1998-2020, expressed as percentage changes from baseline cultivar, for early- (blue), mid- (black) and late-season (red) harvested crops. ADMh = 'Above-ground dry biomass at harvest' (t/ha), SDMh = 'Stalk dry mass at harvest' (t/ha), SUCh = 'Sucrose dry mass at harvest' (t/ha). Genetic adaptations to temperature sensitivity, canopy growth rate and senescence threshold are illustrated

6.4 Discussion

6.4.1 Temperature impacts

Overall, increasing the temperature thresholds – making the crop less tolerant of cold conditions – reduced biomass and stalk yields but increased sucrose yields. The basic mechanism is as follows:

1. Higher base and optimal temperatures for germination delayed emergence, which reduced PAR interception and this reduced biomass accumulation and yield;
2. Higher base and optimal temperatures for canopy growth slowed canopy development and so also reduced PAR interception and biomass accumulation; the combination of less biomass and delayed onset of stalk growth (which is linked to FIPAR) reduced stalk dry mass yields.
3. Higher base and optimal temperatures for stalk growth slowed stalk elongation and reduced the sink strength for stalk fibre, permitting greater accumulation of

'excess' photoassimilate to be stored as sucrose. This effect outweighed the losses of biomass and stalk mass.

From a sucrose accumulation perspective, the interaction with early-/mid-/late-season climatic conditions is manifested as a balance between these opposing forces of biomass accumulation and sucrose accumulation. In general, increasing temperature thresholds favoured late- and mid-season crops more than early-season crops, because this adaptation – under cool conditions at crop start for early-season crops – penalised biomass accumulation relatively more than mid- and late-season crops. Similarly, reducing the temperature thresholds (i.e. adapting the crop to a cooler E) was most unfavourable for late-season crops, as the additional partitioning to structural growth in the late winter and spring did not materially increase radiation interception (and therefore had limited benefit in terms of additional biomass accumulation), but nevertheless suppressed natural ripening.

Singels & Bezuidenhout (2002) described a biomass partitioning model that uses a temperature inflection point (calibrated as 25 °C) to define the temperature above which sugarcane (NCo376) partitions less to sucrose and more to non-sucrose sugars and stalk fibre. The CaneGEM results presented here are consistent with this, and indeed the temperature range 20-25 °C seemed to be particularly beneficial for sucrose accumulation for NCo376; higher temperatures favoured greater stalk mass but with less sucrose at Pongola. In the South African industry, late-season crops are considered to have low sucrose content and are often sprayed with chemical ripeners to increase sucrose content. In these simulations, sucrose yields were highest in late-season crops. This may indicate that the default base temperature for stalk expansive growth is too high, resulting in overestimated sucrose accumulation with the temperature adaptations.

A chemical ripener, trinexapac-ethyl, is sometimes applied to sugarcane to increase sucrose content (Van Heerden et al., 2015a). The mode of action is transient suppression of stalk elongation, the consequence of which is (when used appropriately) increased sucrose content and mass. Again, this model seems to emulate the impact of reduced stalk elongation on sucrose yield as an outcome rather than a prescribed descriptive behaviour.

In South Africa, cultivar N31 is considered suitable for growth in the high-altitude Midlands region, characterised by relatively cool winter conditions (SASRI, 2006). N31 is a high-biomass, low-sucrose cultivar. It is speculated that N31 was selected on the basis of cane yield, and that the favourable cane yields under Midlands conditions are due to being adapted to lower temperatures. The CaneGEM model captures this dynamic – lower sucrose yields in a cold-adapted cultivar.

The reduction of stalk length, 'dwarfism', has been exploited as a grain yield-enhancing mechanism in breeding crops such as wheat and rice (Peng et al., 2008; Richards, 2000), via reduced lodging and increased harvest index (implying biomass partitioning toward harvestable parts rather than the plant stem).

In sorghum, Hammer et al. (2010) simulated differences between Gs differing in canopy height: additional partitioning to grain was a consequence of a shorter growth traits, via reduced sink strength for stalk fibre and greater N available to the leaves. The greater leaf N concentration in shorter Gs meant that the leaf senescence was slowed. It is possible then that genetic adaptations to increase temperature thresholds for stalk structural growth would indirectly result in a higher leaf senescence threshold,

which was shown in the present study to be beneficial to sugarcane biomass and sucrose yields.

On this basis of the results presented in this chapter, with the possible caveat of late-season sucrose yields, the CaneGEM model outcomes and underlying mechanisms appear to be reasonably realistic and credible.

6.4.2 Faster canopy development and higher senescence threshold

Faster relative canopy growth rate was advantageous to biomass, stalk and sucrose yields, through increased radiation interception. This is consistent with Sinclair et al. (2004) who indicated that Gs with large early leaves appeared to yield well, and suggested that early leaf size could be used as a screening trait in selection. Singels et al. (2005c) reported that leaf size appeared to be determined by the size of the preceding leaf, indicating a source feedforward effect. Zhao et al. (2019) reported that normalised different vegetation index (NDVI) measurements taken during the partial canopy phase were strongly correlated with final yield in irrigated sugarcane in Florida, providing further evidence that early growth vigour favours yield accumulation. The relative canopy development parameter is an integrated trait parameter, representing the effects of leaves and tillers – so tillering and leaf size ought to be considered together for screening. One of the findings of Alam et al. (2017) was that tillering in sorghum is a growth process, resulting in more rapid canopy growth under radiation conditions and vice versa, but with clear genotypic propensity to tiller evident as well. The CaneGEM model simulation results revealed a weak yield interaction between relative canopy development rate and harvest season. Yields either increased or did not change in response to lower relative canopy development rate for late-season crops. For early-season crops, yields generally increased slightly in response to higher relative canopy development rate.

Increasing the canopy senescence threshold also increased yields, by increasing radiation interception. This aspect of the model is relatively weak, with unresolved E and possibly GxE interaction effects. There is most probably a strong interaction with plant N status as well. It is also possible that the range of *GLAIst* values considered in this case study was too narrow. Liu & Bull (2001) simulated leaf senescence by considering the carbon balance of each leaf, by comparing its photosynthetic income (in a layered radiation interception model) with its respiration costs. The authors also noted that leaf senescence occurs when the attached internode matures; if this maturity is determined by the sucrose content of the internode, it is possible that the senescence threshold would decrease as an emergent consequence of higher sucrose content in a G adapted to higher temperatures. This possibility was not explored directly, but the results indicate slight reductions in sucrose mass with decreasing senescence threshold. However, a 1 °C change in temperature sensitivity had a much greater impact on sucrose yields than a 10% change in senescence threshold, suggesting that if such a trade-off exists, it would be worth risking.

6.4.3 Significance and provisos for sugarcane breeding

<u>Cold tolerance is advantageous for increasing biomass and stalk yields</u>

Across all Es and harvest times of year, reducing cardinal temperatures by 1-2 °C resulted in increased biomass and stalk yields. This was due to more rapid canopy formation, which resulted in increased radiation interception and increased biomass

accumulation rates. Key to this is the assumption that temperature adaptation is coordinated across all processes in the plant (Parent and Tardieu, 2012).

Early growth vigour and can be phenotyped relatively easily in breeding trials, and could be rapidly phenotyped effectively in germination chambers (Poser et al., 2019) from the second selection stage onwards, when several buds for each G are available.

Breeding for high sucrose content is considered easier than breeding for high biomass, due to the high heritability of sucrose content. In principle then, selecting cold-tolerant Gs and then breeding into these high sucrose content traits would result in high-biomass Gs with high sucrose content – a very appealing prospect. Acreche et al. (2015) attributed genetic gain in sucrose yields to increasing cane yields in Argentina, rather than increasing sucrose content; and concluded that scope for further gains in sucrose yields was possible via increasing sucrose content.

This insight, that breeding for cold tolerance and high sucrose content together could increase both biomass and sucrose yields, can be seen as the major finding of this case study.

Adaptation to higher temperatures may be advantageous for increasing sucrose yields

The CaneGEM model makes use of a novel (for sugarcane) yet fairly simplistic model of sucrose accumulation. In this model, sucrose accumulation is entirely a passive outcome of source strength and competing structural sinks. If structural sink strengths are high (due, perhaps, to lower cardinal temperatures) then less biomass is 'left over' each day to be stored as sucrose. If this model is correct, raising (or even maintaining) sucrose content in new cold-adapted Gs will be difficult unless other genotypic changes are made that result in decreased carbon costs of structural sinks – via, for example, decreasing stalk fibre content, stalk elongation rates, or perhaps increasing specific leaf area.

This new sucrose partitioning model has not been as extensively tested and proven as the canopy development and aerial/stalk biomass partitioning aspects of the CaneGEM model. For this reason, the prediction of sucrose yields may be less accurate than predictions of biomass and stalk yields.

Acknowledging this, a second, more tentative take-home point from this work is then: adapting Gs to warmer environments – making them less tolerant of cold – *all else being equal,* may penalise biomass and stalk yields but considerably favour sucrose yields. Such an adaption might be necessary for maintaining current sucrose yields into a future with warming of 1-2 °C. This discussion is from the perspective of industries where sugar is the primary product of sugarcane, as this is often the case for irrigated sugarcane. For industries that produce bio-ethanol from sugarcane, the model outputs for stalk sugars can be used, although as there is a linear relationship between stalk sugar and sucrose, the outcomes are likely to be similar. For industries that produce, or will one day produce, cellulosic ethanol, or use sugarcane fibre for production of paper, plastics, or biomass energy, this model is similarly useful for identifying favourable traits and exploring new and changing growth environments.

6.4.4 Recommendations for model improvement

A key assumption in this work was that base and optimal temperatures for germination, canopy development and stalk elongation are coordinated – meaning that although these temperatures can be different between processes, they all increase or decrease to similar extent when the G is adapted to higher or lower temperatures. It is possible

however, that these cardinal temperatures are independently genotypically-controlled. Detailed experimentation is required to assess this. If independent G control is shown to be possible, the model could be used to explore a greater universe of adaptation possibilities. On the basis of the outcomes of this study, lower base temperatures for germination and canopy development combined with higher base temperatures for stalk elongation would result in favourable sucrose accumulation outcomes across most environments. The optimal extent of this needs to be determined for the time of harvest and temperature characteristics of the TPE. Given the critical importance of the base and optimal temperatures for stalk expansive growth (via its impact on sucrose accumulation) underlying the tradeoff with biomass accumulation, existing Gs need to be characterised carefully in this regard to ensure that simulations produce accurate recommendations – either for deploying existing varieties to specific environments, or for defining breeding targets for TPEs.

The dynamics of senescence appear to be poorly understood and the model's simulation of senescence is relatively weak. Senescence is likely to be an emergent and GxExM-sensitive consequence of canopy-related sub-traits such as leaf angle (more erect leaves would allow radiation to penetrate deeper into the canopy, possibly requiring less frequent leaf replacement), N concentration (more N-efficient Gs would support larger canopies with greater senescence thresholds; less fibre synthesis with higher temperature thresholds might allow more N to the canopy as the stalk would require less), and stalk elongation rate (more rapid elongation might shade lower leaves more quickly, requiring more frequent leaf replacement), and E traits such as latitude (solar declination angle) and the proportion of direct to diffuse radiation. Dedicated experimentation and more detailed canopy modelling are required to explore these dynamics.

The CaneGEM model is limited by not having a water balance or nitrogen balance, and it is recommended the model is incorporated into the DSSAT and/or APSIM systems, as these allow relatively easy addition of new plant modules (Jones et al., 2003; Jones, 2013). The CaneGEM model algorithms could also be incorporated into the Mosicas model, as it has a simple software structure, although this model does not have an N model. A further advantage of undertaking this integration would be to increase accessibility of the model to existing users familiar with the software interfaces.

6.4.5 Additional model applications

This model, with its unique carbon-linked canopy model and source:sink-based competitive biomass partitioning was demonstrated to have value for exploring G traits controlling temperature, canopy growth rate and senescence adaptations – resulting in outcomes significant and perhaps challenging to sugarcane breeding.

The model would also be useful for application in harvest planning and climate impact assessments, characterisation of current and future irrigated sugarcane growth environments, as well as ideotyping for future irrigated environments – for high-sucrose and high-biomass sugarcane Gs.

6.5 Conclusion

A new sugarcane model, which simulates canopy growth rate on the basis of a carbon-linked relative growth rate concept and partitioning of biomass based on source:sink competition, was used to evaluate genotypic adaptations to temperature sensitivity (1 and 2 °C above and below the cardinal temperature parameter values for the reference

cultivar NCo376), canopy growth rate and canopy senescence (NCo376 ±5 and 10%) for early-, mid- and late-season harvesting at four sites around the world: Belle Glade, Florida, USA; Chiredzi, Zimbabwe; La Mare, Reunion island, France; and Pongola, South Africa.

The temperature adaptations had the greatest impact, with decreases in the thresholds for temperature response (i.e. adaptation to cooler environments) resulting in increased in biomass and stalk yields at harvest. A feasible breeding strategy for increasing sucrose yields in irrigated sugarcane may be to focus on cold adaptation traits and then pursue increased sucrose content. The CaneGEM model also indicated that adaptation to warmer environments may enhance sucrose yields by reducing structural sink activity. This is a more tentative outcome considering that the focus of the model development work was on the canopy development and above-ground biomass partitioning processes, and less on sucrose accumulation.

The new CaneGEM sugarcane model has been demonstrated to be valuable for exploring G adaptations to temperature in particular, taking into account trade-offs that appear to be realistic, that would be difficult or impossible with other sugarcane models. The model has potential for application in identifying temperature- and canopy-related traits that lend favourable adaptability to different irrigated sugarcane production environments. From a model development perspective, it is recommended (a) that the assumption of coordination of temperature thresholds across germination, canopy development and stalk elongation processes is tested more rigorously with carefully-designed experiments; and, (b) that the model is incorporated into a modelling system such as DSSAT or APSIM to link to water and nitrogen balances and permit greater accessibility for users already familiar with such modelling system interfaces.

7. ANSWERS TO RESEARCH QUESTIONS, AND RECOMMENDATIONS

TABLE OF CONTENTS

7.1 Answers to research questions.....139
7.1.1 Are observed genotypic differences adequately captured by process concepts and their respective G-specific input parameters in existing sugarcane models?.....139
7.1.2 To what extent are existing sugarcane crop growth models suitable for supporting plant breeding applications? 139
7.1.3 To what extent can the prediction of these GxE interaction effects be improved if the identified shortcomings are addressed in a new sugarcane model? 140
7.1.4 How would such an improved model deliver value to sugarcane breeding? 140
7.1.5 How has this study contributed to an improved understanding of the physiological basis for GxE interactions in irrigated sugarcane? 141
7.2 Recommendations.....142
7.2.1 Sugarcane breeding.....142
7.2.2 Sugarcane physiology and crop growth model development 143
7.2.3 Model application in other domains.....144

7.1 Answers to research questions

Four research questions were posed at the outset of this study; a set of four broad research activities, described in Chapters 3-6, respectively addressed these questions.

7.1.1 Are observed genotypic differences adequately captured by process concepts and their respective G-specific input parameters in existing sugarcane models?

The analysis of the ICSM IGEP dataset (Chapter 3) revealed shortcomings in three process concepts used in established sugarcane crop growth simulation models: the use of thermal time (based on air temperature) to model germination rate; and the absence of solar radiation influences in the prediction of time to onset of stalk growth, and unstressed tillering rate. The use of G-specific cardinal temperatures for germination and canopy development-related model processes was recommended (Chapter 4) for capturing GxE interactions in canopy development and by implication, biomass yields.

7.1.2 To what extent are existing sugarcane crop growth models suitable for supporting plant breeding applications?

Three widely-used sugarcane crop growth models (DSSAT-Canegro, APSIM-Sugar and Mosicas) were assessed for their abilities to predict observed E, G and GxE

interaction effects in the ICSM IGEP dataset (Chapter 4). All three models were capable of predicting E and G effects reasonably accurately, which suggests that they could be used effectively for environmental characterisation applications in sugarcane breeding. They were however unable to predict GxE interaction effects in seasonal radiation interception (FIPARa), radiation use efficiency (RUEa), and biomass yields (ADMh). Recommendations for model improvement arising from the model assessment included using simulated soil temperatures for calculating germination rate, and linking canopy development with biomass accumulation via a carbon balance in order to account for source:sink dynamics affecting radiation interception and biomass partitioning.

7.1.3 To what extent can the prediction of these GxE interaction effects be improved if the identified shortcomings are addressed in a new sugarcane model?

Following the recommendations outlined in Chapters 3 and 4, recommendations made in previous research for more accurate representation of sugarcane physiology, and general requirements of crop growth models for credible application in plant breeding, a new sugarcane crop growth simulation model, CaneGEM, was developed. This model differs from other sugarcane models in three key respects, which address the identified shortcomings described in Chapters 3 and 4. Firstly, it predicts the transition from tillering to stalk growth on the basis of radiation interception rather than thermal time. Secondly, the canopy is represented in an integrated manner (and does not simulate individual shoots and leaves), with canopy development based on a relative growth rate concept, restricted by temperature and source strength. Thirdly, biomass partitioning, particularly within the stalk, is determined as a consequence of source strength and sink strength for structural stalk fibre, which is determined by the degree of transition to stalk growth, temperature and specific stalk fibre density. G-specific parameters regulate the range in specific leaf area, specific stalk volume, relative canopy growth rates, canopy senescence radiation threshold and relative senescence rate, and cardinal temperatures governing temperature responses for germination, canopy development, biomass accumulation and stalk expansion.

The model was able to predict GxE interaction effects on FIPARa more accurately than the DSSAT-Canegro model in a direct comparison, and more accurately than the three models assessed in Chapter 4. With G-specific calibration of cardinal temperatures for several plant processes, accurate and statistically-significant prediction of GxE interaction effects in FIPARa and ADMh was achieved. However, it was discovered that a considerable proportion of GxE interaction in FIPARa and (consequently) ADMh appears to be driven by GxE interactions in the duration of the germination phase (leading to differences in date of primary shoot emergence).

The CaneGEM model was not able to predict GxE interaction effects accurately unless accurate dates of emergence were correctly specified (not simulated) in advance. Attempts to improve the prediction of duration of germination phase, including using simulated soil temperature instead of air temperature, were unsuccessful.

7.1.4 How would such an improved model deliver value to sugarcane breeding?

The case study described in Chapter 6 explored three types of genotypic adaptations for each of the ICSM IGEP study sites, for early-, mid- and late-season harvests, using long-term weather datasets. Increasing relative canopy development rate and canopy senescence thresholds by 5 and 10% resulted in small increases in radiation

interception, biomass, stalk and sucrose yields across all sites and dates of harvest, while decreasing these had the opposite effect. These responses were limited in magnitude, probably due to the canopy development plasticity permitted by variable specific leaf area and the large green leaf area index characteristic of irrigated sugarcane.

The response to temperature adaptation was most interesting and perhaps shows where the real value of the model lies for breeding. Relatively small changes to cardinal temperature parameters resulted in substantial differences across several complex traits. G adaptation to lower temperatures resulted in increased biomass and stalk yields, as expected, but also considerable sucrose yield penalties, a consequence of greatly increased sink strength for stalk fibre. Targeting cold adaptation in breeding may however provide parent material for subsequent breeding of higher sucrose content, which may permit the development of Gs with high biomass and sucrose yields. The source:sink-based partitioning, and linking of biomass accumulation with canopy development and phenology, means that this model has built-in inter-process couplings that result in natural, biologically-realistic tradeoffs that can be explored *in silico* for trait dissection and ideotyping. Ideally these interactions should be confirmed and/or refined experimentally.

If the assumption, that temperature adaptation is coordinated across plant processes, proves to be correct, it might be possible to associate Quantitative Trait Loci with temperature adaptations. Running the model with QTL-predicted temperature adaptations may permit early screening of new crosses for suitability to different Es with the target populations of environments. Phenotyping for cold tolerance and early growth vigour – which could be done relatively easily – holds promise for early screening in breeding programmes.

More broadly, the CaneGEM model fulfils the requirements of models for supporting plant breeding: complex traits are predicted as the emergent consequences of lower-level processes regulated by simple, stable traits; most trait parameters can be related to or inferred from phenotypic measurements; the model was demonstrated to have the potential to predict GxE interaction effects; and the model is parsimonious, being relatively simple and computationally undemanding.

7.1.5 How has this study contributed to an improved understanding of the physiological basis for GxE interactions in irrigated sugarcane?

This question is perhaps least convincingly answered. Observed GxE crossover effects in seasonal radiation interception and biomass yields were accurately predicted by the CaneGEM model and explained in terms of: (unresolved) GxE differences in duration of germination phase; cardinal temperatures for canopy development; and relative canopy growth rate.

The outcomes of the analysis of the ICSM IGEP dataset, literature, outcomes of the CaneGEM validation and GxE interactions assessment, as well as the case study, can be synthesised as follows:

- Higher radiation use efficiency is favourable for biomass, cane and sucrose yields – but it remains unclear whether there is much genotypic variability in this trait amongst elite sugarcane genotypes. The CaneGEM model was able to predict GxE interactions in biomass yields accurately without genotype-specific maximum radiation use efficiency values.

- Rapid germination is favourable for more rapid canopy development; more rapid canopy development results in increased seasonal radiation interception and earlier onset of stalk growth, permitting an earlier start to cane yield accumulation and greater opportunity for sucrose accumulation.
- Early canopy growth vigour is favourable for yield accumulation, for similar reasons as outlined above. It is noted that from a phenotyping perspective, rapid germination and high early canopy growth vigour may be difficult to distinguish, but that this does not matter – the outcomes are similar.
- Genotypic adaptation to lower temperatures is favourable for biomass and cane yields. This is, all else being equal, more important in cooler environments, particularly environments with a cool start to the season: sugarcane is able to photosynthesise at lower temperatures than it is able to grow expansively; by ensuring that the crop has a larger canopy earlier on, more radiation can be intercepted and larger yields will result. A unit change in temperature response will have a proportionally greater impact in a cool environment than a warm one – meaning that prospects for genetic gain (via temperature adaptations) in biomass and cane yields is likely to be greater in cooler production areas than warm ones.
- Genotypic adaptation to higher temperatures may be favourable for sucrose yields. A genotype adapted to higher temperatures is able to exert weaker structural sink strengths, which has the disadvantage of decreasing canopy development rate and radiation interception; but, the weaker sink strength for structural growth means that additional source is available to be stored as sucrose, with the result that the plant has a higher natural ripening ability. This effect is exacerbated in environments where the pre-harvest period is relatively warm, such as late-season crops. In this case, the optimal tradeoff between canopy vigour and natural ripening ability needs to be determined, perhaps using a carefully-calibrated model such as CaneGEM. This outcome can be considered tentative, as reduced sink strength for stalk fibre as a result of higher cardinal temperatures has not be demonstrated empirically.

7.2 Recommendations

7.2.1 Sugarcane breeding

Four strong recommendations can be made from this study, that could inform future breeding of irrigated sugarcane:

- Firstly, all three of the existing sugarcane models tested (DSSAT-Canegro, APSIM-Sugar and Mosicas), with their standard trait parameter values, are sufficiently sensitive to E differences for environmental characterisation applications in breeding. It is recommended these models are used in breeding programmes for applications such as characterising abiotic stress patterns for weighting selection and cultivar evaluation and identifying suitable sites for inclusion in multi-environment trials.
- Secondly, a considerable proportion of the GxE interactions in seasonal radiation interception and biomass accumulation appears to be driven by GxE interactions in duration of the germination phase. It follows then that germination rate is an overwhelmingly powerful trait for determining sugarcane biomass accumulation potential. It is recommended that phenotyping of date of primary shoot emergence (or possibly early canopy development rate) be explored for assisting selection in irrigated sugarcane.

- Thirdly, it was revealed in the case study in Chapter 6 that adapting the plant to grow more vigorously at lower temperatures resulted in enhanced radiation interception and increased biomass yields, but generally also resulted in very significant decreases in sucrose yields, particularly for cooler environments (and *vice versa*). Projected global warming over the next 30 years will have the same practical effect as the adaptations to lower temperatures described in this study.
- Fourthly, following the previous point, sucrose accumulation does of course depend on biomass accumulation, notwithstanding the trade-off between these in terms of temperature adaptation. It is recommended that the new CaneGEM sugarcane model developed in this study (or some derivative thereof) is used to explore the biomass-sucrose mass trade-off space for each E and M combination in sugarcane breeding programmes, in order to identify optimal temperature adaptation breeding targets for current and/or expected future Es.

7.2.2 Sugarcane physiology and crop growth model development

In Chapters 5 (new model development) and 6 (case study), it was assumed that temperature responses are coordinated across processes. This makes intuitive physiological sense and is supported in broad terms by literature. However, given that sugarcane Gs are interspecific hybrids, it is recommended that this is demonstrated experimentally.

It is clear from Chapters 3 and 5 that the standard model of predicting date of emergence based on thermal time accumulation (from air temperature) is inadequate, particularly for ratoon crops, especially given the importance of this trait in driving GxE interactions in biomass yields. It appears that initiation of ratoon crop germination starts before harvest of the previous crop, and so an improved model of ratoon crop germination will need to consider radiation, temperature, water status and perhaps other conditions prior to harvest of the previous crop.

The simulation (and understanding) of leaf/canopy senescence remains simplistic. There is likely to be value in better understanding and elucidating the mechanisms underlying senescence, as replacing green leaf area in a mature crop comes at the expense of stalk and sucrose mass accumulation.

The new sugarcane crop growth model was developed and validated using a relatively large number of NCo376 trial datasets. The calibration and assessment of the model for the ICSM IGEP cultivars (N41, R570 and CP88-1762) relied on a relatively small dataset, with some shortcomings (as identified in Chapter 3). Repeating (with more consistent sampling protocols) or extending the experiment (by adding more Es, including date of harvest or formal water stress treatments) will lead to greater clarity and permit further model refinement.

The CaneGEM model developed in this study does not simulate water or nitrogen balances. Additionally, users of existing sugarcane crop models have already invested considerable time and effort into training and preparation of model inputs and software workflows for processing model outputs. Rather than adding algorithms for water and nitrogen dynamics, it is recommended that the novel algorithms developed in this project should be integrated into codebases for existing mature sugarcane crop models. Water and N-stress linkages to these processes are expected to be simple to implement, but further model calibration and testing would be required.

7.2.3 Model application in other domains

The newly-developed sugarcane model will be valuable outside of operational plant breeding, particularly if the novel algorithms are incorporated into an established modelling system. The temperature sensitivity and coupled biomass partitioning makes the model eminently suitable for exploring date of harvest strategies, agronomic management for biofuels and biomass production from sugarcane, climate change impacts assessment, and genotypic adaptations to climate change. The CaneGEM model is unlikely however to offer much advantage over well-calibrated existing models for applications where E differences are of primary importance, such as crop forecasting and irrigation scheduling.

8. CONCLUSION

The adequacy of common sugarcane crop model concepts for capturing genotype-by-environment (GxE) interaction effects on sugarcane growth parameters was assessed using an international multi-E, multi-G dataset collected as part of the International Consortium for Sugarcane Modelling's "International GxE Project" (ICSM IGEP). Genotype-specific phenotypic parameters based on process-level concepts from several sugarcane crop growth models were evaluated for their stability in value and consistency in ranking between environments. Key outcomes included: maximum radiation use efficiency appeared to be genotype-specific and consistent between environments; leaf appearance rate is stably genotype-specific; duration of the germination phase was poorly predicted by the standard thermal time model, with unresolved GxE effects, but also that final biomass yields were correlated with duration of the germination phase; and, the date of onset of stalk growth depends on radiation intensity in addition to temperature.

The same ICSM IGEP dataset was used to evaluate the performance of existing widely-used sugarcane crop growth simulation models for predicting GxE interactions in seasonal radiation interception and final biomass yields. The DSSAT-Canegro, APSIM-Sugar and Mosicas models were calibrated based on the mean phenotypic values per genotype determined in the previous step; performance in predicting emergent outcomes was assessed (as no separate validation dataset was available), as were the models' abilities to predict GxE interaction effects in seasonal radiation interception, apparent radiation use efficiency, biomass yields and stalk yields. All models were found to predict environmental and genotypic main effects reasonably accurately, but were unable to predict GxE interaction effects.

Weaknesses identified during these two activities were used to inform the development of a new sugarcane model with novel algorithms for predicting (1) the transition from tillering to stalk growth in response to fractional radiation interception; (2) leaf area index on the basis of temperature- and carbon-limited relative growth rate; and (3) biomass partitioning as a consequence of competition sink and source strengths. This model was calibrated and validated using existing NCo376 growth analysis datasets, revealing similar performance to the established DSSAT-Canegro model. E, G and GxE interaction effects in seasonal radiation interception and final biomass yields in the ICSM IGEP were also assessed, and prediction performance was shown to be superior to existing models.

This process revealed that the prediction of the date of emergence was (and likely is in general) a powerful driver of these GxE interaction effects. GxE interaction effects were found also to be affected by G-specific cardinal temperatures for canopy development and relative canopy development rate. In warm environments in general, plants germinated and developed canopy cover quickly, leading to greater seasonal radiation interception and greater biomass yields, and vice versa. Lower relative canopy development rate (a G trait) could be compensated by warmer conditions (the case of CP88-1762 at Belle Glade), while a larger relative canopy development rate could compensate to some extent for cool conditions (R570 at Belle Glade) but slow germination in such conditions appeared to have a powerful diminishing effect on yields. The standard thermal-time based germination simulation algorithm, as implemented in this model and many others, was found to be wholly inadequate for predicting GxE interaction effects in duration of germination phase.

Value of the CaneGEM model in sugarcane breeding was demonstrated in a case study, where G adaptations to temperature response, relative canopy development rate and leaf senescence threshold were assessed, for early-, mid-and late-season harvesting, using long-term weather datasets at each of the four sites in the ICSM IGEP dataset. This revealed that biomass and cane yields could be increased via genotypic temperature adaptations to cooler environments (that is, reducing the cardinal temperatures for key plant processes); doing so, however, generally resulted in sucrose yield penalties. The positive impacts on biomass yields were largest for mid-season crops at most sites, where germination rates and canopy development rates were considerably higher during the winter months compared to the non-adapted baseline genotype, leading to additional radiation interception. The negative impacts on sucrose accumulation were strongest for crops harvested late in the season: the same temperature adaptation suppressed natural ripening during late winter and spring, while the concomitant additional structural canopy growth was not beneficial as the crop was already intercepting nearly all the incident radiation.

Several recommendations are made for future work. For sugarcane breeding, existing sugarcane models can and should be used for environmental characterisation applications, as these were shown to be sufficiently accurate; using date of emergence and/or early canopy growth vigour as a screening phenotype for irrigated sugarcane selection should be explored, as this had a powerful bearing on biomass yields; efforts should if possible focus on breeding Gs adapted to higher temperatures and/or with lower stalk fibre content, in order to maintain sucrose yields in a warming climate; and that the new sugarcane model be used to explore GxExM temperature adaptations to inform breeding targets in this regard. From sugarcane physiology/model development perspectives, the coordination of temperature responses across plant processes needs to be evaluated; research ought to be undertaken to properly understand the mechanics of ratoon crop germination; additional GxE experiments with more consistent protocols and/or additional treatments ought to be conducted to support model refinement; and it is recommended that the novel process algorithms developed for the CaneGEM model are included in one or more mainstream crop modelling platforms, in order to take advantage of existing water and nitrogen modules and ensure compatibility with existing users, data and processing workflows.

It was hypothesised that crop simulation modelling capacity to support breeding of irrigated sugarcane could be enhanced by:

- evaluating the strengths and weaknesses of existing sugarcane models and/or their constituent process-level concepts for predicting GxE interaction effects, observed in the ICSM IGEP trials, on radiation interception, radiation use efficiency and biomass yields; and
- if necessary, developing an improved (new or revised) sugarcane crop model to address identified weaknesses

With the possible caveat of the inadequate germination model, this hypothesis has been shown to be true.

This work has resulted in enhanced understanding of the drivers of GxE interactions in biomass yields in irrigated sugarcane, new insights into breeding for current and future irrigated sugarcane growth environments, and the development of a practical crop growth simulation model for supporting breeding of irrigated sugarcane.

9. REFERENCES

Acreche, M.M., 2017. Nitrogen-, water- and radiation-use efficiencies affected by sugarcane breeding in Argentina. Plant Breeding 136, 174–181. https://doi.org/https://doi.org/10.1111/pbr.12440

Acreche, M.M., Saez, J. V, Chalco Vera, J., 2015. Physiological bases of genetic gains in sugarcane yield in Argentina. Field Crops Res 175, 80–86. https://doi.org/https://doi.org/10.1016/j.fcr.2015.02.002

Alam, M.M., Hammer, G.L., van Oosterom, E.J., Cruickshank, A.W., Hunt, C.H., Jordan, D.R., 2014. A physiological framework to explain genetic and environmental regulation of tillering in sorghum. New Phytologist 203, 155–167. https://doi.org/10.1111/nph.12767

Alam, M.M., van Oosterom, E.J., Cruickshank, A.W., Jordan, D.R., Hammer, G.L., 2017. Predicting Tillering of Diverse Sorghum Germplasm across Environments. Crop Sci 57, 78–87. https://doi.org/10.2135/cropsci2016.04.0262

Allison, J.C.S., Pammenter, N.W., Haslam, R.J., 2007. Why does sugarcane (Saccharum sp. hybrid) grow slowly? South African Journal of Botany 73, 546–551. https://doi.org/10.1016/j.sajb.2007.04.065

Anderson, R.G., Tirado-Corbalá, R., Wang, D., Ayars, J.E., 2015. Long-rotation sugarcane in Hawaii sustains high carbon accumulation and radiation use efficiency in 2nd year of growth. Agric Ecosyst Environ 199, 216–224. https://doi.org/https://doi.org/10.1016/j.agee.2014.09.012

Balsalobre, T.W.A., da Silva Pereira, G., Margarido, G.R.A., Gazaffi, R., Barreto, F.Z., Anoni, C.O., Cardoso-Silva, C.B., Costa, E.A., Mancini, M.C., Hoffmann, H.P., de Souza, A.P., Garcia, A.A.F., Carneiro, M.S., 2017. GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. BMC Genomics 18, 72. https://doi.org/10.1186/s12864-016-3383-x

Basnayake, J., Jackson, P.A., Inman-Bamber, N.G., Lakshmanan, P., 2015. Sugarcane for water-limited environments. Variation in stomatal conductance and its genetic correlation with crop productivity. J Exp Bot 66, 3945–3958. https://doi.org/10.1093/jxb/erv194

Basnayake, J., Jackson, P.A., Inman-Bamber, N.G., Lakshmanan, P., 2012. Sugarcane for water-limited environments. Genetic variation in cane yield and sugar content in response to water stress. J Exp Bot 63, 6023–6033. https://doi.org/10.1093/jxb/ers251

Baucum, L.E., Rice, R.W., 2009. An overview of Florida sugarcane. EDIS 2009.

Bell, M.J., Garside, A.L., 2005. Shoot and stalk dynamics and the yield of sugarcane crops in tropical and subtropical Queensland, Australia. Field Crops Res 92, 231–248. https://doi.org/https://doi.org/10.1016/j.fcr.2005.01.032

Bezuidenhout, C.N., O'Leary, G.J., Singels, A., Bajic, V.B., 2003. A process-based model to simulate changes in tiller density and light interception of sugarcane crops. Agric Syst 76, 589–599. https://doi.org/10.1016/S0308-521X(02)00076-8

Bezuidenhout, C.N., Singels, A., 2007. Operational forecasting of South African sugarcane production: Part 1 - System description. Agric Syst 92, 23–38. https://doi.org/10.1016/j.agsy.2006.02.001

Bojanowski, J.S., 2016. sirad: Functions for Calculating Daily Solar Radiation and Evapotranspiration.

Bonnett, G., 2014. Developmental stages (phenology), in: Botha, P.H.M. and F.C. (Ed.), Sugarcane: Physiology, Biochemistry, and Functional Biology. Oxford: John Wiley & Sons, Ltd, pp. 35–53.

Bonnett, G.D., 1998. Rate of leaf appearance in sugarcane, including a comparison of a range of varieties. Functional Plant Biology 25, 829–834.

Boote, K.J., Jones, J.W., Hoogenboom, G., 2021. Incorporating Realistic Trait Physiology into Crop Growth Models to Support Genetic Improvement. In Silico Plants. https://doi.org/10.1093/insilicoplants/diab002

Boote, K.J., Jones, J.W., Hoogenboom, G., Pickering, N.B., 1998. The CROPGRO model for grain legumes, in: Understanding Options for Agricultural Production. Springer, pp. 99–128.

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère, J.P., Hénault, C., Maraux, F., Seguin, B., Sinoquet, H., 2003. An overview of the crop model stics. European Journal of Agronomy 18, 309–332. https://doi.org/https://doi.org/10.1016/S1161-0301(02)00110-7

Buis, S., Wallach, D., Guillaume, S., Varella, H., Lecharpentier, P., Launay, M., Guérif, M., Bergez, J.-E., Justes, E., 2011. The STICS Crop Model and Associated Software for Analysis, Parameterization, and Evaluation. Methods of Introducing System Models into Agricultural Research, Advances in Agricultural Systems Modeling. https://doi.org/https://doi.org/10.2134/advagricsystmodel2.c14

Campbell, J.A., Robertson, M.J., Grof, C.P.L., 1998. Temperature effects on node appearance in sugarcane. Functional Plant Biology 25, 815–818.

Chenu, K., Chapman, S.C., McLean, G., Lush, D., Hammer, G.L., Dreccer, F., 2009a. Crop modelling as an aid for environmental characterisation and crop improvement. Comp Biochem Physiol A Mol Integr Physiol 153, S227.

Chenu, K., Chapman, S.C., Tardieu, F., McLean, G., Welcker, C., Hammer, G.L., 2009b. Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: a “gene-to-phenotype” modeling approach. Genetics 183, 1507–1523.

Christina, M., Jones, M.-R.M.-R., Versini, A., Mézino, M., Le Mézo, L., Auzoux, S., Soulié, J.C., Poser, C., Gérardeaux, E., Le Mezo, L., Auzoux, S., Soulie, J.-C., Poser, C., Gérardeaux, E., 2021. Impact of climate variability and extreme rainfall events on sugarcane yield gap in a tropical Island. Field Crops Res 274, 108326. https://doi.org/https://doi.org/10.1016/j.fcr.2021.108326

Coelho, A.P., Dalri, A.B., Fischer Filho, J.A., Faria, R.T. de, Silva, L.S., Gomes, R.P., 2020. Calibration and evaluation of the DSSAT/Canegro model for sugarcane cultivars under irrigation managements. Revista Brasileira de Engenharia Agrícola e Ambiental.

Cooper, M., Byth, D.E., 2002. Understanding Plant Adaptation to Achieve Systematic Applied Crop Improvement – A Fundamental Challenge, in: Cooper, M., Hammer, G.L. (Eds.), Plant Adaptation and Crop Improvement (2002). CAB International, Wallingford, Oxon OX10 8DE, United Kingdom.

Cooper, M., Gho, C., Leafgren, R., Tang, T., Messina, C., 2014. Breeding drought-tolerant maize hybrids for the US corn-belt: discovery to product. J Exp Bot 65, 6191–6204.

Cooper, M., Messina, C.D., 2021. Can We Harness "Enviromics" to Accelerate Crop Improvement by Integrating Breeding and Agronomy? Front Plant Sci.

Cooper, M., Powell, O., Voss-Fels, K., Messina, C., Gho, C., Podlich, D., Technow, F., Chapman, S., Beveridge, C., Ortiz-Barrientos, D., Hammer, G., 2020a. Modelling selection response in plant breeding programs using crop models as mechanistic gene-to-phenotype (CGM-G2P) multi-trait link functions. In Silico Plants. https://doi.org/10.1093/insilicoplants/diaa016

Cooper, M., Tang, T., Gho, C., Hart, T., Hammer, G., Messina, C., 2020b. Integrating genetic gain and gap analysis to predict improvements in crop productivity. Crop Sci 60, 582–604. https://doi.org/https://doi.org/10.1002/csc2.20109

Cooper, M., Voss-Fels, K.P., Messina, C.D., Tang, T., Hammer, G.L., 2021. Tackling G × E × M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. Theoretical and Applied Genetics 134, 1625–1644. https://doi.org/10.1007/s00122-021-03812-3

Crossa, J., Fritsche-Neto, R., Montesinos-Lopez, O.A., Costa-Neto, G., Dreisigacker, S., Montesinos-Lopez, A., Bentley, A.R., 2021. The Modern Plant Breeding Triangle: Optimizing the Use of Genomics, Phenomics, and Enviromics Data. Front Plant Sci.

Cursi, D.E., Hoffmann, H.P., Barbosa, G.V.S., Bressiani, J.A., Gazaffi, R., Chapola, R.G., Fernandes Junior, A.R., Balsalobre, T.W.A., Diniz, C.A., Santos, J.M., Carneiro, M.S., 2022. History and Current Status of Sugarcane Breeding, Germplasm Development and Molecular Genetics in Brazil. Sugar Tech 24, 112–133. https://doi.org/10.1007/s12355-021-00951-1

de Leon, N., Jannink, J., Edwards, J.W., Kaeppler, S.M., 2016. Introduction to a special issue on genotype by environment interaction. Crop Sci 56, 2081–2089.

De Silva, A.L.C., De Costa, W.A.J.M., 2012. Growth and Radiation Use Efficiency of Sugarcane Under Irrigated and Rain-fed Conditions in Sri Lanka. Sugar Tech 14, 247–254. https://doi.org/10.1007/s12355-012-0148-y

Dias, H.B., Inman-Bamber, G., 2020. Sugarcane: Contribution of Process-Based Models for Understanding and Mitigating Impacts of Climate Variability and Change on Production, in: Systems Modeling. Springer, pp. 217–260.

Dias, H.B., Inman-Bamber, G., Bermejo, R., Sentelhas, P.C., Christodoulou, D., 2019. New APSIM-Sugar features and parameters required to account for high sugarcane yields in tropical environments. Field Crops Res 235, 38–53. https://doi.org/10.1016/j.fcr.2019.02.002

Dias, H.B., Inman-Bamber, G., Everingham, Y., Sentelhas, P.C., Bermejo, R., Christodoulou, D., 2020. Traits for canopy development and light interception by

twenty-seven Brazilian sugarcane varieties. Field Crops Res 249, 107716. https://doi.org/https://doi.org/10.1016/j.fcr.2020.107716

Dias, H.B., Sentelhas, P.C., 2018. Drying-Off Periods for Irrigated Sugarcane to Maximize Sucrose Yields Under Brazilian Conditions. Irrigation and Drainage 67, 527–537. https://doi.org/https://doi.org/10.1002/ird.2263

Diepenbrock, C., Tang, T., Jines, M., Technow, F., Lira, S., Podlich, D., Cooper, M., Messina, C., 2021. Can we harness digital technologies and physiology to hasten genetic gain in U.S. maize breeding? https://doi.org/10.1101/2021.02.23.432477

Donald, C.M. t, 1968. The breeding of crop ideotypes. euphytica 17, 385–403.

Donaldson, R.A., 2009. Season effects on the potential biomass and sucrose accumulation of some commercial cultivars of sugarcane. University of KwaZulu-Natal, Pietermaritzburg.

Dumont, T., Barau, L., Thong-Chane, A., Dijoux, J., Mellin, M., Daugrois, J., Hoarau, J.-Y., 2021. Sugarcane Breeding in Reunion: Challenges, Achievements and Future Prospects. Sugar Tech. https://doi.org/10.1007/s12355-021-00998-0

Engelbrecht, F.A., Monteiro, P.M.S., 2021. The IPCC Assessment Report Six Working Group 1 report and southern Africa: Reasons to take action. S Afr J Sci 117.

Everingham, Y., Inman-bamber, G., Sexton, J., Stokes, C., 2015. A Dual Ensemble Agroclimate Modelling Procedure to Assess Climate Change Impacts on Sugarcane Production in Australia. Agricultural Sciences 870–888. https://doi.org/10.4236/as.2015.68084

FAO, 2021. FAOSTAT Statistical Database.

Gers, C.J., 2003. Remotely sensed sugarcane phenological characteristics at Umfolozi South Africa, in: IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477). pp. 1010–1012 vol.2. https://doi.org/10.1109/IGARSS.2003.1293995

Gifford, R.M., 2003. Plant respiration in productivity models: conceptualisation, representation and issues for global terrestrial carbon-cycle research. Functional Plant Biology 30, 171–186.

Godwin, D.C., Jones, C.A., 1991. Nitrogen Dynamics in Soil-Plant Systems. Modeling plant and soil systems 31, 287–321.

Grassini, P., Eskridge, K.M., Cassman, K.G., 2013. Distinguishing between yield advances and yield plateaus in historical crop production trends. Nat Commun 4, 2918. https://doi.org/10.1038/ncomms3918

Gu, J., Yin, X., Zhang, C., Wang, H., Struik, P.C., 2014. Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (Oryza sativa) under drought stress. Ann Bot 114, 499–511.

Hammer, G., 2020. The roles of credibility and transdisciplinarity in modelling to support future crop improvement. In Silico Plants 2, diaa004. https://doi.org/10.1093/insilicoplants/diaa004

Hammer, G., McLean, G., Doherty, A., van Oosterom, E., Chapman, S., 2019a. Sorghum Crop Modeling and Its Utility in Agronomy and Breeding. Sorghum, Agronomy Monographs. https://doi.org/doi:10.2134/agronmonogr58.c10

Hammer, G., Messina, C., Wu, A., Cooper, M., 2019b. Biological reality and parsimony in crop models—why we need both in crop improvement! In Silico Plants 1, 1–21. https://doi.org/10.1093/insilicoplants/diz010

Hammer, G.L., Jordan, D.R., 2007. An integrated systems approach to crop improvement., in: Scale and Complexity in Plant Systems Research: Gene-Plant-Crop Relations. Springer-Verlag GmbH, Heidelberg, pp. 45–61.

Hammer, G.L., Van Oosterom, E., McLean, G., Chapman, S.C., Broad, I., Harland, P., Muchow, R.C., 2010. Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J Exp Bot 61, 2185–2202. https://doi.org/10.1093/jxb/erq095

Heslot, N., Yang, H.-P., Sorrells, M.E., Jannink, J.-L., 2012. Genomic Selection in Plant Breeding: A Comparison of Models. Crop Sci 52, 146–160. https://doi.org/https://doi.org/10.2135/cropsci2011.06.0297

Hoffman, N., 2017. Pot trial phenotyping to predict sugarcane genotype field performance with the Canegro Model. University of KwaZulu-Natal.

Hoffman, N., Singels, A., Patton, A., Ramburan, S., 2018. Predicting genotypic differences in irrigated sugarcane yield using the Canegro model and independent trait parameter estimates. European Journal of Agronomy 96, 13–21. https://doi.org/10.1016/j.eja.2018.01.005

Hoffmann, W.A., Poorter, H., 2002. Avoiding Bias in Calculations of Relative Growth Rate. Ann Bot 90, 37–42.

Hunt, R., 1982. Plant growth curves. The functional approach to plant growth analysis. Edward Arnold Ltd.

Inman-Bamber, N.G., 2004. Sugarcane water stress criteria for irrigation and drying off. Field Crops Res 89, 107–122. https://doi.org/https://doi.org/10.1016/j.fcr.2004.01.018

Inman-Bamber, N.G., 1994. Temperature and seasonal effects on canopy development and light interception of sugarcane. Field Crops Res 36, 41–51. https://doi.org/10.1016/0378-4290(94)90051-5

Inman-Bamber, N.G., 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 1862, 37–41. https://doi.org/10.1080/02571862.1991.10634587

Inman-Bamber, N.G., Attard, S.J., Verrall, S.A., Webb, W.A., Baillie, C., 2007. A web-based system for scheduling irrigation in sugarcane, in: Proceedings of the 2007 International Society of Sugar Cane Technologists Congress. International Society of Sugar Cane Technologists, pp. 459–464.

Inman-Bamber, N.G., Bonnett, G.D., Singels, A., Thorburn, P.J., 2005. Directions for R&D and cane growing from an international review on sugarcane physiology. Proceedings of the 79th Annual Congress of South African Sugar Technologists' Association, held at Kwa-Shukela, Mount Edgecombe, South Africa, 19-22 July 2005.

Inman-Bamber, N.G., Bonnett, G.D., Spillman, M.F., Hewitt, M.H., Glassop, D., 2010. Sucrose accumulation in sugarcane is influenced by temperature and genotype through the carbon sourcesink balance. Crop Pasture Sci 61, 111–121.

Inman-Bamber, N.G., Bonnett, G.D., Spillman, M.F., Hewitt, M.L., Xu, J., 2009. Sourcesink differences in genotypes and water regimes influencing sucrose accumulation in sugarcane stalks. Crop Pasture Sci 60, 316–327.

Inman-Bamber, N.G., Jackson, P.A., Stokes, C.J., Verrall, S., Lakshmanan, P., Basnayake, J., 2016. Sugarcane for water-limited environments: Enhanced capability of the APSIM sugarcane model for assessing traits for transpiration efficiency and root water supply. Field Crops Res 196, 112–123. https://doi.org/10.1016/j.fcr.2016.06.013

Inman-Bamber, N.G., Lakshmanan, P., Park, S., 2012. Sugarcane for water-limited environments: Theoretical assessment of suitable traits. Field Crops Res 134, 95–104. https://doi.org/10.1016/j.fcr.2012.05.004

Inman-Bamber, N.G., Muchow, R.C., Robertson, M.J., 2002. Dry matter partitioning of sugarcane in Australia and South Africa. Field Crops Res 76, 71–84. https://doi.org/10.1016/S0378-4290(02)00044-8

Inman-Bamber, N.G., Thompson, G.D., 1989. Models of dry matter accumulation by sugarcane. Proceedings of the South African Sugar Technologists Association, Durban 63, 212–216.

Jackson, P., Basnayake, J., Inman-Bamber, N., Lakshmanan, P., 2014. Selecting sugarcane varieties with higher transpiration efficiency, Proceedings of the 36th Conference of the Australian Society of Sugar Cane Technologists, ASSCT 2014.

Jackson, P., Deomano, E., Wei, X., 2021. An Introduction to Some Concepts in Statistical Analysis and Quantitative Genetics for Sugarcane Breeding Programs. Sugar Tech 1, 298–319. https://doi.org/10.1007/s12355-021-01071-6

Jackson, P.A., 2005. Breeding for improved sugar content in sugarcane. Field Crops Res 92, 277–290. https://doi.org/https://doi.org/10.1016/j.fcr.2005.01.024

Jones, C.A., Kiniry, J.R., Dyke, P.T., 1986. CERES-Maize: a simulation model of maize growth and development.

Jones, J.W., He, J., Boote, K.J., Wilkens, P., Porter, C.H., Hu, Z., 2011. Estimating DSSAT cropping system cultivar-specific parameters using Bayesian techniques. Methods of introducing system models into agricultural research 2, 365–393.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model, European Journal of Agronomy. https://doi.org/10.1016/S1161-0301(02)00107-7

Jones, J.W., Luyten, J.C., 1998. Simulation of biological processes, in: Agricultural Systems Modeling and Simulation. CRC Press, pp. 19–62.

Jones, M.R., 2013. Incorporating the Canegro sugarcane model into the DSSAT v4 Cropping System Model framework. University of KwaZulu-Natal, Pietermaritzburg.

Jones, M.R., Singels, A., 2018. Refining the Canegro model for improved simulation of climate change impacts on sugarcane. European Journal of Agronomy. https://doi.org/10.1016/j.eja.2017.12.009

Jones, M.R., Singels, A., 2015. Analysing yield trends in the South African sugar industry. Agric Syst 141. https://doi.org/10.1016/j.agsy.2015.09.004

Jones, M.R., Singels, A., 2014. A preliminary assessment of mid-century climate change impacts on sugarcane production in South Africa. Proceedings of the Annual Congress - South African Sugar Technologists' Association 290–297.

Jones, M.R., Singels, A., Chinorumba, S., Patton, A., Poser, C., Singh, M., Martiné, J.F.F., Christina, M., Shine, J., Annandale, J., Hammer, G., 2019. Exploring process-level genotypic and environmental effects on sugarcane yield using an international experimental dataset. Field Crops Res 244, 107622. https://doi.org/10.1016/j.fcr.2019.107622

Jones, M.R., Singels, A., Chinorumba, S., Poser, C., Christina, M., Shine, J., Annandale, J., Hammer, G.L., 2021. Evaluating process-based sugarcane models for simulating genotypic and environmental effects observed in an international dataset. Field Crops Res 260, 107983. https://doi.org/https://doi.org/10.1016/j.fcr.2020.107983

Jones, M.R., Singels, A., Inman-Bamber, N.G., 2011. Simulating Source and Sink Control of Structural Growth and Development and Sugar Accumulation in Sugarcane. Proc S Afr Sug Technol Ass 84, 157–163.

Jones, M.R., Singels, A., Ruane, A.C., 2015. Simulated impacts of climate change on water use and yield of irrigated sugarcane in South Africa. Agric Syst 139, 260–270. https://doi.org/10.1016/j.agsy.2015.07.007

Jovanovic, N.Z., Annandale, J.G., 1998. Measurement of radiant interception of crop canopies with the LAI-2000 plant canopy analyzer. South African Journal of Plant and Soil 15, 6–13. https://doi.org/10.1080/02571862.1998.10635107

Keating, B.A., Robertson, M.J., Muchow, R.C., Huth, N.I., 1999. Modelling sugarcane production systems I. Development and performance of the sugarcane module. Field Crops Res 61, 253–271. https://doi.org/10.1016/S0378-4290(98)00167-1

Kim, H., Oosterom, E. van, Dingkuhn, M., Luquet, D., Hammer, G., 2010. Regulation of tillering in sorghum: environmental effects. Ann Bot 106, 57–67. https://doi.org/10.1093/aob/mcq079

Knox, J.W., Rodríguez Díaz, J.A., Nixon, D.J., Mkhwanazi, M., 2010. A preliminary assessment of climate change impacts on sugarcane in Swaziland. Agric Syst 103, 63–72. https://doi.org/10.1016/j.agsy.2009.09.002

Lejars, C., Siegmund, B., 2004. Overview of Réunion sugar industry, in: Proc S Afr Sug Technol Ass. Citeseer, p. 78.

Lingle, S.E., 1999. Sugar Metabolism during Growth and Development in Sugarcane Internodes. Crop Sci 39, 480–486. https://doi.org/https://doi.org/10.2135/cropsci1999.0011183X0039000200030x

Liu, D.L., Bull, T.A., 2001. Simulation of biomass and sugar accumulation in sugarcane using a process-based model. Ecol Modell 144, 181–211. https://doi.org/10.1016/S0304-3800(01)00372-6

Lorenz, A.J., Chao, S., Asoro, F.G., Heffner, E.L., Hayashi, T., Iwata, H., Smith, K.P., Sorrells, M.E., Jannink, J.-L., 2011. Genomic selection in plant breeding: knowledge and prospects. Advances in agronomy 110, 77–123.

Luo, J., Que, Y., Zhang, H., Xu, L., 2013. Seasonal Variation of the Canopy Structure Parameters and Its Correlation with Yield-Related Traits in Sugarcane. The Scientific World Journal 2013, 801486. https://doi.org/10.1155/2013/801486

Marin, F.R., Jones, J.W., 2014. Process-based simple model for simulating sugarcane growth and production . Scientia Agricola .

Marin, F.R., Jones, J.W., Royce, F., Suguitani, C., Donzeli, J.L., Pallone Filho, W.J., Nassif, D.S.P., 2011. Parameterization and evaluation of predictions of DSSAT/CANEGRO for Brazilian sugarcane. Agron J 103, 304–315.

Marin, F.R., Jones, J.W., Singels, A., Royce, F., Assad, E.D., Pellegrino, G.Q., Justino, F., 2013. Climate change impacts on sugarcane attainable yield in southern Brazil. Clim Change 117, 227–239. https://doi.org/10.1007/s10584-012-0561-y

Marshall-Colon, A., Long, S.P., Allen, D.K., Allen, G., Beard, D.A., Benes, B., Von Caemmerer, S., Christensen, A.J., Cox, D.J., Hart, J.C., 2017. Crops in silico: generating virtual crops using an integrative and multi-scale modeling platform. Front Plant Sci 8, 786.

Martiné, J.-F., Todoroff, P.R.I., 2004. Le modèle de croissance mosicas et sa plateforme de simulation simulex: état des lieux et perspectives. Revue Agricole et Sucrière de l'Ile Maurice 80 LB-U1, 133–147.

McCown, R.L., Hammer, G.L., Hargreaves, J.N.G., Holzworth, D.P., Freebairn, D.M., 1996. APSIM: a novel software system for model development, model testing and simulation in agricultural systems research. Agric Syst 50, 255–271. https://doi.org/10.1016/0308-521X(94)00055-V

McGlinchey, M.G., Inman-Bamber, N.G., 2002. Robust estimates of evapotranspiration for sugarcane, in: Proc S Afr Sug Technol Ass. p. 245.

Messina, C.D., Technow, F., Tang, T., Totir, R., Gho, C., Cooper, M., 2018. Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP). European Journal of Agronomy 100, 151–162. https://doi.org/https://doi.org/10.1016/j.eja.2018.01.007

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Mohamed, H.A., Clark, J.A., Ong, C.K., 1988a. Genotypic differences in the temperature responses of tropical crops: I. Germination characteristics of groundnut (Arachis hypogaea L.) and pearl millet (Pennisetum typhoides S. & H.). J Exp Bot 39, 1121–1128.

Mohamed, H.A., Clark, J.A., Ong, C.K., 1988b. Genotypic Differences in the Temperature Responses of Tropical Crops: II. SEEDLING EMERGENCE AND LEAF GROWTH OF GROUNDNUT (ARACHIS HYPOGAEA L.) AND PEARL MILLET (PENNISETUM TYPHOIDES S. & H.). J Exp Bot 39, 1129–1135.

Moor, G.M., Wynne, A.T., 2001. Economic maximisation of grower and miller sugar cane profits: optimising the length of milling season at South African sugar factories., in: International Society of Sugar Cane Technologists. Proceedings of the XXIV Congress, Brisbane, Australia, 17-21 September 2001. Volume 1. Australian Society of Sugar Cane Technologists, pp. 245–249.

Moore, P.H., Paterson, A.H., Tew, T., 2013. Sugarcane: the crop, the plant, and domestication. Sugarcane: physiology, biochemistry, and functional biology 1–17.

Muchow, R.C., Spillman, M.F., Wood, A.W., Thomas, M.R., 1994. Radiation interception and biomass accumulation in a sugarcane crop grown under irrigated

tropical conditions. Aust J Agric Res 45, 37–49. https://doi.org/10.1071/AR9940037

Ngobese, I., Ramburan, S., Labuschagne, M., 2018. Quantifying sugarcane cultivar differences in tiller and stalk phenology: identifying traits suited to crop model-assisted breeding. J Crop Improv 32, 847–860. https://doi.org/10.1080/15427528.2018.1534762

O'Leary, G.J., 2000. A review of three sugarcane simulation models with respect to their prediction of sucrose yield. Field Crops Res 68, 97–111. https://doi.org/10.1016/S0378-4290(00)00112-X

Oliveira, F.A.A., Jones, J.W., Pavan, W., Bhakta, M., Vallejos, C.E., Correll, M.J., Boote, K.J., Fernandes, J.M.C., Hölbig, C.A., Hoogenboom, G., 2021. Incorporating a dynamic gene-based process module into a crop simulation model. In Silico Plants 3, diab011. https://doi.org/10.1093/insilicoplants/diab011

Ortiz, A.C., de Smet, I., Sozzani, R., Locke, A.M., 2022. Field-grown soybean shows genotypic variation in physiological and seed composition responses to heat stress during seed development. Environ Exp Bot 195, 104768. https://doi.org/https://doi.org/10.1016/j.envexpbot.2021.104768

Pagani, V., Stella, T., Guarneri, T., Finotto, G., Berg, M. van den, Marin, F.R., Acutis, M., Confalonieri, R., 2017. Forecasting sugarcane yields using agro-climatic indicators and Canegro model: a case study in the main production region in Brazil. Agric Syst 154, 45–52. https://doi.org/10.1016/j.agsy.2017.03.002

Parent, B., Tardieu, F., 2014. Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? J Exp Bot 65, 6179–6189. https://doi.org/10.1093/jxb/eru223

Parent, B., Tardieu, F., 2012. Temperature responses of developmental processes have not been affected by breeding in different ecological areas for 17 crop species. New Phytologist 194, 760–774. https://doi.org/10.1111/j.1469-8137.2012.04086.x

Parent, B., Turc, O., Gibon, Y., Stitt, M., Tardieu, F., 2010. Modelling temperature-compensated physiological rates, based on the co-ordination of responses to temperature of developmental processes. J Exp Bot 61, 2057–2069. https://doi.org/10.1093/jxb/erq003

Park, S.E., Robertson, M., Inman-Bamber, N.G., 2005. Decline in the growth of a sugarcane crop with age under high input conditions. Field Crops Res 92, 305–320. https://doi.org/10.1016/j.fcr.2005.01.025

Pastina, M.M., Malosetti, M., Gazaffi, R., Mollinari, M., Margarido, G.R.A., Oliveira, K.M., Pinto, L.R., Souza, A.P., van Eeuwijk, F.A., Garcia, A.A.F., 2012. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. Theoretical and Applied Genetics 124, 835–849. https://doi.org/10.1007/s00122-011-1748-8

Peng, S., Khush, G.S., Virk, P., Tang, Q., Zou, Y., 2008. Progress in ideotype breeding to increase rice yield potential. Field Crops Res 108, 32–38. https://doi.org/https://doi.org/10.1016/j.fcr.2008.04.001

Poser, C., Barau, L., Mezino, M., 2019. Effect of the emergence threshold temperature on the geographical distribution of the variety R583 in Reunion. International

Society of Sugar Cane Technologists. Proceedings of the XXX congress, Tucuman, Argentina.

R Core Team, 2016. R: A Language and Environment for Statistical Computing.

Ramburan, S., 2012. The nature and causes of sugarcane genotype x environment interactions: Integrated approaches to analysis and interpretation. University of the Free State.

Ramburan, S., 2011. Sugarcane cultivar x time of harvest interactions in South Africa. South African Journal of Plant and Soil 28, 75–84. https://doi.org/10.1080/02571862.2011.10640016

Richards, R.A., 2000. Selectable traits to increase crop photosynthesis and yield of grain crops. J Exp Bot 51, 447–458. https://doi.org/10.1093/jexbot/51.suppl_1.447

Rizzo, G., Monzon, J.P., Tenorio, F.A., Howard, R., Cassman, K.G., Grassini, P., 2022. Climate and agronomy, not genetics, underpin recent maize yield gains in favorable environments. Proceedings of the National Academy of Sciences 119, e2113629119. https://doi.org/10.1073/pnas.2113629119

Robertson, M.J., Bonnett, G.D., Hughes, R.M., Muchow, R.C., Campbell, J.A., 1998. Temperature and leaf area expansion of sugarcane: integration of controlled-environment, field and model studies. Functional Plant Biology 25, 819–828.

Robertson, M.J., Donaldson, R.A., 1998. Changes in the components of cane and sucrose yield in response to drying-off of sugarcane before harvest. Field Crops Res 55, 201–208. https://doi.org/https://doi.org/10.1016/S0378-4290(97)00065-8

Robertson, M.J., Wood, A.W., Muchow, R.C., 1996. Growth of sugarcane under high input conditions in tropical Australia. I. Radiation use, biomass accumulation and partitioning. Field Crops Res 48, 11–25. https://doi.org/https://doi.org/10.1016/0378-4290(96)00041-X

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Mutter, C.Z., Adiku, S.G.K., Ahmad, A., Beletse, Y., Gangwar, B., Guntuku, D., Kihara, J., Masikati, P., Paramasivan, P., Rao, K.P.C., Zubair, L., 2012. The Agricultural Model Intercomparison and Improvement Project (AgMIP): Integrated regional assessment projects, in: Hillel, D., Rosenzweig, C. (Eds.), Handbook of Climate Change and Agroecosystems: Global and Regional Aspects and Implications, ICP Series on Climate Change Impacts, Adaptation, and Mitigation Vol. 2. Imperial College Press, London, pp. 263–280.

Rotili, D.H., de Voil, P., Eyre, J., Serafin, L., Aisthorpe, D., Maddonni, G.Á., Rodríguez, D., 2020. Untangling genotype x management interactions in multi-environment on-farm experimentation. Field Crops Res 255, 107900. https://doi.org/https://doi.org/10.1016/j.fcr.2020.107900

Ruane, A.C., Goldberg, R., Chryssanthacopoulos, J., 2015. Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. Agric For Meteorol 200, 233–248. https://doi.org/https://doi.org/10.1016/j.agrformet.2014.09.016

Salmerón, M., Purcell, L.C., Vories, E.D., Shannon, G., 2017. Simulation of genotype-by-environment interactions on irrigated soybean yields in the U.S. Midsouth. Agric Syst 150, 120–129. https://doi.org/10.1016/j.agsy.2016.10.008

SASRI, 2006. N31 Information Sheet [WWW Document]. URL https://sasri.org.za/wp-content/uploads/Information_Sheets/IS_13.18-Variety-N31.pdf (accessed 11.15.21).

Sebastião de Oliveira Maia Júnior, J.R. de A., Claudiana Moura dos Santos Karolyne P. O. d. Santos, J.A.C.S., Silva, J.V., Endres, L., 2019. Leaf thickness and gas exchange are indicators of drought stress tolerance of sugarcane. Emir J Food Agric 31. https://doi.org/https://doi.org/10.9755/ejfa.2019.v31.i1.1897

Sexton, J., Everingham, Y.L., Inman-Bamber, G., 2017. A global sensitivity analysis of cultivar trait parameters in a sugarcane growth model for contrasting production environments in Queensland, Australia. European Journal of Agronomy 88, 96–105. https://doi.org/10.1016/j.eja.2015.11.009

Sinclair, T.R., 2012. Is transpiration efficiency a viable plant trait in breeding for crop improvement? Functional Plant Biology 39, 359–365.

Sinclair, T.R., Gilbert, R.A., Perdomo, R.E., Shine, J.M., Powell, G., Montes, G., 2004. Sugarcane leaf area development under field conditions in Florida, USA. Field Crops Res 88, 171–178. https://doi.org/10.1016/j.fcr.2003.12.005

Singels, A., 2017. ICSM report on Trait modelling workshop held at Mount Edgecombe from 26-27 June 2017.

Singels, A., 2014. Crop Models, in: P.H. Moore and F.C. Botha (Ed.), Sugarcane: Physiology, Biochemistry, and Functional Biology. Wiley Online Books. https://doi.org/doi:10.1002/9781118771280.ch20

Singels, A., 2007. A new approach to implementing computer-based decision support for sugarcane farmers and extension staff: the case of My Canesim., in: XXVI Congress, International Society of Sugar Cane Technologists, ICC, Durban, South Africa, 29 July-2 August, 2007. International Society Sugar Cane Technologists (ISSCT), pp. 211–214.

Singels, A., Berg, M. van den, Antwerpen, R. van, 2005a. Climate change and yield decline: an analysis of actual and simulated yield data from the BT1 field experiment. Proceedings of the 79th Annual Congress of South African Sugar Technologists' Association, held at Kwa-Shukela, Mount Edgecombe, South Africa, 19-22 July 2005.

Singels, A., Bezuidenhout, C.N., 2002. A new method of simulating dry matter partitioning in the Canegro sugarcane model. Field Crops Res 78, 151–164. https://doi.org/10.1016/S0378-4290(02)00118-1

Singels, A., Donaldson, R.A., 2000. A simple model of unstressed sugarcane canopy development. Proceedings of the South African Sugar Technologists' Association 74, 151–154.

Singels, A., Donaldson, R.A., Smit, M.A., 2005b. Improving biomass production and partitioning in sugarcane: Theory and practice. Field Crops Res 92, 291–303. https://doi.org/10.1016/j.fcr.2005.01.022

Singels, A., Hoffman, N., Paraskevepoulos, A.L., Ramburan, S., 2016. Sugarcane genetic trait parameter estimation., in: Proceedings of the ICROPM2016 International Crop Modelling Symposium Held from 15 to 17 March 2016 in Berlin. pp. 143–144.

Singels, A., Inman-Bamber, N.G., 2011. Modelling genetic and environmental control of biomass partitioning at plant and phytomer level of sugarcane grown in controlled environments. Crop & Pasture SciencePasture Science 62, 66–81. https://doi.org/10.1071/CP10182

Singels, A., Jackson, P., Inman-Bamber, G., 2021. Chapter 21 - Sugarcane, in: Sadras, V.O., Calderini, D.F.B.T.-C.P.C.H. for M.C. (Eds.), . Academic Press, pp. 674–713. https://doi.org/https://doi.org/10.1016/B978-0-12-819194-1.00021-9

Singels, A., Jones, M., Marin, F., Ruane, A., Thorburn, P., 2014. Predicting Climate Change Impacts on Sugarcane Production at Sites in Australia, Brazil and South Africa Using the Canegro Model. Sugar Tech 16, 347–355. https://doi.org/10.1007/s12355-013-0274-1

Singels, A., Jones, M., Marin, F., Ruane, A.C., Thorburn, P., 2013. Predicting climate change impacts on sugarcane production at sites in Australia, Brazil and South Africa using the Canegro model. International Sugar Journal 115, 874–881.

Singels, A., Jones, M., van Den Berg, M., 2008. DSSAT v4. 5-Canegro Sugarcane Plant Module: Scientific Documentation. South African Sugarcane Research Institute.

Singels, A., Kennedy, A.J., Bezuidenhout, C.N., 1998. Irricane: a simple computerised irrigation scheduling method for sugarcane. Proceedings of the Annual Congress - South African Sugar Technologists' Association 117–122.

Singels, A., Paraskevepoulos, A.L.L., 2017. The Canesim Sugarcane Model: Scientific Documentation. South African Sugarcane Research Institute.

Singels, A., Smit, M.A., 2009. Sugarcane response to row spacing-induced competition for light. Field Crops Res 113, 149–155. https://doi.org/10.1016/j.fcr.2009.04.015

Singels, A., Smit, M.A., Butterfield, M.K., Van Heerden, P.D.R., Van Den Berg, M., 2010a. Identifying quantitative trait alleles for physiological traits in sugarcane: an exploratory study, in: Proc. Int. Soc. Sugar Cane Technol.

Singels, A., Smit, M.A., Redshaw, K.A., Donaldson, R.A., 2005c. The effect of crop start date, crop class and cultivar on sugarcane canopy development and radiation interception. Field Crops Res 92, 249–260. https://doi.org/10.1016/j.fcr.2005.01.028

Singels, A., van den Berg, M., Smit, M.A., Jones, M.R., van Antwerpen, R., 2010b. Modelling water uptake, growth and sucrose accumulation of sugarcane subjected to water stress. Field Crops Res 117, 59–69. https://doi.org/10.1016/j.fcr.2010.02.003

Smit, M.A., 2010. Characterising the factors that affect germination and emergence in sugarcane. South African Sugarcane Technologists Association Annual Congress 83, 230–234.

Smit, M.A., Singels, A., 2007. Quantifying the effects of environment and genotype on stalk elongation rate in sugarcane., in: Proceedings of the International Society of Sugar Cane Technologists. pp. 568–572.

Smit, M.A., Singels, A., 2006. The response of sugarcane canopy development to water stress. Field Crops Res 98, 91–97. https://doi.org/10.1016/j.fcr.2005.12.009

Som-ard, J., Atzberger, C., Izquierdo-Verdiguier, E., Vuolo, F., Immitzer, M., 2021. Remote sensing applications in sugarcane cultivation: A review. Remote Sens (Basel) 13, 4040.

Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. Agron J 101, 426–437. https://doi.org/10.2134/agronj2008.0139s

Stöckle, C.O., Kemanian, A.R., 2020. Can Crop Models Identify Critical Gaps in Genetics, Environment, and Management Interactions? Front Plant Sci.

Tao, F., Rötter, R.P., Palosuo, T., Díaz-Ambrona, C.G.H., Mínguez, M.I., Semenov, M.A., Kersebaum, K.C., Nendel, C., Cammarano, D., Hoffmann, H., Ewert, F., Dambreville, A., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Höhn, J.G., Salo, T., Ferrise, R., Bindi, M., Schulman, A.H., 2017. Designing future barley ideotypes using a crop model ensemble. European Journal of Agronomy 82, 144–162. https://doi.org/10.1016/j.eja.2016.10.012

Tardieu, F., 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8, 9–14. https://doi.org/https://doi.org/10.1016/S1360-1385(02)00008-0

Tardieu, F., Granato, I., Oosterom, E., Parent, B., Hammer, G., 2020. Are crop and detailed physiological models equally “mechanistic” for predicting the genetic variability of whole plant behaviour? - the nexus between mechanisms and adaptive strategies. In Silico Plants. https://doi.org/10.1093/insilicoplants/diaa011

Tardieu, F., Granier, C., Muller, B., 1999. Modelling leaf expansion in a fluctuating environment: are changes in specific leaf area a consequence of changes in expansion rate? New Phytol 143, 33–43.

Technow, F., Messina, C.D., Totir, L.R., Cooper, M., 2015. Integrating Crop Growth Models with Whole Genome Prediction through Approximate Bayesian Computation. PLoS One 10, e0130855.

Terauchi, T., Matsuoka, M., 2001. Analysis of the slow growth of sugarcane at the early stage. Proc. Int. Soc. Sugar Cane Technol. 24, 149–151.

Tirfessa, A., McLean, G., Mace, E., van Oosterom, E., Jordan, D., Hammer, G., 2020. Differences in temperature response of phenological development among diverse Ethiopian sorghum genotypes are linked to racial grouping and agroecological adaptation. Crop Sci 60, 977–990.

Tsutsumi-Morita, Y., Heuvelink, E., Khaleghi, S., Bustos-Korts, D., Marcelis, L.F.M., Vermeer, K.M.C.A., van Dijk, H., Millenaar, F.F., Van Voorn, G.A.K., Van Eeuwijk, F.A., 2021. Yield dissection models to improve yield: a case study in tomato. In Silico Plants 3, diab012. https://doi.org/10.1093/insilicoplants/diab012

Van den Berg, M., Singels, A., 2013. Modelling and monitoring for strategic yield gap diagnosis in the South African sugar belt. Field Crops Res 143, 143–150. https://doi.org/10.1016/j.fcr.2012.10.009

van Dillewijn, C., 1952. Botany of Sugarcane. Chronica Botanica; New York: Stechert-Hafner.

Van Heerden, P.D.R., 2014. Differential acclimation capacity to frost in sugarcane varieties grown under field conditions. Plant Growth Regul 72, 181–187. https://doi.org/10.1007/s10725-013-9850-3

Van Heerden, P.D.R., Donaldson, R.A., Watt, D.A., Singels, A., 2010. Biomass accumulation in sugarcane: Unravelling the factors underpinning reduced growth phenomena. J Exp Bot 61, 2877–2887. https://doi.org/10.1093/jxb/erq144

van Heerden, P.D.R., Eggleston, G., Donaldson, R.A., 2013. Ripening and Postharvest Deterioration. Sugarcane: Physiology, Biochemistry, and Functional Biology, Wiley Online Books. https://doi.org/https://doi.org/10.1002/9781118771280.ch4

Van Heerden, P.D.R., Mbatha, T.P., Ngxaliwe, S., 2015a. Chemical ripening of sugarcane with trinexapac-ethyl (Moddus®)—Mode of action and comparative efficacy. Field Crops Res 181, 69–75.

Van Heerden, P.D.R., Singels, A., Paraskevopoulos, A., Rossler, R., 2015b. Negative effects of lodging on irrigated sugarcane productivity - an experimental and crop modelling assessment. Field Crops Res 180, 135–142. https://doi.org/10.1016/j.fcr.2015.05.019

Van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance-A review. Field Crops Res 143, 4–17. https://doi.org/10.1016/j.fcr.2012.09.009

Vanuytrecht, E., Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., Heng, L.K., Garcia Vila, M., Mejias Moreno, P., 2014. AquaCrop: FAO's crop water productivity and yield response model. Environmental Modelling & Software 62, 351–360. https://doi.org/https://doi.org/10.1016/j.envsoft.2014.08.005

Venkataramana, S., Shunmugasundaram, S., Naidu, K.M., 1984. Growth behaviour of field grown sugarcane varieties in relation to environmental parameters and soil moisture stress. Agric For Meteorol 31, 251–260. https://doi.org/https://doi.org/10.1016/0168-1923(84)90039-X

Villegas, F.D., Daza, O.H., Jones, J.W., Royce, F.S., 2005. CASUPRO: An industry-driven sugarcane model, in: 2005 ASAE Annual Meeting. American Society of Agricultural and Biological Engineers, p. 1.

von Caemmerer, S., 2021. Updating the steady-state model of C4 photosynthesis. J Exp Bot 72, 6003–6017. https://doi.org/10.1093/jxb/erab266

Waclawovsky, A.J., Sato, P.M., Lembke, C.G., Moore, P.H., Souza, G.M., 2010. Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. Plant Biotechnol J 8, 263–276. https://doi.org/https://doi.org/10.1111/j.1467-7652.2009.00491.x

Wang, E., Brown, H.E., Rebetzke, G.J., Zhao, Z., Zheng, B., Chapman, S.C., 2019. Improving process-based crop models to better capture genotype×environment×management interactions. J Exp Bot 70, 2389–2401. https://doi.org/10.1093/jxb/erz092

White, J.W., Hoogenboom, G., 2003. Gene-based approaches to crop simulation: Past experiences and future opportunities. Agron J 95, 52–64.

White, J.W., Hoogenboom, G., 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. Agron J 88, 416–422.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., Prasanna, B.M., Olsen, M.S., 2017. Enhancing genetic gain in the era of molecular breeding. J Exp Bot 68, 2641–2666. https://doi.org/10.1093/jxb/erx135

Yadav, R., Gupta, S., Gaikwad, K.B., Bainsla, N.K., Kumar, M., Babu, P., Ansari, R., Dhar, N., Dharmateja, P., Prasad, R., 2021. Genetic Gain in Yield and Associated Changes in Agronomic Traits in Wheat Cultivars Developed Between 1900 and 2016 for Irrigated Ecosystems of Northwestern Plain Zone of India . Frontiers in Plant Science .

Yadav, S., Jackson, P., Wei, X., Ross, E.M., Aitken, K., Deomano, E., Atkin, F., Hayes, B.J., Voss-Fels, K.P., 2020. Accelerating Genetic Gain in Sugarcane Breeding Using Genomic Selection. Agronomy. https://doi.org/10.3390/agronomy10040585

Zhao, D., Irey, M., LaBorde, C., Hu, C.-J., 2019. Physiological and Yield Characteristics of 18 Sugarcane Genotypes Grown on a Sand Soil. Crop Sci 59, 2741–2751. https://doi.org/https://doi.org/10.2135/cropsci2019.02.0107

Zhao, P., Jackson, P.A., Basnayake, J., Liu, J., Chen, X., Zhao, J., Zhao, X., Bai, Y., Yang, L., Zan, F., Yang, K., Xia, H., Qin, W., Zhao, L., Yao, L., Lakshmanan, P., Fan, Y., 2017. Genetic variation in sugarcane for leaf functional traits and relationships with cane yield, in environments with varying water stress. Field Crops Res 213, 143–153. https://doi.org/https://doi.org/10.1016/j.fcr.2017.08.004

Zheng, B., Biddulph, B., Li, D., Kuchel, H., Chapman, S., 2013. Quantification of the effects of VRN1 and Ppd-D1 to predict spring wheat (Triticum aestivum) heading time across diverse environments. J Exp Bot 64, 3747–3761.

Zhi, X., Tao, Y., Jordan, D., Borrell, A., Hunt, C., Cruickshank, A., Potgieter, A., Wu, A., Hammer, G., George-Jaeggli, B., Mace, E., 2021. Genetic control of leaf angle in sorghum and its effect on light interception. J Exp Bot. https://doi.org/10.1093/jxb/erab467

Zhou, H., Zhou, G., He, Q., Zhou, L., Ji, Y., Zhou, M., 2020. Environmental explanation of maize specific leaf area under varying water stress regimes. Environ Exp Bot 171, 103932. https://doi.org/https://doi.org/10.1016/j.envexpbot.2019.103932

Zhou, M., 2017. Cultivar genetic gains from 50 years of irrigated sugarcane breeding in South Africa. South African Journal of Plant and Soil 34, 167–174.

Zhou, M., 2013. Conventional Sugarcane Breeding in South Africa: Progress and Future Prospects. Am J Plant Sci 04, 189–197. https://doi.org/10.4236/ajps.2013.42025

Zhou, M., 2005. Strategies used for variety selection in the breeding programme at the Zimbabwe Sugar Association Experiment Station. Proceedings of the 78th Annual Congress of South African Sugar Technologists' Association, held at Kwa-Shukela, Mount Edgecombe, South Africa, 27-30 July 2004.

Zhou, M., Gwata, T., 2015. Quantifying Sugarcane Cultivar Genetic Gains in the Midlands Region of South Africa. Agron J 108. https://doi.org/10.2134/agronj2015.0141

Zhou, M.M., 2013. Realised selection gains for cane yield, sucrose content and sugar yield among South African breeding programmes, in: Proceedings of the South African Sugar Technologists' Association. pp. 273–285.

Zhou, M.M., Singels, A., Savage, M.J., 2003. Physiological parameters for modelling differences in canopy development between sugarcane cultivars. Proceedings of the Annual Congress - South African Sugar Technologists' Association 610–621.

10. APPENDIX A – INVESTIGATING THE USE OF SIMULATED SOIL TEMPERATURE IN DRIVING GERMINATION

10.1 Methodology

Thermal time from crop start to observed date of 50% primary shoot emergence (*TTem_P* for plant crops, *TTem_R* for ratoon crops) was calculated using a range of base temperatures (0-20 °C, 0.5 °C increments) calculated using air and soil temperatures simulated by the DSSAT-Canegro model, for each G-E combination for the ICSM IGEP dataset (Table 5-3). The coefficient of variation (cv%) was calculated per G (across Es), for each base temperature and temperature source tested. The hypothesis in this regard was that cv% would be lower for soil temperatures than air temperatures, and it would be possible to identify optimal base temperature (Tb_germ, °C) and *TTem_P* / *TTem_R* parameter values for each G.

10.2 Results

Per-G minima of cv% of thermal time from crop start to date of primary shoot emergence, calculated using air and DSSAT-simulated soil temperatures (Figure 10-1), revealed the base temperature that minimised *TTem_P* and *TTem_R* between Es.

Calculated using air temperatures, *TTem_P* N41 plant crops had a clear minimum at 14.5 °C; variability with R570 was high (≈ 35%) with similar cv% values for *Tb_germ* between 10 and 15 °C; and CP88-1762 Tb_germ was minimised in the range 10-15 °C as well, although the variability between Es was lower than with R570 (about 15%). Ratoon crops showed much clearer minima than plant crops, at higher base temperatures than plant crops.

Using soil temperatures, however, plant crops showed clearer *TTem_P* cv% minima than with air temperatures, although the cv% values were higher for N41, similar for CP88-1762 and lower for R570. For ratoon crops, only R570 showed a minimum value (around 22 °C), while for N41 and R570 cv% increased gradually from 0 °C, with a sharp increase from 23 °C upwards. Minimum cv% values were higher than with air temperatures, strongly indicating that ratoon crop germination ought to be simulated with air rather than soil temperatures.

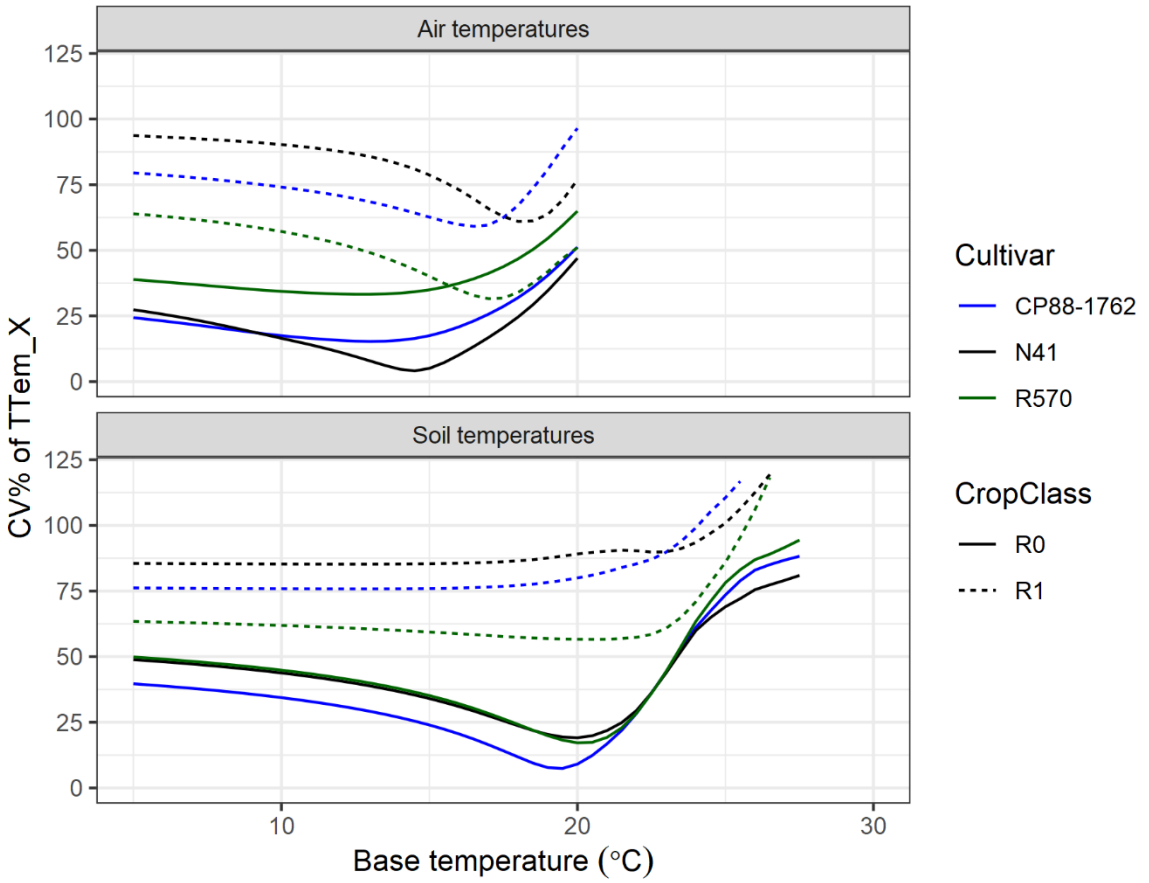

Figure 10-1. Coefficient of variation (%) of thermal time from crop start to emergence (TTem_X) for ICSM cultivars and sites (Belle Glade, Chiredzi, La Mare and Pongola), at different base temperatures. Soil temperatures were simulated by DSSAT-Canegro, while air temperatures were recorded.

This analysis was not repeated using soil temperatures simulated by APSIM-Sugar, but DSSAT and APSIM simulations were compared to an observed multi-year sugarcane soil temperature dataset. Model performance for predicting soil temperatures was found to be similar for both crop models.

10.3 Conclusion

As using soil temperatures to drive germination was not clearly beneficial overall, the decision was made to use air temperatures, as this is simpler to implement and is the standard way that duration of the germination phase is calculated in most sugarcane models.

11. APPENDIX B – SUPPLEMENTARY DATA

11.1 Supplementary CaneGEM calibration results

Results for the model assessment scenario code “GEM_376C_S” – i.e. the CaneGEM model, with the Nco376 calibration dataset, with duration of the germination phase simulated – are shown in Figure 11-1, Figure 11-2 and Table 11-1. These are supplementary to the results presented in Section 5.4.

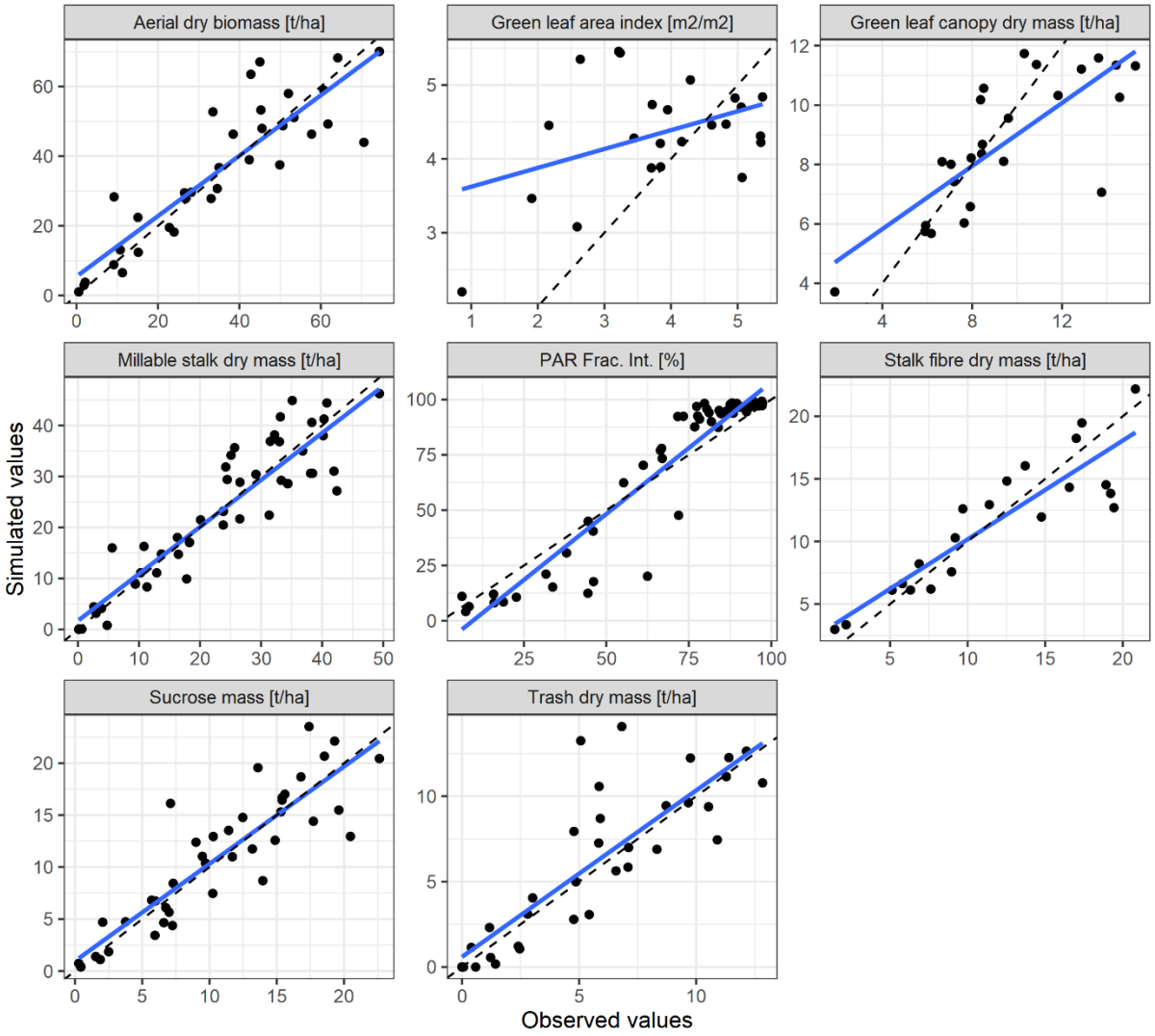

Figure 11-1. Scatter plot of time-series data for crop phenology- and biomass-related variables, for the CaneGEM model, calibration dataset for NCo376. Durations of germination phases were simulated.

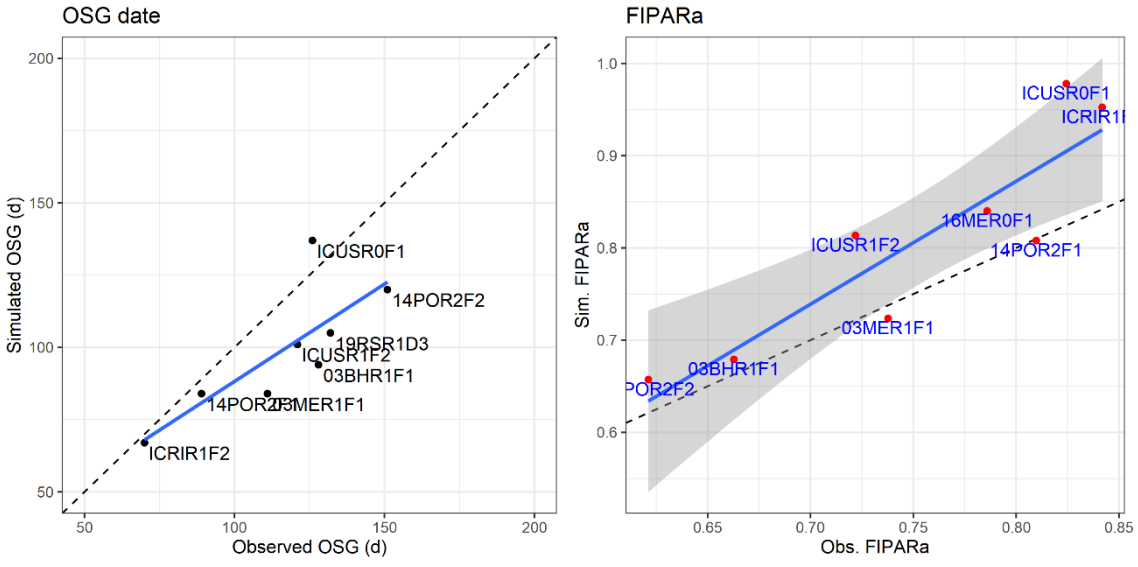

Figure 11-2. Scatter plots of age of onset of stalk growth (days since crop start) and seasonal PAR interception, simulated (CaneGEM model) and observed, for the calibration dataset for NCo376. Durations of germination phases were simulated.

Table 11-1. CaneGEM model performance statistics, for calibration with NCo376 data and simulated durations of growth phase. N = "number of observations" , "p" = p-value, "Y-int" is the y intercept of linear regression between observed and simulated values, RMSE is the root mean squared error, and APE is the average prediction error (explained in the text).

Variable	N	R^2	p	Y-int	Slope	RMSE	APE
Time-series values							
Aerial dry biomass [t/ha]	34	0.78	0.00	5.48	0.87	9.73	0.81
Green leaf area index [m2/m2]	23	0.17	0.05	3.37	0.26	1.23	0.51
Green leaf canopy dry mass [t/ha]	25	0.60	0.00	3.70	0.53	2.19	-0.70
Millable stalk dry mass [t/ha]	51	0.85	0.00	1.71	0.92	5.36	-0.12
PAR Frac. Int. [%]	59	0.90	0.00	-11.31	1.19	12.36	2.27
Stalk fibre dry mass [t/ha]	21	0.79	0.00	2.28	0.79	2.66	-0.18
Sucrose mass [t/ha]	40	0.78	0.00	0.96	0.94	3.05	0.27
Trash dry mass [t/ha]	34	0.71	0.00	0.59	0.98	2.48	0.45
Trash dry mass [t/ha]	34	0.71	0.00	0.59	0.98	2.48	0.45
Time-series biomass fractions							
"Trash" / ADM	33	0.65	0.00	-0.02	1.13	0.04	0.00
Green leaf canopy dry mass / ADM	25	0.82	0.00	0.00	1.06	0.10	0.01
Stalk dry mass / ADM	39	0.67	0.00	0.02	0.94	0.11	-0.01
Sucrose / stalk dry mass	43	0.70	0.00	0.00	1.00	0.08	0.00
Values at harvest							
Seasonal PAR interception fraction	8	0.80	0.00	-0.19	1.33	0.08	0.06
Onset of stalk growth (d)	8	0.61	0.02	20.84	0.67	22.72	-17.00
Aerial dry biomass [t/ha], at harvest	9	0.22	0.21	33.13	0.43	12.06	2.90
Green leaf area index [m2/m2], at harvest	6	0.27	0.29	6.29	-0.43	1.47	0.89
Green leaf canopy dry mass [t/ha], at harvest	6	0.47	0.13	3.36	0.49	2.99	-1.96
Millable stalk dry mass [t/ha], at harvest	14	0.34	0.03	18.61	0.52	6.86	1.87
PAR Frac. Int. [%], at harvest	8	0.00	0.88	97.07	0.01	6.40	5.43
Stalk fibre dry mass [t/ha], at harvest	5	0.65	0.10	-0.99	0.97	3.36	-1.54
Sucrose mass [t/ha], at harvest	13	0.34	0.03	9.58	0.50	4.17	1.90
Trash dry mass [t/ha], at harvest	9	0.01	0.85	11.87	-0.06	4.33	2.58
Biomass fractions at harvest							
"Trash" / ADM	9	0.03	0.67	0.23	-0.19	0.06	0.04
Green leaf canopy dry mass / ADM	6	0.01	0.89	0.15	0.02	0.08	-0.04
Stalk dry mass / ADM	10	0.27	0.12	0.54	0.15	0.08	0.01
Sucrose / stalk dry mass	14	0.13	0.21	0.24	0.52	0.09	0.03

11.2 Supplementary CaneGEM validation results

Model performance outcomes for the CaneGEM model, the NCo376 validation dataset, with simulated durations of germination phase, are shown in Figure 11-3, Figure 11-4 and Figure 11-5.

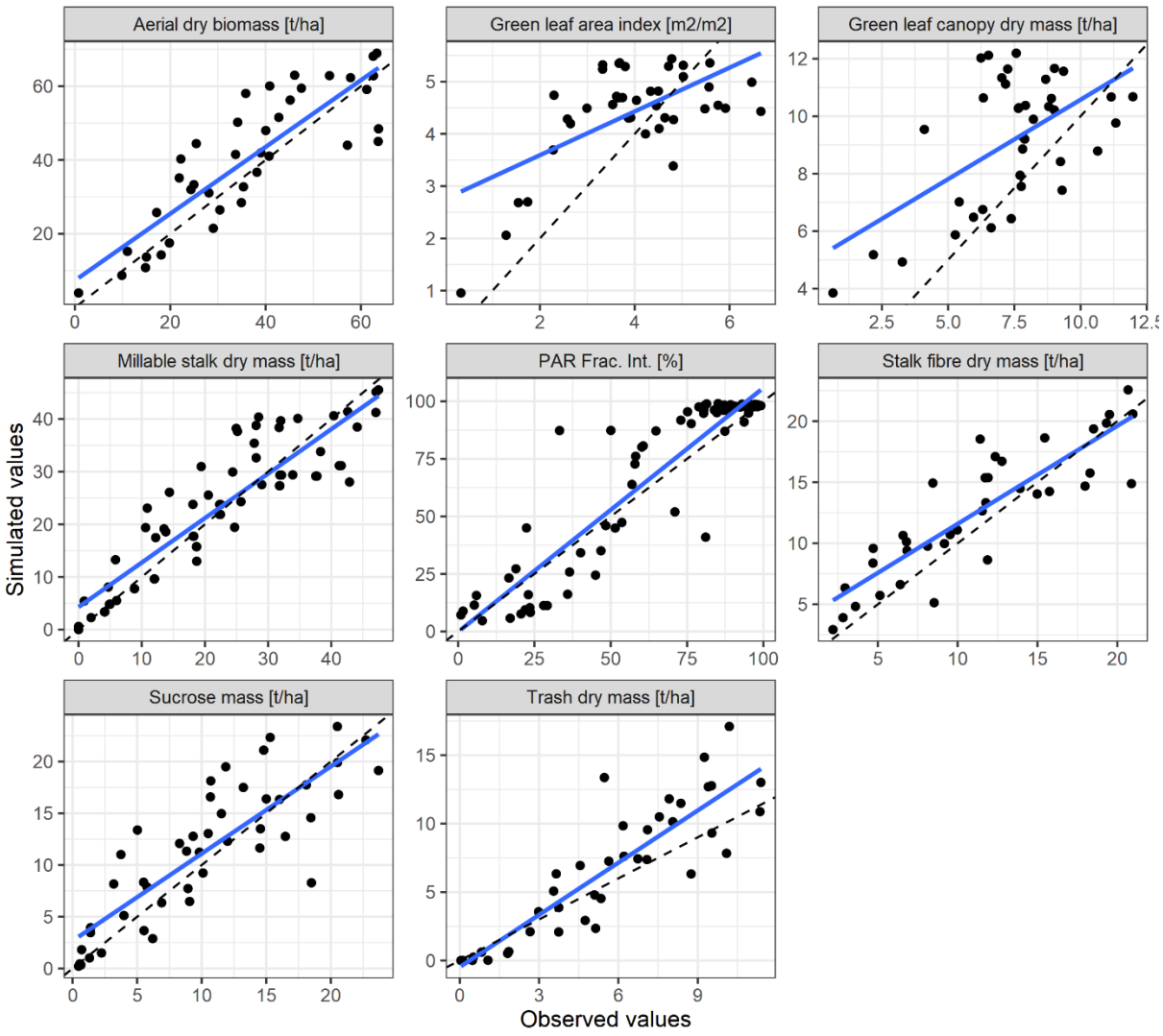

Figure 11-3. Scatter plot of time-series data for crop phenology- and biomass-related variables, for the CaneGEM model and the validation dataset for NCo376, simulated durations of germination phase (model assessment scenario code "GEM_376V_S").

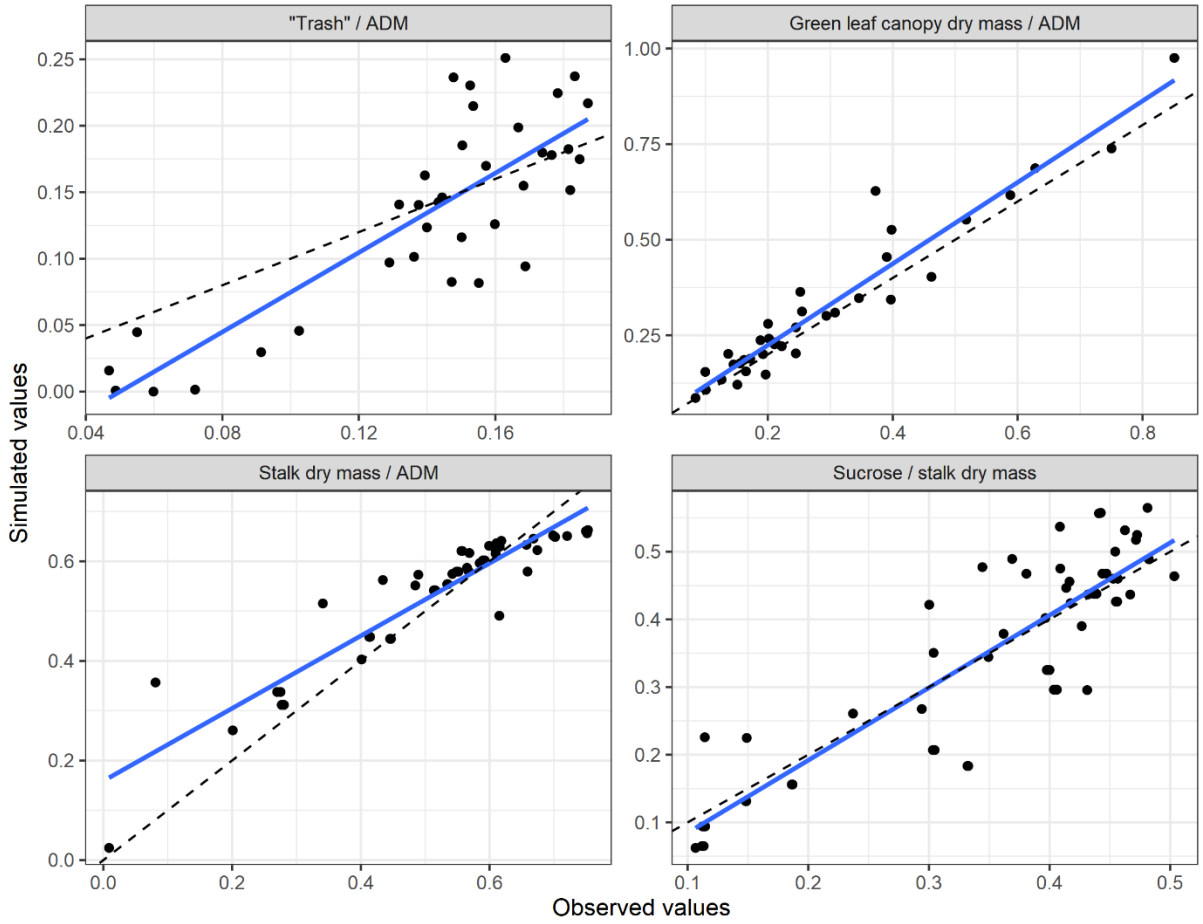

Figure 11-4. Scatter plot of time-series data, showing simulated (CaneGEM model) and observed biomass fractions, for the validation dataset for NCo376, simulated duration of germination phase (model assessment scenario code "GEM_376V_S"). "ADM" is above-ground dry biomass (t/ha), while "Trash" refers to senesced leaf and stalk dry mass.

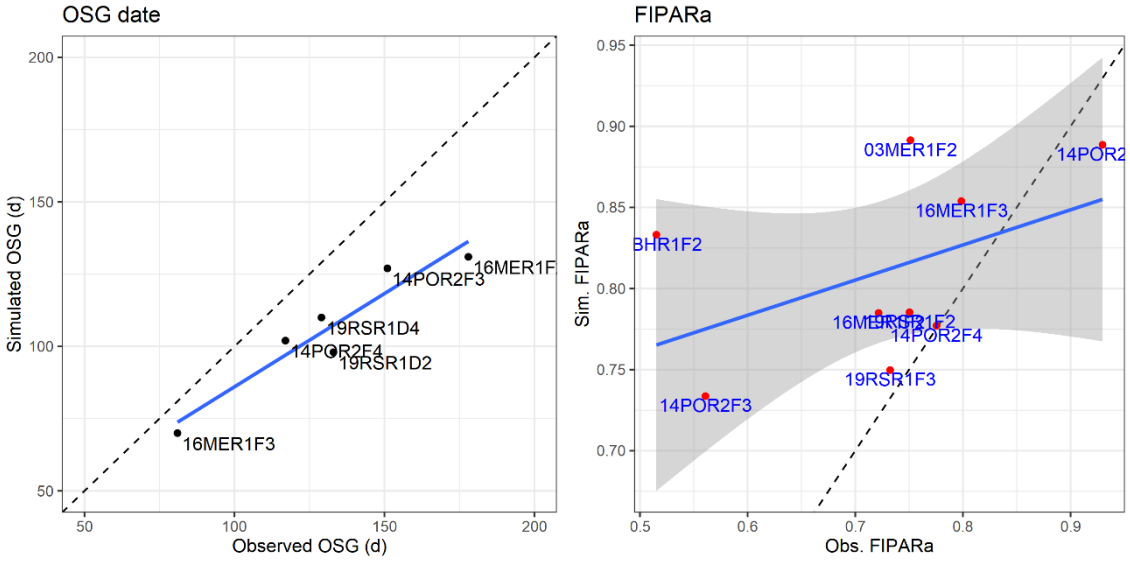

Figure 11-5. Simulated vs observed duration of tillering phase (labelled as date of onset of stalk growth, "OSG date", left), and seasonal PAR interception (FIPARa, right), for the NCo376 validation dataset, simulated duration of germination phase

(model assessment scenario code "GEM_376V_S"). The shaded region shows values within one standard error range.

11.3 Supplementary CaneGEM results for the ICSM IGEP dataset

CaneGEM-simulated ("GEM_IGEP_P" scenario) and observed FIPAR are shown in Figure 11-6. Additional results for scenario "GEM_IGEP_S" are shown in Figure 11-7 and Figure 11-8.

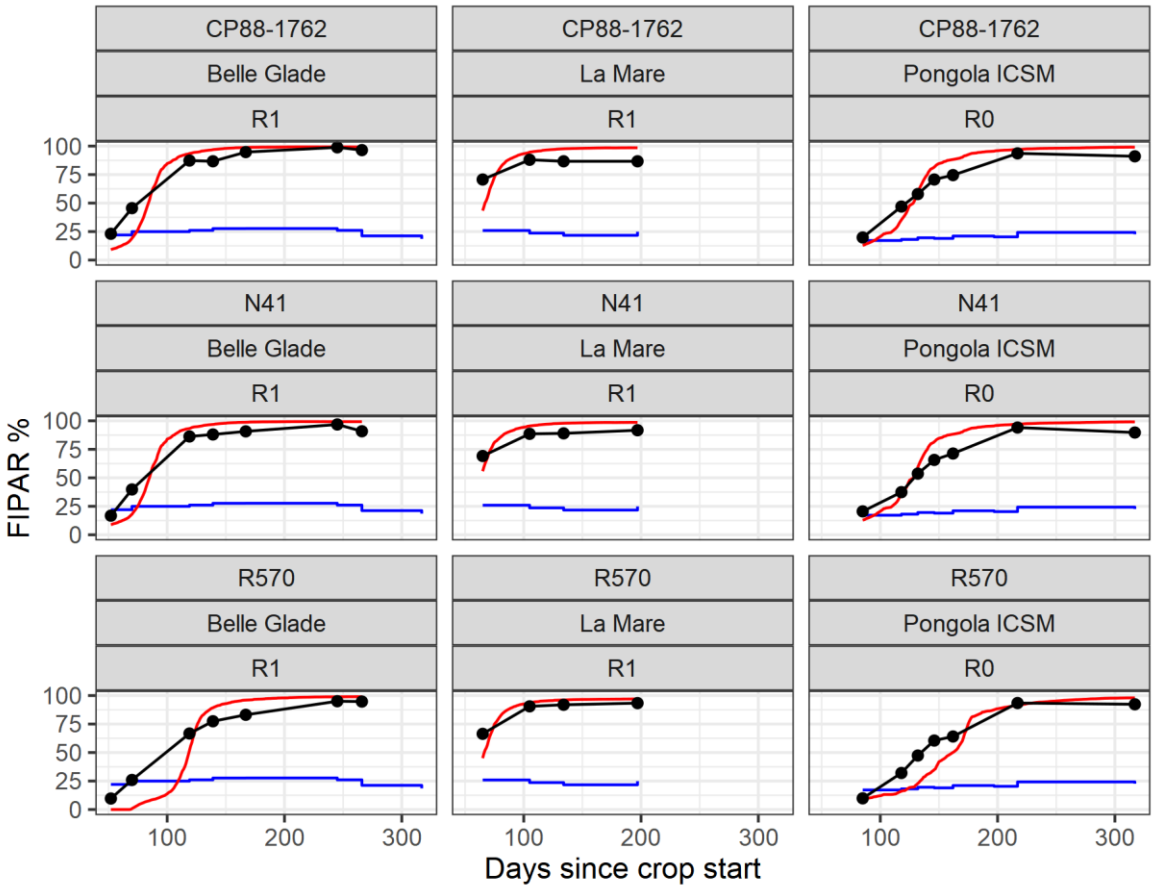

Figure 11-6. CaneGEM-simulated (red line) and observed (black points and linearly-interpolated lines) FIPAR, with predetermined duration of growth phase and G-specific calibration of canopy development parameters (model assessment scenario "GEM_IGEP_P"). Inter-sample period average daily air temperature is also shown (stepped blue line).

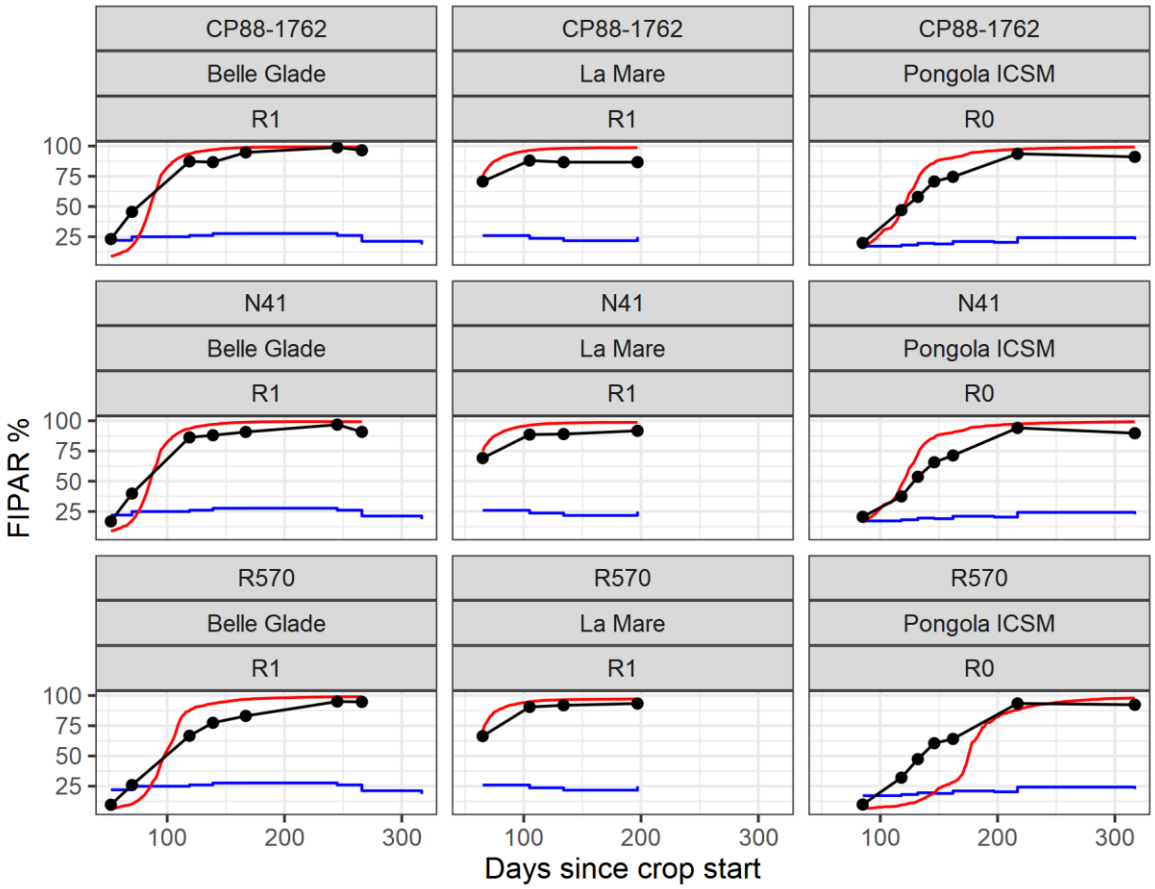

Figure 11-7. Simulated (CaneGEM model, with simulated emergence dates and G-specific calibration) vs observed crop development and biomass components, for the ICSM IGEP dataset (model assessment scenario "GEM_IGEP_S").

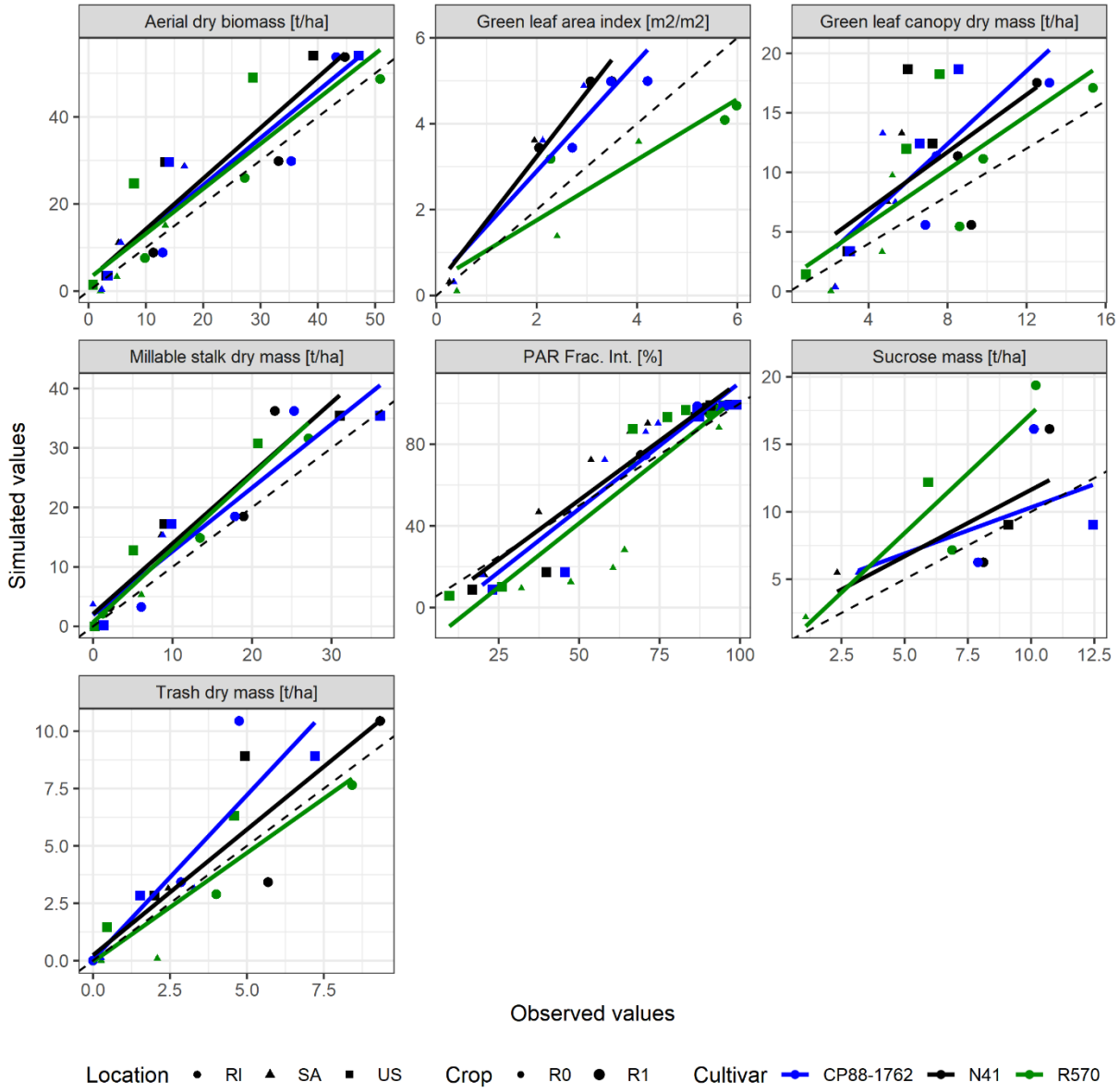

Figure 11-8. Simulated (CaneGEM model, with simulated duration of germination phase and G-specific calibration of canopy development parameters, model assessment scenario code "GEM_IGEP_P") vs observed crop development and biomass components, for the ICSM IGEP dataset.

11.4 Weather data graphs for case study

Weather data showing periods where data were patched (in red) are shown in Figure 11-9-Figure 11-13.

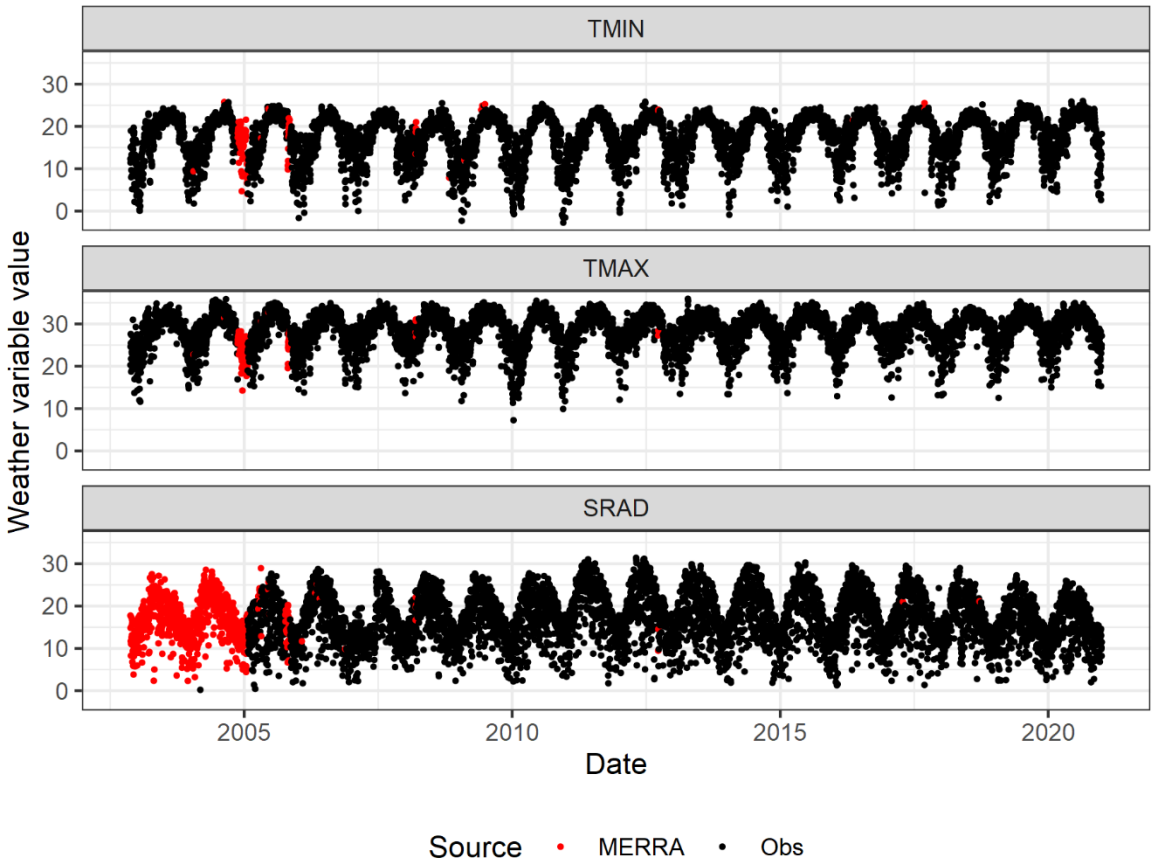

Figure 11-9. Daily weather data for Belle Glade, Florida. Observed data ("Obs", black points) were replaced with MERRA synthesized data (red points) where observed data were missing or out of expected ranges.

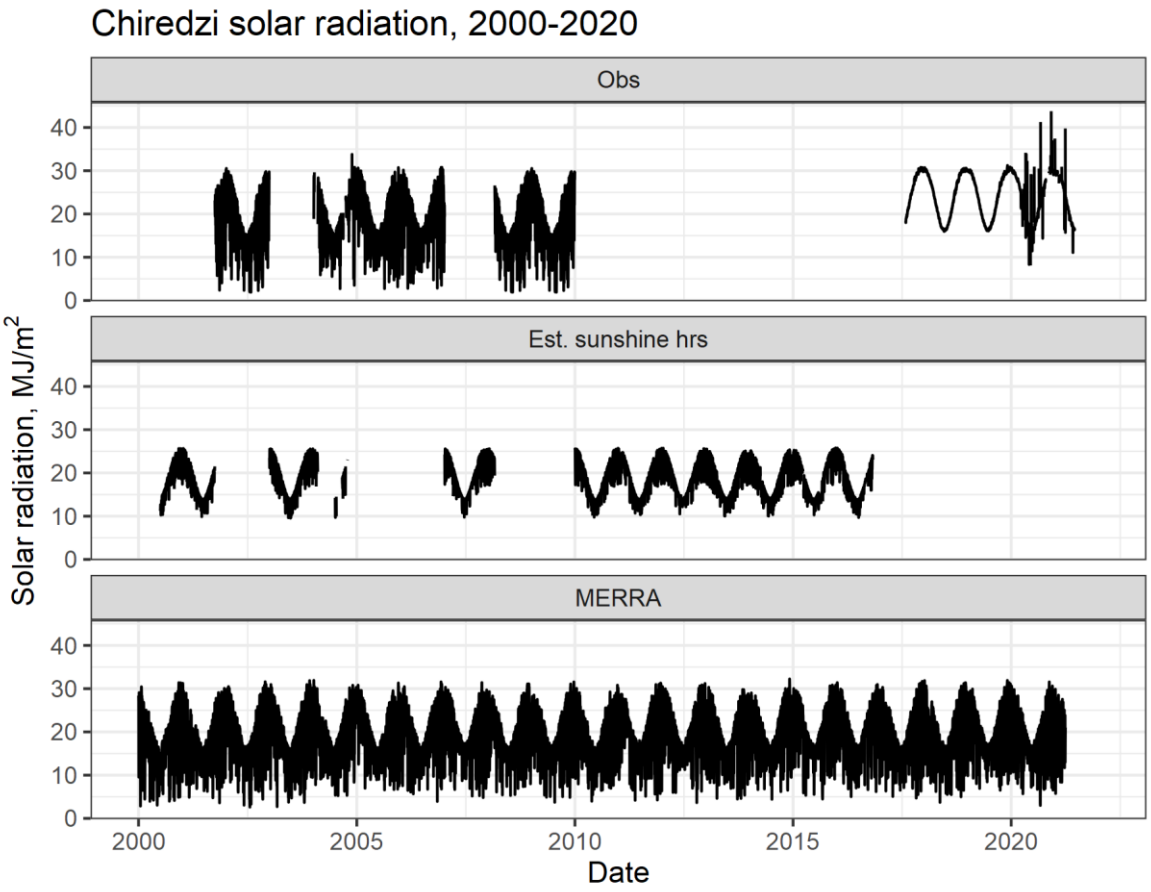

Figure 11-10. Solar radiation data for Chiredzi, Zimbabwe, 2000-2020, observed, estimated from sunshine hours, and estimated by NASA Power (MERRA).

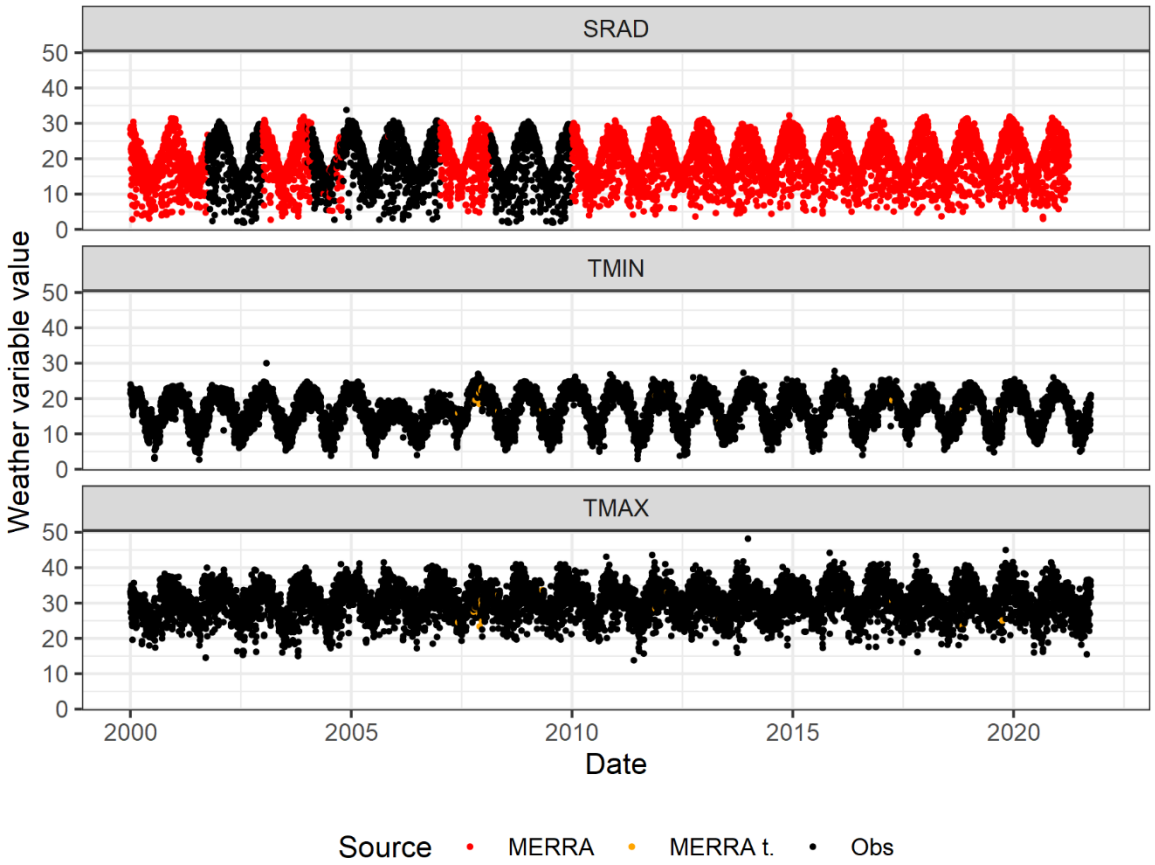

Figure 11-11. Daily weather data for Chiredzi, Zimbabwe. Observed data ("Obs", black points) were replaced with MERRA synthesized data (red points) or bias-adjusted MERRA data ("MERRA t.", orange points) where observed data were missing or out of expected ranges.

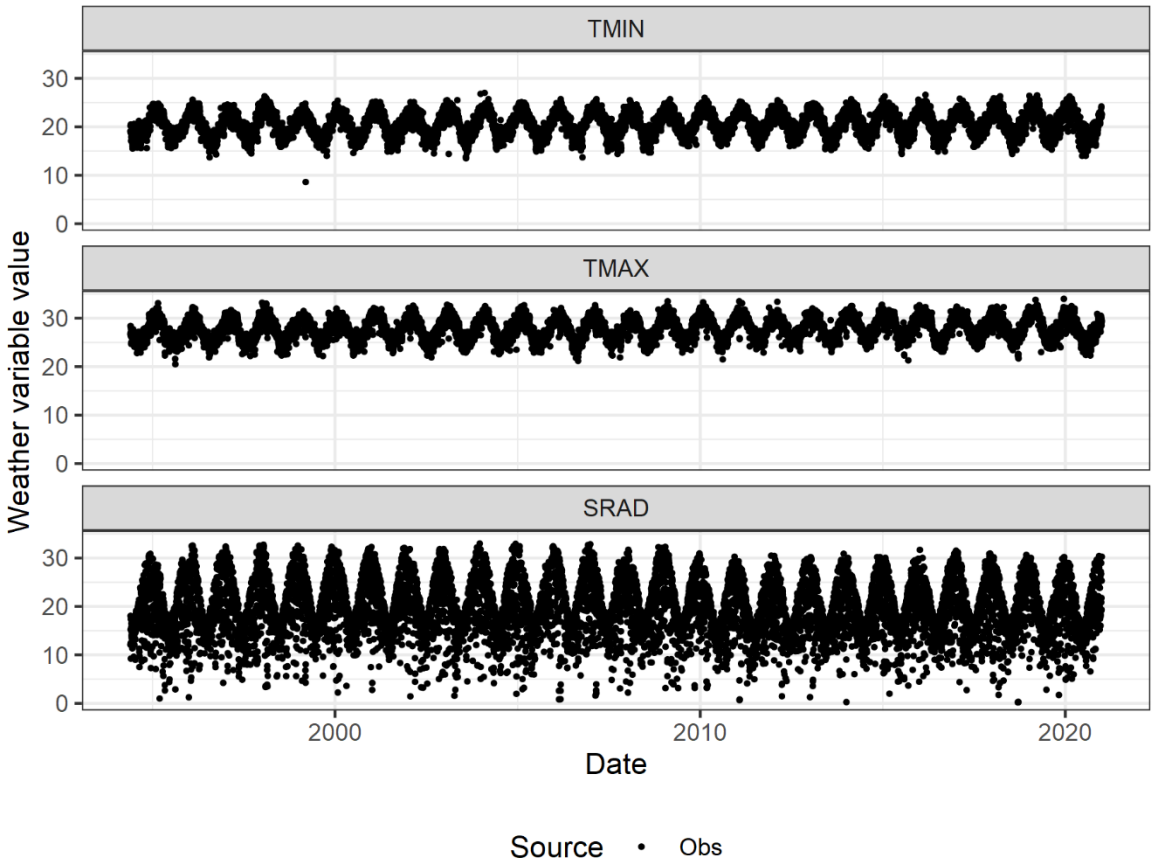

Figure 11-12. Daily weather data for La Mare, Reunion Island, France.

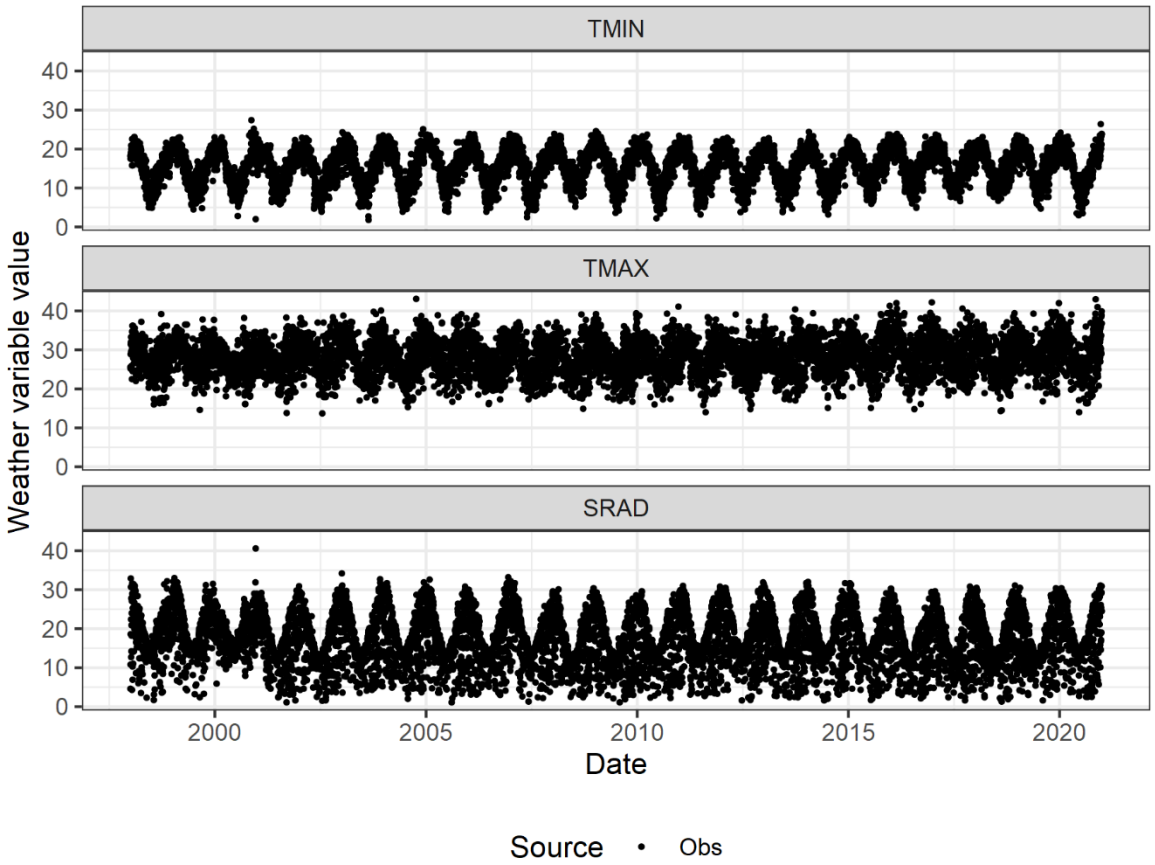

Figure 11-13. Daily weather data for Pongola, South Africa.

11.5 Results for case study

Table 11-2 shows a summary of results for the case study described in Chapter 6. Daily fractional interception of photosynthetically-active radiation, for early-season (E), mid-season (M) and late-season (L) crops at Belle Glade, 2003-2020, shown in Figure 11-14.

Table 11-2. Mean simulated duration of germination (DoGP, d after crop start), date of onset of stalk growth (doOSG, d after crop start), seasonal photosynthetically-active radiation interception fraction (FIPARa), seasonal apparent radiation use efficiency (RUEa, g/MJ), above-ground dry biomass at harvest (ADMh, t/ha), stalk dry mass at harvest (SDMh, t/ha) and sucrose mass at harvest (SUCh, t/ha), for the baseline cultivar NCo376, and three classes of genetic adaptations: modification of temperature sensitivity thresholds, relative canopy growth rate, and the green leaf area index threshold at which senescence starts. Data are shown for three times of harvest (early-, mid- and late-season) and the average of the three ('All').

		Base-line NCo	Temperature adaptation (°C)				Canopy growth rate adaptation (% change from baseline G)				Senescence threshold (% change from baseline G)			
H.S.	Var	376	-2	-1	+1	+2	-10%	-5%	+5%	+10%	-10%	-5%	+5%	+10%
Belle Glade, Florida, USA, 2002-2020														
All	DoGP	48.1	37.1	39.4	55.8	65.7	48.1	48.1	48.1	48.1	48.1	48.1	48.1	48.1
All	DoOSG	126.9	114.1	119.6	135.9	146.1	130.5	128.5	126.0	125.0	126.9	126.9	126.9	126.9
All	FIPARa	0.65	0.68	0.67	0.62	0.59	0.64	0.64	0.65	0.65	0.64	0.64	0.65	0.65
All	RUEa	1.69	1.81	1.76	1.62	1.53	1.68	1.69	1.70	1.70	1.68	1.69	1.70	1.71
All	ADMh	52.7	56.2	54.6	50.3	47.5	52.3	52.5	52.8	52.8	52.2	52.4	52.9	53.1
All	SDMh	37.6	40.7	39.2	35.6	33.3	37.0	37.3	37.7	37.9	37.1	37.3	37.8	38.0
All	SUCh	12.3	9.2	10.8	13.6	14.5	12.2	12.2	12.4	12.4	12.0	12.2	12.5	12.7
Early	DoGP	43.4	33.6	30.4	50.8	63.7	43.4	43.4	43.4	43.4	43.4	43.4	43.4	43.4
Early	DoOSG	160.4	143.5	150.7	174.3	187.5	167.1	163.6	158.8	157.2	160.4	160.4	160.4	160.4
Early	FIPARa	0.65	0.69	0.67	0.62	0.58	0.64	0.64	0.65	0.66	0.65	0.65	0.65	0.65
Early	RUEa	1.73	1.83	1.79	1.65	1.55	1.71	1.72	1.73	1.73	1.71	1.72	1.73	1.74
Early	ADMh	53.4	56.8	55.3	51.1	48.1	52.9	53.1	53.6	53.7	53.1	53.3	53.6	53.8
Early	SDMh	37.2	40.7	39.2	34.7	32.1	36.4	36.8	37.4	37.7	36.8	37.0	37.4	37.6
Early	SUCh	10.3	7.8	9.1	11.0	11.4	9.9	10.1	10.4	10.4	10.0	10.2	10.5	10.6
Mid	DoGP	69.5	51.4	59.2	80.2	93.1	69.5	69.5	69.5	69.5	69.5	69.5	69.5	69.5
Mid	DoOSG	133.2	117.4	124.2	142.8	154.2	136.5	134.6	132.1	130.8	133.2	133.2	133.2	133.2
Mid	FIPARa	0.62	0.67	0.64	0.58	0.54	0.61	0.61	0.62	0.62	0.61	0.61	0.62	0.62
Mid	RUEa	1.63	1.78	1.72	1.53	1.42	1.62	1.63	1.64	1.64	1.62	1.63	1.64	1.65
Mid	ADMh	50.9	55.6	53.5	47.7	44.1	50.6	50.8	51.1	51.0	50.5	50.7	51.1	51.3
Mid	SDMh	36.2	40.0	38.2	33.7	31.0	35.6	35.9	36.4	36.5	35.7	35.9	36.4	36.5
Mid	SUCh	11.6	8.9	10.4	12.5	13.1	11.5	11.6	11.7	11.6	11.3	11.5	11.8	11.9

H.S.	Var	Base-line NCo 376	Temperature adaptation (°C)				Canopy growth rate adaptation (% change from baseline G)				Senescence threshold (% change from baseline G)			
			-2	-1	+1	+2	-10%	-5%	+5%	+10%	-10%	-5%	+5%	+10%
Late	DoGP	31.5	26.3	28.5	35.8	40.4	31.5	31.5	31.5	31.5	31.5	31.5	31.5	31.5
Late	DoOSG	87.1	81.5	83.8	91.5	96.6	87.7	87.3	87.1	87.1	87.1	87.1	87.1	87.1
Late	FIPARa	0.67	0.70	0.69	0.66	0.64	0.67	0.67	0.68	0.68	0.67	0.67	0.68	0.68
Late	RUEa	1.72	1.80	1.76	1.67	1.61	1.72	1.72	1.72	1.72	1.70	1.71	1.73	1.74
Late	ADMh	53.6	56.1	54.9	52.0	50.3	53.5	53.6	53.7	53.7	53.0	53.3	53.9	54.1
Late	SDMh	39.3	41.4	40.3	38.1	37.0	39.2	39.3	39.4	39.4	38.7	39.0	39.6	39.8
Late	SUCh	15.1	10.8	12.9	17.1	19.0	15.1	15.1	15.1	15.1	14.6	14.9	15.3	15.5
Chiredzi, Zimbabwe, 2000-2020														
All	DoGP	36.7	30.4	33.0	41.5	47.6	36.7	36.7	36.7	36.7	36.7	36.7	36.7	36.7
All	DoOSG	108.9	92.4	98.0	119.7	130.4	113.5	110.9	106.0	104.1	108.9	108.9	108.9	108.9
All	FIPARa	0.71	0.73	0.72	0.68	0.65	0.70	0.70	0.71	0.71	0.70	0.70	0.71	0.71
All	RUEa	1.90	1.97	1.93	1.83	1.75	1.89	1.89	1.90	1.90	1.88	1.89	1.90	1.91
All	ADMh	66.6	69.2	67.9	64.3	61.3	66.4	66.5	66.6	66.6	66.0	66.3	66.8	67.1
All	SDMh	47.9	50.7	49.4	45.7	43.3	47.4	47.7	48.1	48.2	47.3	47.6	48.2	48.4
All	SUCh	21.4	17.4	19.4	22.6	23.5	21.3	21.3	21.4	21.5	20.9	21.2	21.6	21.8
Early	DoGP	26.7	22.8	24.1	30.9	37.8	26.7	26.7	26.7	26.7	26.7	26.7	26.7	26.7
Early	DoOSG	131.3	99.2	108.5	150.3	166.8	140.9	135.6	124.0	119.2	131.3	131.3	131.3	131.3
Early	FIPARa	0.71	0.75	0.74	0.68	0.64	0.70	0.71	0.72	0.72	0.71	0.71	0.71	0.72
Early	RUEa	1.95	2.00	1.97	1.88	1.76	1.94	1.94	1.95	1.95	1.93	1.94	1.96	1.96
Early	ADMh	68.3	70.0	69.1	65.9	61.5	67.9	68.1	68.4	68.4	67.7	68.0	68.5	68.7
Early	SDMh	48.4	51.3	49.9	45.8	42.0	47.6	48.0	48.8	49.0	47.8	48.1	48.6	48.9
Early	SUCh	20.5	16.2	18.3	21.2	20.6	20.0	20.2	20.6	20.7	20.0	20.3	20.7	20.9
Mid	DoGP	58.8	44.5	51.0	68.0	77.3	58.8	58.8	58.8	58.8	58.8	58.8	58.8	58.8
Mid	DoOSG	117.2	102.8	109.6	127.7	138.3	119.2	118.0	116.9	116.4	117.2	117.2	117.2	117.2
Mid	FIPARa	0.67	0.71	0.70	0.64	0.61	0.67	0.67	0.67	0.68	0.67	0.67	0.68	0.68
Mid	RUEa	1.80	1.94	1.88	1.70	1.61	1.80	1.80	1.80	1.80	1.79	1.79	1.81	1.81
Mid	ADMh	63.2	68.2	65.9	59.7	56.5	63.1	63.1	63.2	63.2	62.7	63.0	63.5	63.7
Mid	SDMh	45.7	50.1	48.0	42.6	40.1	45.3	45.5	45.8	45.8	45.1	45.4	46.0	46.2
Mid	SUCh	20.5	18.0	19.4	21.1	21.9	20.4	20.4	20.5	20.5	20.0	20.3	20.7	20.9

H.S.	Var	Base-line NCo 376	Temperature adaptation (°C)				Canopy growth rate adaptation (% change from baseline G)				Senescence threshold (% change from baseline G)			
			-2	-1	+1	+2	-10%	-5%	+5%	+10%	-10%	-5%	+5%	+10%
Late	DoGP	25.1	24.2	24.4	26.2	28.3	25.1	25.1	25.1	25.1	25.1	25.1	25.1	25.1
Late	DoOSG	77.1	75.0	75.5	79.7	84.5	79.1	77.8	76.2	75.9	77.1	77.1	77.1	77.1
Late	FIPARa	0.73	0.74	0.74	0.72	0.71	0.73	0.73	0.73	0.73	0.72	0.73	0.73	0.74
Late	RUEa	1.94	1.97	1.95	1.91	1.87	1.94	1.94	1.94	1.94	1.92	1.93	1.95	1.95
Late	ADMh	68.1	69.2	68.7	67.1	65.9	68.2	68.1	68.2	68.1	67.4	67.8	68.5	68.7
Late	SDMh	49.6	50.9	50.2	48.6	47.7	49.4	49.5	49.7	49.7	48.8	49.2	49.9	50.2
Late	SUCh	23.3	17.9	20.5	25.7	28.2	23.4	23.3	23.3	23.3	22.7	23.0	23.6	23.8
La Mare, Reunion island, 1994-2020														
All	DoGP	32.4	25.8	28.5	37.7	45.3	32.4	32.4	32.4	32.4	32.4	32.4	32.4	32.4
All	DoOSG	92.5	82.4	86.3	100.6	110.6	96.8	94.4	91.1	90.0	92.5	92.5	92.5	92.5
All	FIPARa	0.72	0.75	0.74	0.70	0.68	0.72	0.72	0.73	0.73	0.72	0.72	0.73	0.73
All	RUEa	2.03	2.10	2.07	1.97	1.90	2.03	2.03	2.03	2.03	2.01	2.02	2.04	2.05
All	ADMh	72.3	74.8	73.8	70.1	67.5	72.0	72.2	72.2	72.3	71.6	71.9	72.6	72.8
All	SDMh	52.5	55.1	54.0	50.4	48.0	51.9	52.2	52.6	52.7	51.7	52.1	52.8	53.1
All	SUCh	22.9	18.3	20.7	24.4	25.6	22.7	22.8	22.8	22.8	22.3	22.6	23.1	23.4
Early	DoGP	40.1	29.6	34.0	48.5	61.5	40.1	40.1	40.1	40.1	40.1	40.1	40.1	40.1
Early	DoOSG	113.7	95.6	102.8	127.5	142.8	120.3	116.6	110.9	108.5	113.7	113.7	113.7	113.7
Early	FIPARa	0.72	0.76	0.74	0.69	0.65	0.71	0.71	0.72	0.73	0.71	0.72	0.72	0.72
Early	RUEa	2.00	2.09	2.06	1.92	1.80	1.99	2.00	2.00	2.00	1.98	1.99	2.01	2.01
Early	ADMh	71.1	74.4	73.2	68.1	63.9	70.8	71.1	71.1	71.2	70.5	70.8	71.4	71.6
Early	SDMh	50.6	54.2	52.8	47.6	44.1	49.8	50.3	50.8	51.1	49.9	50.3	50.8	51.1
Early	SUCh	20.7	16.9	19.1	21.5	21.6	20.4	20.7	20.6	20.7	20.1	20.4	20.9	21.1
Mid	DoGP	34.7	27.2	30.5	39.8	46.8	34.7	34.7	34.7	34.7	34.7	34.7	34.7	34.7
Mid	DoOSG	89.3	79.3	83.4	96.6	105.6	93.3	91.3	88.1	87.1	89.3	89.3	89.3	89.3
Mid	FIPARa	0.71	0.74	0.73	0.69	0.66	0.70	0.71	0.71	0.71	0.70	0.71	0.71	0.71
Mid	RUEa	1.98	2.07	2.03	1.92	1.85	1.97	1.98	1.98	1.98	1.96	1.97	1.99	2.00
Mid	ADMh	70.5	73.5	72.1	68.3	65.8	70.1	70.4	70.4	70.4	69.8	70.1	70.8	71.0
Mid	SDMh	51.3	54.2	52.8	49.2	46.9	50.5	51.0	51.3	51.4	50.5	50.9	51.6	51.8
Mid	SUCh	21.9	17.5	19.8	23.5	24.7	21.6	21.8	21.8	21.8	21.3	21.6	22.2	22.4

H.S.	Var	Base-line NCo 376	Temperature adaptation (°C)				Canopy growth rate adaptation (% change from baseline G)				Senescence threshold (% change from baseline G)			
			-2	-1	+1	+2	-10%	-5%	+5%	+10%	-10%	-5%	+5%	+10%
Late	DoGP	22.5	20.7	21.0	24.8	27.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5
Late	DoOSG	74.6	72.3	72.6	77.7	83.5	76.8	75.2	74.5	74.4	74.6	74.6	74.6	74.6
Late	FIPARa	0.74	0.75	0.75	0.73	0.72	0.74	0.74	0.75	0.75	0.74	0.74	0.75	0.75
Late	RUEa	2.11	2.15	2.14	2.08	2.05	2.11	2.11	2.11	2.11	2.09	2.10	2.12	2.13
Late	ADMh	75.2	76.6	76.0	74.0	72.8	75.2	75.2	75.2	75.2	74.4	74.8	75.5	75.8
Late	SDMh	55.6	57.0	56.4	54.4	53.1	55.3	55.5	55.6	55.7	54.7	55.2	56.0	56.3
Late	SUCh	26.0	20.5	23.2	28.4	30.6	26.1	26.0	26.0	26.0	25.3	25.7	26.3	26.6
Pongola, South Africa, 1998-2020														
All	DoGP	36.5	28.7	31.8	43.1	54.5	36.5	36.5	36.5	36.5	36.5	36.5	36.5	36.5
All	DoOSG	119.9	102.3	107.6	132.7	148.6	125.3	122.3	117.6	116.2	119.9	119.9	119.9	119.9
All	FIPARa	0.67	0.71	0.70	0.64	0.60	0.66	0.67	0.68	0.68	0.67	0.67	0.67	0.68
All	RUEa	1.72	1.82	1.78	1.64	1.52	1.72	1.72	1.73	1.73	1.71	1.72	1.73	1.74
All	ADMh	53.3	56.2	55.0	50.5	46.9	53.1	53.2	53.4	53.3	52.8	53.0	53.5	53.7
All	SDMh	37.5	40.6	39.3	35.0	32.1	37.0	37.3	37.7	37.8	37.0	37.3	37.7	37.9
All	SUCh	17.3	13.7	15.7	18.1	18.5	17.2	17.2	17.3	17.3	16.8	17.1	17.5	17.6
Early	DoGP	30.8	24.1	26.7	37.7	56.4	30.8	30.8	30.8	30.8	30.8	30.8	30.8	30.8
Early	DoOSG	156.1	122.2	130.9	178.7	204.8	165.4	160.7	151.5	148.8	156.1	156.1	156.1	156.1
Early	FIPARa	0.66	0.72	0.70	0.62	0.55	0.65	0.66	0.67	0.68	0.66	0.66	0.67	0.67
Early	RUEa	1.78	1.87	1.83	1.65	1.48	1.76	1.78	1.79	1.79	1.77	1.77	1.79	1.79
Early	ADMh	55.1	57.7	56.7	51.1	45.6	54.6	55.0	55.4	55.3	54.7	54.9	55.3	55.4
Early	SDMh	38.3	41.8	40.5	34.4	29.8	37.4	38.0	38.8	38.8	37.8	38.1	38.5	38.7
Early	SUCh	16.3	13.1	15.0	15.8	14.5	15.9	16.2	16.5	16.5	15.9	16.1	16.5	16.6
Mid	DoGP	52.9	38.6	44.5	62.9	75.5	52.9	52.9	52.9	52.9	52.9	52.9	52.9	52.9
Mid	DoOSG	120.6	106.6	112.1	131.4	145.3	124.4	122.4	119.5	118.6	120.6	120.6	120.6	120.6
Mid	FIPARa	0.65	0.70	0.68	0.62	0.58	0.65	0.65	0.66	0.66	0.65	0.65	0.66	0.66
Mid	RUEa	1.64	1.78	1.72	1.55	1.43	1.64	1.64	1.64	1.64	1.63	1.63	1.65	1.65
Mid	ADMh	50.7	54.9	53.1	47.8	44.0	50.6	50.6	50.6	50.6	50.2	50.5	50.9	51.0
Mid	SDMh	35.7	39.4	37.9	33.3	30.3	35.3	35.4	35.8	35.8	35.2	35.4	35.9	36.1
Mid	SUCh	16.8	13.9	15.6	17.7	18.0	16.8	16.7	16.7	16.6	16.4	16.6	16.9	17.1

H.S.	Var	Base-line NCo 376	Temperature adaptation (°C)				Canopy growth rate adaptation (% change from baseline G)				Senescence threshold (% change from baseline G)			
			-2	-1	+1	+2	-10%	-5%	+5%	+10%	-10%	-5%	+5%	+10%
Late	DoGP	26.0	23.4	24.1	28.6	31.7	26.0	26.0	26.0	26.0	26.0	26.0	26.0	26.0
Late	DoOSG	82.9	78.2	80.0	87.9	95.8	86.1	83.8	81.9	81.2	82.9	82.9	82.9	82.9
Late	FIPARa	0.70	0.71	0.71	0.69	0.67	0.69	0.70	0.70	0.70	0.69	0.69	0.70	0.70
Late	RUEa	1.75	1.82	1.79	1.70	1.66	1.75	1.75	1.75	1.75	1.73	1.74	1.76	1.77
Late	ADMh	54.0	56.1	55.1	52.6	51.1	54.1	54.1	54.1	54.1	53.5	53.8	54.3	54.5
Late	SDMh	38.6	40.6	39.6	37.3	36.2	38.3	38.5	38.7	38.7	37.9	38.2	38.8	39.1
Late	SUCh	18.8	14.1	16.4	20.9	23.0	18.9	18.8	18.8	18.7	18.2	18.5	19.0	19.2

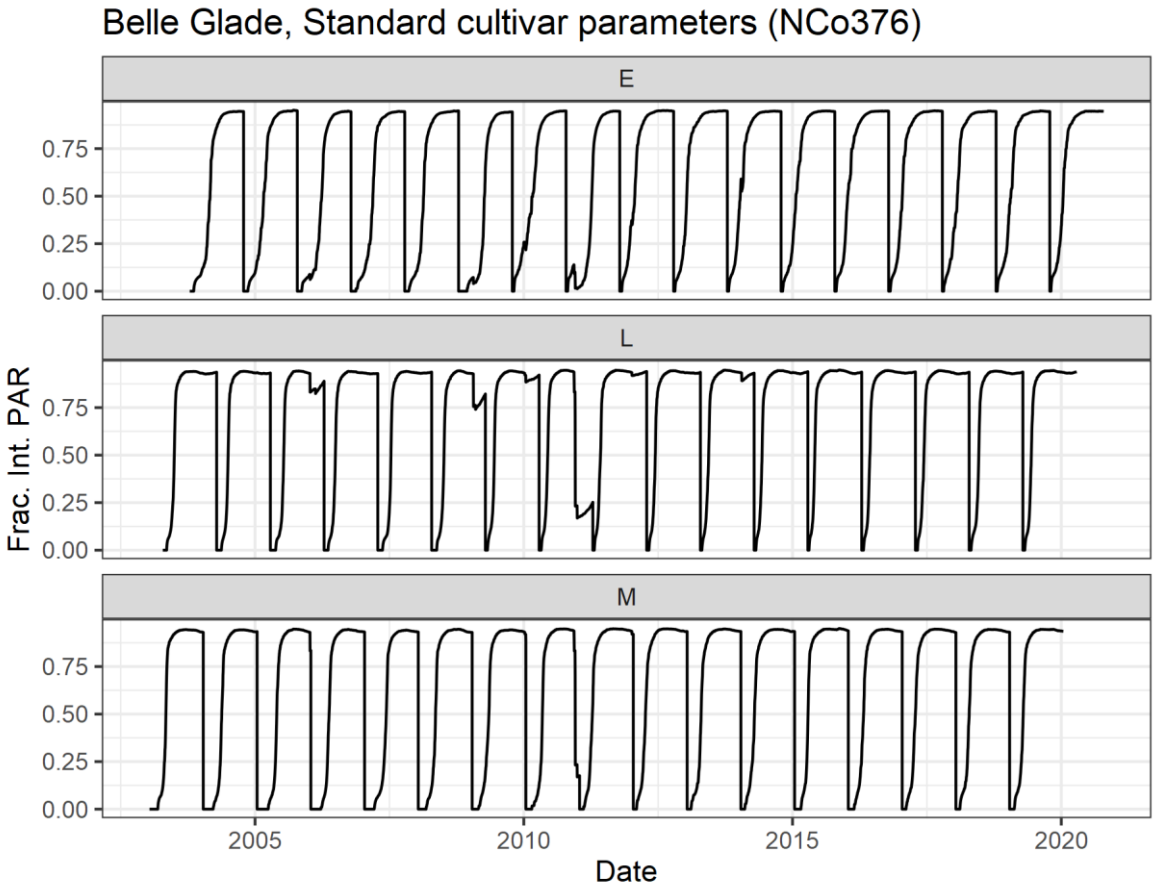

Figure 11-14. Daily fractional interception of photosynthetically-active radiation, for early-season (E), mid-season (M) and late-season (L) crops at Belle Glade, 2003-2020. Impacts of frost events, which reduce green leaf area index, are evident, e.g. 2011.

11.6 CaneGEM model source code file structure and operation

11.6.1 Access to the CaneGEM model

Please contact Matthew Jones (matthewjones0001@gmail.com) or Abraham Singels (abraham.singels@gmail.com) to request access to the CaneGEM model.

11.6.2 Setting up the R environment for running the CaneGEM model

The CaneGEM model is written in the R language.

1. It is recommended to install R and "RStudio" – both free downloads.
2. In RStudio, install the libraries tidyverse, broom, knitr, reshape2, kableExtra, ggpubr, chillR, lubridate and readxl, using the Tools menu.

11.6.3 How to simulate the ICSM experiments (Chapter 5) with the CaneGEM model

Steps for running the model

1. Open CaneGEM.Rproj in RStudio

2. Open 'CreateCluster.R' and press on source. This sets up parallel processing for fast model runs.
3. Open RunCaneGEM_ICSM.R
4. Click on "Source". This runs the model.
5. Outputs are generated in the ./outputs directory.

File structure

- The top-level R script to run is RunCaneGEM.R. This makes uses of the following additional script files in the root directory:
  - CLAImodelv13.11_func.R – functions used by the CaneGEM model
  - CLAI_utilities.r – additional functions used by the CaneGEM model
  - CreateCluster.R – script to create parallel processing environment to increase the execution speed of the CaneGEM model
  - obssimplot_v13.8_v2LinObsFi.r – functions for processing and visualisation observed and simulated data
- Model inputs are in the ./inputs directory. These are explained in the model source code.
- Genotype parameters are stored in ./modelparams_v13.11.csv (root working directory, not in ./inputs/)
- output is produced in the ./outputs/ directory. modelrun.csv and ModelRunDetails_YYYY-MM-DD HH.MM.SS.xlsx
- The names of the input and output parameters are not consistent with the thesis. An explanation is provided in ./Input_Output_acronym_mapping.xlsx

11.6.4 How to simulate the Case Study experiments (Chapter 6) with the CaneGEM model

1. Open CaneGEM_casestudy.Rproj in RStudio
2. Open 'CreateCluster.R' and press on source. This sets up parallel processing for fast model runs.
3. Open the following files and click ‘source’ to execute each one in sequence. Execution may take many minutes due to the large number of simulations.
   a. RunCaneGEM__casestudy_BelleGladeUSA.R
   b. RunCaneGEM__casestudy_ChiredziZimbabwe.R
   c. RunCaneGEM__casestudy_LaMareReunionFrance.R
   d. RunCaneGEM__casestudy_PongolaSouthAfrica.R
4. The scripts above generate output in the ./outputs directory. Files of approximately 200 Mb each are generated containing model output. These outputs can be reloaded into the R environment later on for further analysis without having to rerun the model. Open the file reprocessRDS.R for guidance.
5. Open and run (with ‘source’ button) WeatherDataSummary.R to generate graphs of weather data.
6. Outputs are stored in the ./outputs directory.