Supplementing Glasshouse Horticulture for Tomatoes with Mathematical Modelling

Brodie Lawson[1] Robert Cope[2] Brendan Harding[3]
Jessica Crawshaw[4] Mark McGuinness[5] Tal Kanety[6]
Keith Chambers[7] Keith Chan[8] Guy Metcalfe[9]
David Morselli[10] Mark Flegg[11]

Received 31 February 1066

DOI:10.21914/anziamproc.v65.20991, © Austral. Mathematical Soc. 2026. Published 2026-11-24, as part of the Proceedings of Mathematics in Industry Study Group, 2023. ISSN 1445-8810. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to the DOI for this article.

Abstract

This article details the findings of the Maths in Industry Study Group 2023 project with Costa Group, regarding commercial tomato cultivation in glasshouses. Through discussions with a lead grower in the company's tomato operation, and data from the company's Guyra glasshouses, several sub-teams investigated the current protocols for climate control inside the glasshouses from a mathematical perspective. Given the current operation supplies additional CO_2 to the tomato plants in order to increase photosynthesis, mathematical models were developed and statistically calibrated to data in order to evaluate the potential benefits of increased CO_2 supplementation. This modelling suggests that additional CO_2 would not contribute significantly to glasshouse interior temperature, and that although it would improve net photosynthesis a small amount, light remains the most important limiting factor. Trends in the site's CO_2 data also suggest that at the time when the data was taken the site's control over CO_2 levels could be improved. Additional modelling of tomato yield suggests an approach that improves calibration by concentrating on only the most meaningful portion of plant CO_2 consumption. Whilst this manuscript contains real data, this data is historical (exact period not disclosed) and all quantities do not accurately reflect the current operations at Costa Group.

MSC codes: 92C80; 00A69; 62F15

keywords: glasshouse horticulture, tomato, yield optimisation, mathematical modelling, Bayesian inference.

Contents

1 Introduction **M162**

2 Understanding Glasshouse Temperature Dynamics **M164**
2.1 Model for glasshouse temperature M164
2.1.1 Simplifying assumptions M165
2.2 Calibration of model to data M166
2.3 Contributors of heat energy M166

3 Understanding Glasshouse CO_2 Dynamics **M170**
3.1 Modelling CO_2 dynamics M170
3.2 Model Calibration M172
3.3 Validation of the Calibrated Model M174
3.4 Learnings M179

4 Modelling Tomato Yield **M185**
4.1 Maturity model M185
4.2 Stored value model M188

5 Conclusions **M193**

References **M198**

A Glasshouse Data **M203**
A.1 External weather conditions M204
A.2 Measured internal atmospheric conditions M205
A.2.1 Critical assumptions: mixing of internal glasshouse conditions M207
A.3 Glasshouse settings M209
A.3.1 Critical assumption: air exchange calculation M213
A.4 Yield measurements M215
A.5 Additional data considerations M217

B Temperature Model Details **M217**
B.1 Solar radiation . . . M217
B.2 Heating system . . . M218
B.3 Injection of flue gas . . . M218
B.4 Ventilation . . . M218
B.5 Conductive heat loss . . . M219
B.6 Misting . . . M219
B.7 Model statement . . . M219
B.8 Parameter inference . . . M220
B.8.1 Estimates for conduction and heating . . . M220
B.8.2 Estimates for the remaining parameters . . . M225
B.9 Contributions to changing glasshouse temperature . . . M229
B.10 Effect of CO2 pumping temperature . . . M231

C CO_2 Model Details **M233**
C.1 Modelling CO2 dynamics . . . M233
C.1.1 Injection . . . M233
C.1.2 Venting . . . M234
C.1.3 Plant activity . . . M235
C.2 Model Calibration and Validation . . . M238
C.3 Mass balance equation . . . M245
C.4 Qualitative analysis and expectations . . . M247
C.5 Glasshouse CO2 data is chaotic . . . M250

D Modelling Yield Details **M254**
D.1 Parameter estimation and indicative model performance . M257

1 Introduction

Glasshouse (and greenhouse) horticulture typically operates under heavily regulated conditions, seeking to maximise the use of light for photosynthesis (and hence yield) by controlling temperature and humidity [23], and potentially also supplying plants with additional carbon dioxide (CO_2) [38]. How these factors should be controlled will depend strongly on the specific growing operation, including for example the species and cultivar of plant and external climate conditions at the site. Furthermore, the nature of this control is often to hold (key aspects of) the glasshouse's interior climate at specific setpoints [7, 20]. Whether these setpoints are optimal is unlikely to be known, and indeed for a strategy to be truly optimum it must vary day-to-day in response to changing factors such as crop age and incident radiation [33]. The "optimum" is also probably best defined as a multi-objective optimisation problem, owing to the need to balance yield against operational costs such as energy consumption [30].

In order to begin trying to determine protocols that are optimum in some sense, one requires a mathematical description of the glasshouse system [14]. However, the sheer complexity of these systems means that no mathematical model can be expected to completely simulate all of their interacting components. As an example, the photosynthetic process through which plants obtain carbon from the air for growth is limited by the availability of electrons from light, the availability of Rubisco receptors, and the ability of CO_2 to compete with oxygen (O_2) to utilise these receptors [12]. Increasing temperature accelerates the rate of photosynthetic reactions until heat stress begins to compromise the process, suggesting an optimal temperature [4, 39]. However, plants also adapt the amount and the precise form of their Rubisco expression in response to changes in temperature, resulting in differences in potential rate and in affinity for CO_2 versus O_2 [6]. If a glasshouse is to be vented (say to reduce humidity, which also affects photosynthetic rate [5, 40]), then this will also affect its internal CO_2 concentration and temperature, producing a mixed effect on photosynthesis. Temperature also affects many other crop

properties, and in a cultivar-dependent fashion, as greenhouse-grown tomatoes demonstrate [35]. Similar webs of interactions can also be expected to characterise other important physiological processes in the plant of interest, such as respiration, resource allocation and fruit development.

Rather than attempt to comprehensively model all of the interdependent pieces of the plant-climate system, and then hope that their interactions will manifest in physiologically realistic predictions, a separate approach is to start from an already successful growing operation and utilise more targeted models that concentrate on individual aspects of the problem. Especially in a greenhouse or glasshouse system, where environmental conditions are tightly controlled and hence their effects less variable, processes not modelled can be absorbed into the parameters that describe the processes that are modelled. Calibration to data from the growing operation can be used to help ensure that these parameters are set correctly. The workshop of the Maths in Industry Study Group (MISG), hosted in 2023 at Monash University, provided a great opportunity for such an approach. Working together with a head grower from the Costa Group, mathematicians gathered together to model aspects of the Costa Group's tomato growing operations in the Guyra region of New South Wales, Australia, which use a number of large glasshouses (about $400\,000\,\mathrm{m}^2$ in footprint).

During the course of this workshop, different subteams constructed mathematical models for the dynamics of interior glasshouse temperature, internal CO_2 concentration (and hence photosynthetic activity), and tomato yield, and calibrated these models to data from some of the Guyra glasshouses. The calibrated models suggest a number of important conclusions about the tomato growing operation, and in particular, how increasing CO_2 supply is expected to affect interior conditions and plant behaviour. Our modelling suggests that the glasshouse temperature is mostly affected by venting and the misting system used for humidity control, along with heating due to incident light, and the effect of injected CO_2 gas on temperature is minimal. However, the expected gains in photosynthesis if CO_2 supply is increased are not large (at least using the current injection strategy), as the current

growing operation is predominantly light-limited. Additional modelling using our calibrated description of photosynthesis quantifies how the cost of CO_2 supply affects these considerations. We also find that by modelling glasshouse yield directly, some of the complexities of the photosynthetic process can be sidestepped, and a model with less mechanistic detail can be calibrated very well to Costa's data, hence providing good predictions of a glasshouse's meaningful output.

2 Understanding Glasshouse Temperature Dynamics

2.1 Model for glasshouse temperature

In the model, we will use $T(t)$ to represent the modelled temperature in the glasshouse at time t. The following energy balance equation asserts that the change in heat energy in the glasshouse is accounted for by contributions by various heat sources and sinks:

$$\frac{d}{dt}(\rho V_g C_p T) = Q_{rad} + Q_{heat} + Q_{flue} + Q_{vents} + Q_{cond} + Q_{mist}. \qquad (1)$$

The left hand side of Equation (1) is the rate of change of glasshouse air internal energy described by temperature T. The quantities ρ, V_g and C_p are respectively the mass density, total volume and specific heat capacity of the glasshouse air. The right hand side represents the various heat exchanges that the glasshouse air is subject to. Specifically, we assume that heat energy is added due to solar radiation, Q_{rad}, the heating system, Q_{heat}, the entry of flue gas, Q_{flue}, is also regulated by air exchange due to ventilation, Q_{vents}, conductive heat transfer to and from the environment, Q_{cond}, and may be lost due to misting, Q_{mist}. Note that gain and loss of energy is indicated by the sign (+ or −) of the terms on the right hand side (RHS) of the equation. The model written in its fullest form is presented in the Appendix as Equation (11). This form contains the expanded version of the various fluxes on the RHS of Equation (1) and these are detailed and explained in the Appendix.

2.1.1 Simplifying assumptions

We will assume the following for simplicity.

- **Equal and constant mass densities:** We assume that the mass density of the glasshouse air, ρ, the injected flue gas, ρ_{flue}, and the outside air, ρ_{atm} are equal and do not change significantly with time:

$$\rho = \rho_{\text{flue}} = \rho_{\text{atm}}, \qquad \frac{d\rho}{dt} = 0. \tag{2}$$

  We argue that the assumption is valid since (i) the Costa Group glasshouses are typically well-ventilated, ensuring that the glasshouse air content is similar to that of the outside air, and (ii) the mass densities of flue gas and outside air are close in value. Indeed, at the typical temperature of injection ($T_{\text{flue}} = 50°C$), we have $\rho_{\text{flue}} \approx 1.125\,\text{kgm}^{-3}$, in comparison to $\rho_{\text{atm}} \approx 1.204\,\text{kgm}^{-3}$ at typical outside temperature ($T = 20°C$) [27, 34].

  The assumption of time independence is common among established greenhouse temperature models [26, 32]. We justify this approximation for a Costa Group glasshouse since the relative humidity is tightly controlled (typically only ranging from $\sim 70\%$ to $\sim 91\%$) and glasshouse temperature typically only varies within a $\sim 20°C$ range each day-night cycle.

- **Equal and constant heat capacities:** We likewise assume that the specific heat capacities of the glasshouse air, C_p, the flue gas, $C_p^{(\text{flue})}$, and outside air, $C_p^{(\text{atm})}$, are equal and independent of time:

$$C_p = C_p^{(\text{flue})} = C_p^{(\text{atm})}, \qquad \frac{dC_p}{dt} = 0. \tag{3}$$

  Indeed the specific heat of flue gas, $C_p^{(\text{flue})} = 1.05\ \text{kJK}^{-1}\text{kg}^{-1}$ at $T = 50°C$, is approximately that of air, $C_p^{(\text{atm})} = 1.006\ \text{kJK}^{-1}\text{kg}^{-1}$ at $T = 20°C$. The assumption of time independence is reasonable since the specific heat capacities changes negligibly in the small temperature intervals we consider.

2.2 Calibration of model to data

We use Bayesian inference to get best estimates for the parameters of the temperature model. For time-dependent input functions for which data exists at discrete times, we use a smoothing spline interpolation. The method details are presented in the Appendix.

Figure 1 demonstrates the accuracy of the fitted model to the data provided by Costa group. In this figure we show the fitted model output for glasshouse temperature $\mathsf{T}(\mathsf{t})$, the confidence intervals corresponding to one and two standard deviations, and the measured data. The interval of time corresponds to a 5-day interval starting at 0:00 on the 23rd of January. Importantly, this model is quite capable of capturing rapid temperature changes as a result of switching the heating system on. This rapid change can be seen for example in Figure 2 which focuses on a short period of time when the temperature in the glasshouse is dropping and the heating system is suddenly switched on.

The model is essential to optimisation of the glasshouse. The parameters of the model are derived from the calibration process and present a good summary of the quantitative relationships between operational factors. Here is a summary of the parameters used in the model, their meaning, and the value of these parameters as determined by Bayesian fitting with the data. These parameters provide a quantitative summary of the thermodynamic properties of the glasshouse. Importantly, the confidence in these parameters is codependent and the distributions for these parameters can be found in the Appendix.

2.3 Contributors of heat energy

One of the key outcomes of the heat model that was investigated at MISG is the use of the model to shed insight into the key contributors to heat energy in the glasshouse. To do this, we assume typical values for operational conditions. That is, we assume that

1. Solar radiation is at $500\mathrm{W/m^2}$,

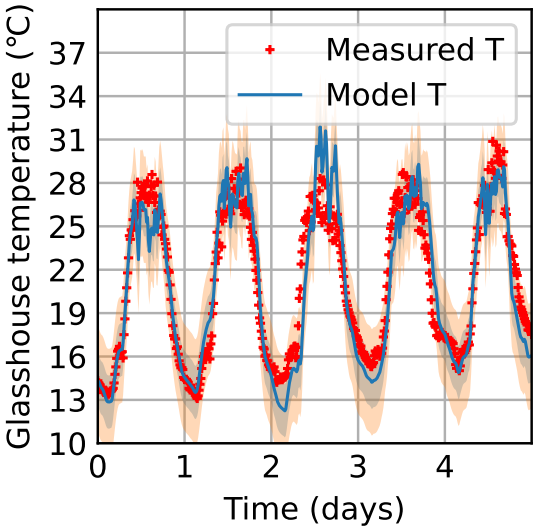


Figure 1: Plot of the solution to the full model (1) (detailed in the Appendix Equation (11)). The Model temperature presented here in blue is the output of the model which uses the mode of the posterior distribution for model parameters and is plotted beside the temperature measured at the glasshouse. The darker shaded region indicates the envelope for one standard deviation noise while the lighter region corresponds to two standard deviations.

2. The temperature in the glasshouse is 25 degrees cooler than the water in the heating system and the flue gas but 5 degrees warmer than the outside temperature,

3. The rate of injected flue gas by volume is 0.13 of the glasshouse volume per day whilst the rate of exchanged air through the vents by volume is 200 times the glasshouse volume per day.

Under these typical conditions and parameters found from model fitting, we are able to determine the key contributors to temperature regulation in the glasshouse (see Figure 3). Unsurprisingly, solar radiation is the primary source

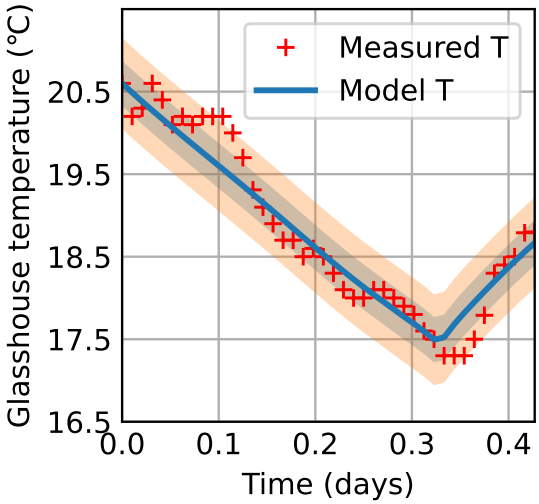


Figure 2: Fit of the model (blue) to measured values of glasshouse temperature (red) close to a rapid temperature change due to the engagement of the heating system. The darker shaded region indicates the envelope for one standard deviation noise while the lighter region corresponds to two standard deviations. Time $t = 0$ corresponds to data taken at 21:30 on 19th February.

Param.	Meaning	Base Value	Std Dev.
k_{rad}	Radiative heat absorption rate	$0.65°C\ day^{-1}(W/m^2)^{-1}$	0.12
k_{vents}	Air mixing inefficiency (floor to ceiling)	0.66	0.16
k_{mist}	Rate of temperature cooling from misting (valve fully open)	$870°C\ day^{-1}$	206
k_{heat}	Efficacy of the heating system	$1.38\ day^{-1}$	0.11
k_{cond}	Efficacy of conduction with environment	$1.87\ day^{-1}$	0.09
σ	Measured temperature noise (Std Dev.)	$1.84\ °C$	0.11

Table 1: Mode parameter values found from calibration of heat model. The precise definitions of these parameters can be inferred from the model equation in the Appendix Section B.7. Also in the Appendix is a detailed presentation of the inferred distribution of these parameters.

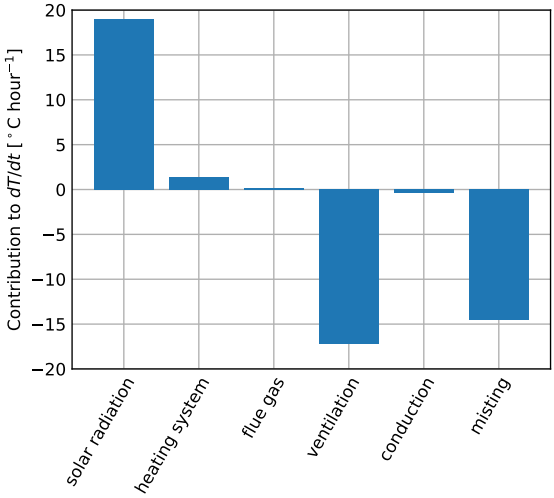


Figure 3: Typical impact of the heat sources/sinks on changing glasshouse temperature. These are the magnitudes of the terms on the right side of equation (11) using the parameter estimates (17) and reference values (18).

of heat during the day and the greatest sinks of heat are from ventilation and misting. Interestingly, the flue gas (due to the small volumes) provide very little overall heat energy to the glasshouse. It is estimated that if the flue gas is to contribute heat energy on the scale of the heating system, the flue gas temperature would need to exceed around 270 degrees. Therefore we conclude that the flue gas has a negligible impact on the temperature of the glasshouse. An open question remains 'If the flue gas is pumped in at lower temperatures, would less CO_2 be ventilated and therefore more CO_2 be uptaken by the crop?'. A preliminary finite element computation of the air flow in the glasshouse was conducted to investigate this question but more work is required to calibrate this model to give conclusive results.

3 Understanding Glasshouse CO_2 Dynamics

3.1 Modelling CO_2 dynamics

CO_2 , along with light, is the key ingredient for plant photosynthesis, which converts CO_2 in the air into usable carbon for compounds the plant requires such as cellulose [24]. For this reason, the Costa glasshouses use a CO_2 injection system to increase the rate of photosynthesis, and hence the growth rate of the tomato plants and their fruits. Other sources of CO_2 inside the glasshouse include plant respiration (where plants uptake O_2 and release CO_2), and CO_2 that enters through the glasshouse venting system when it is active. CO_2 is lost from the glasshouse atmosphere when it is taken up by plants, and when it leaves through the glasshouse venting system. As the Costa glasshouses use fans for air circulation, we make the simplifying assumption that the spatial variation in CO_2 concentration is low (a well-mixed assumption), hence avoiding the need for spatial coordinates or partial differential equations. Conservation of mass for CO_2 , m_{CO_2} , then results in an ordinary differential equation,

$$\frac{dm_{CO_2}}{dt} = \dot{m}_{\text{inject}} + \dot{m}_{\text{vent-in}} + \dot{m}_{\text{resp}} - \dot{m}_{\text{vent-out}} - \dot{m}_{\text{uptake}},$$

where $\dot{m}_{\text{process}}$ denotes the rate of change of mass of CO_2 that is transferred (in $\text{kg}\,\text{h}^{-1}$) in or out of the glasshouse due to the different processes. Our consideration of additional CO_2 supplementation corresponds to increasing $\dot{m}_{\text{inject}}$, although it should be noted that this has a monetary cost associated with it that needs to be taken into account in determination of an optimum strategy. The equation also highlights that reducing loss of CO_2 through the vents is a way to achieve the same thing, without this cost.

The tomato plants' ability to photosynthesise depends on the CO_2 concentration in the air around them, and if it is assumed that the air pressure inside the glasshouse can be treated as constant, then the net effect of the two terms related to venting becomes a push towards equilibrium of CO_2 concentration between the glasshouse and its surrounding environment, occurring at a rate

defined by the rate of air exchange. These observations motivate expressing the model in terms of CO_2 concentration, C, where it becomes

$$\frac{dC}{dt} = \alpha_{inject}\frac{\dot{m}_{inject}}{V_g} + \alpha_{vent}\frac{\dot{V}_{vent}}{V_g}(C_{atm} - C) + \frac{A_{plants}}{V_g}(-\mathcal{R}_{photo} + \mathcal{R}_{darkresp}).$$

Here, V_g denotes the glasshouse's (fixed) volume, C_{atm} describes the (time-varying) atmospheric concentration of CO_2 outside of it, and A_{plants} describes the active surface area of all the plants inside the glasshouse, which is treated as a fixed constant to be learned via calibration due to this not being present in the data. Each $\mathcal{R}$ term denotes the rate (per surface area) of CO_2 exchange between plant tissue and the surrounding environment due to a different physiological process. Data *is* available for the amount of CO_2 injected into the glasshouse ($\dot{m}_{inject}$), and the rate of air exchange through the vents ($\dot{V}_{vent}$), however additional flexibility is lent to the model through the constants α_{inject} and α_{vent}, which are learned via calibration and permit the contribution of these terms to the CO_2 dynamics to be scaled.

Full details regarding the CO_2 model are given in Appendix C, but in brief, the model is based largely around the Farquhar *et al.* [12] model of photosynthesis in C3 plants, which provides a description of the net effect of photosynthesis and photorespiration, $\mathcal{R}_{photo}$ that includes dependencies on incident light, CO_2 concentration, and interior temperature. We also use their approach for dark respiration (that is, release of CO_2 and uptake of O_2 by plants in a non-light-dependent fashion), which assigns it a dependence on temperature, and also include a parameter that changes the baseline rate of dark respiration between nighttime and daytime. The oxygen concentration inside the glasshouse is treated as fixed, as although O_2 is output by photosynthesis and taken in from the air by photorespiration and dark respiration, the baseline O_2 concentration is still one hundred times larger than the largest CO_2 concentration recorded in the provided data, such that O_2 changes would be minimal.

3.2 Model Calibration

In total, the model features 24 parameters, which are listed in Table 2 (details on how these parameters feature in the model are given in Appendix C). Suggested parameter values for their photosynthesis formulation were published by Farquhar *et al.* [12], however these were chosen in response to separate experiments, and not in tomato plants (let alone the specific variety in use at the Guyra glasshouses). Our model also includes additional parameters, including the data-correction parameters for injection and venting, the multiplier that scales nighttime dark respiration, and the surface area of active plant material. For these reasons, we again use Bayesian model calibration and data provided by Costa Group's representative to determine parameter values (for the 17 parameters that are not fixed *a priori*) that plausibly describe their specific growing operation. This generates an ensemble of plausible models, each with its own set of calibrated parameter values.

Full details of the Bayesian calibration process are given in Appendix C, but we briefly describe the important details here. Five days of data from a Guyra glasshouse in late June (including CO_2 concentrations, internal and external temperature, and rates of venting and CO_2 injection) were used for calibration. However, a direct application of the Bayesian calibration routine (sequential Monte Carlo [11] in this case) to this data using a non-informative prior was found to produce a good agreement between data and model, but physiologically unfeasible predictions about photosynthesis dynamics. Therefore, a two-stage calibration process was adopted instead, allowing a more restrictive prior to be assigned to important model parameters in a data-informed manner.

In this approach, only the Farquhar *et al.* model parameters that govern plant CO_2 exchange were first separately calibrated to published data for how photosynthetic rate depends on CO_2 concentration, light, and temperature in tomatoes [8]. Again using the Bayesian calibration routine with a wide prior to obtain plausible calibrations, we chose that which achieved the best fit to the tomato photosynthesis data and used those parameter values as

Variable Parameters		
Param.	**Meaning**	**Best Value**
α_{inject}	Scaling factor for CO_2 injection data	1.516 *(no dim.)*
α_{vent}	Scaling factor for vent air exchange data	1.553 *(no dim.)*
C^*_{atm}	Atmospheric CO_2 level (at $T = T^*$)	$5.00 \times 10^{-4}\,\text{kg m}^{-3}$
A_{plants}	Surface area of active plant matter	$3.698 \times 10^{5}\,\text{m}^2$
r_{cmax}	CO_2-saturated photosynthesis rate	$5.804 \times 10^{-3}\,\text{kg m}^{-2}\,\text{h}^{-1}$
E_c	r_{cmax} activation energy (T-dependence)	$7.257 \times 10^{4}\,\text{J mol}^{-1}$
K_c	Michaelis–Menten parameter for CO_2	$9.384 \times 10^{-4}\,\text{kg m}^{-3}$
E_{Kc}	K_c activation energy (T-dependence)	$4.759 \times 10^{4}\,\text{J mol}^{-1}$
K_o	Michaelis–Menten parameter for O_2	$0.167\,\text{kg m}^{-3}$
E_{Ko}	K_o activation energy (T-dependence)	$4.217 \times 10^{4}\,\text{J mol}^{-1}$
β	Potential CO_2 production from light	$1.446 \times 10^{-4}\,\text{kg W}^{-1}\,\text{h}^{-1}$
$I_{\text{sat-max}}$	Light level of photosynthesis saturation	$307.3\,\text{W m}^{-2}$
E_I	$I_{\text{sat-max}}$ activation energy (T-dependence)	$2.632 \times 10^{4}\,\text{J mol}^{-1}$
Z	Curvature parameter for light saturation	$202.8\,\text{W m}^{-2}$
r^*_d	Respiration rate at baseline temperature	$8.516 \times 10^{-5}\,\text{kg m}^{-2}\,\text{h}^{-1}$
E_d	r_d activation energy (T-dependence)	$4.634 \times 10^{4}\,\text{J mol}^{-1}$
γ_{night}	Increase in nighttime respiration	0.203 *(no dim.)*
Fixed Parameters		
V_g	Volume of the glasshouse	$3.28 \times 10^{5}\,\text{m}^3$
T^*	Baseline temperature	$11.5\,^{\circ}\text{C}$
O	Concentration of O_2 in glasshouse	$0.2577\,\text{kg m}^{-3}$
R	Universal gas constant	$8.314\,\text{J mol}^{-1}\,\text{K}^{-1}$
S	Entropy for light saturation "reaction"	$690.27\,\text{J K}^{-1}\,\text{mol}^{-1}$
H	Enthalpy for light saturation "reaction"	$2.277 \times 10^{5}\,\text{J mol}^{-1}$
I_{day}	Threshold irradiance delineating daytime	$5\,\text{W m}^{-2}$

Table 2: Parameter values for which the CO_2 model provides the best fit to the five days of Costa's CO_2 data used for calibration. Variable parameters are those that were allowed to vary during the calibration process (further details in the Appendix), while fixed parameters were pre-specified.

a sensible physiological baseline (see Table 4 in Appendix C). Calibration to the five days of Guyra glasshouse data then used a more restrictive prior on these pre-calibrated parameters, such that model predictions would not be allowed to stray too far from a physiologically plausible baseline. Details about this prior, including the window of values for each parameter that it enforces, are given in Appendix C.

The result of our calibration process is a set of fitted models that attain a good, but imperfect fit to Costa's CO_2 readings in the glasshouse. In particular, the model shows a marked ability to capture the complex *trends* in CO_2 across the five days, faithfully representing the effects of injection and venting events, and correctly recovering both the heavy consumption of CO_2 by plants during the day when light feeds the photosynthetic process, and their slower release of CO_2 during the night. However, the precise heights of peaks, or depths of post-venting troughs in the CO_2 data are not always faithfully captured. Given that the accuracy of the data is not guaranteed (in particular, the venting rate data is reportedly estimated by Costa's software, rather than directly measured), and the requirement that the model deal with very rapid changes in CO_2 concentration due to venting and injection, we consider the model's quality sufficient for its predictions to be meaningful.

3.3 Validation of the Calibrated Model

Rather than relying solely on the accuracy of the fitted curves shown in Figure 5 as a measure of the CO_2 model's quality, we also judge it using several additional validation metrics. Firstly, we consider whether or not the model still predicts physiologically reasonable photosynthesis curves, after it has been calibrated to the Guyra glasshouse data. As shown in Figure 5, the best-fit model still does a good job of representing experimentally-observed trends in tomato photosynthesis curves [8], producing saturating curves with plausible rates and capturing the subtle effect that higher temperatures are detrimental to CO_2 intake at low ambient CO_2 levels, but beneficial at higher CO_2 levels. However, the model does under-predict the overall rate of

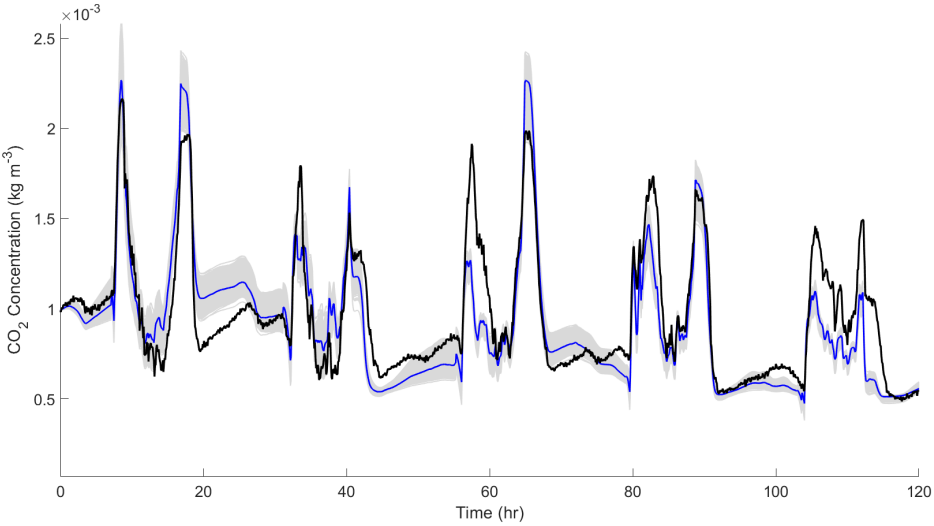


Figure 4: The predictions of the model ensemble (grey), including the best-fitting member (blue), as compared to Costa's sensor data (black) for five days of glasshouse operation during late June. The qualitative behaviour of CO_2 is represented well by the model, despite the precise value not always being accurately captured.

photosynthesis at low light levels (bottom few curves in left panel, and all curves in right panel which are for a fixed, moderately low level of light). We are satisfied the model remains physiologically plausible, and note that an exact match to experimental data should not necessarily be expected given potential differences between cultivars.

To further validate the model, we also evaluated its predictions against data that was provided by Costa, but unseen by the calibration process (which only used the five days in late June). For a fair test, we systematically select three more five-day periods all exactly three months apart, in order to have an example from each season without cherry-picking. For each five-day test period, we supply the model with Costa's data for the time series for CO_2 injection and venting during the period, as well as interior and exterior

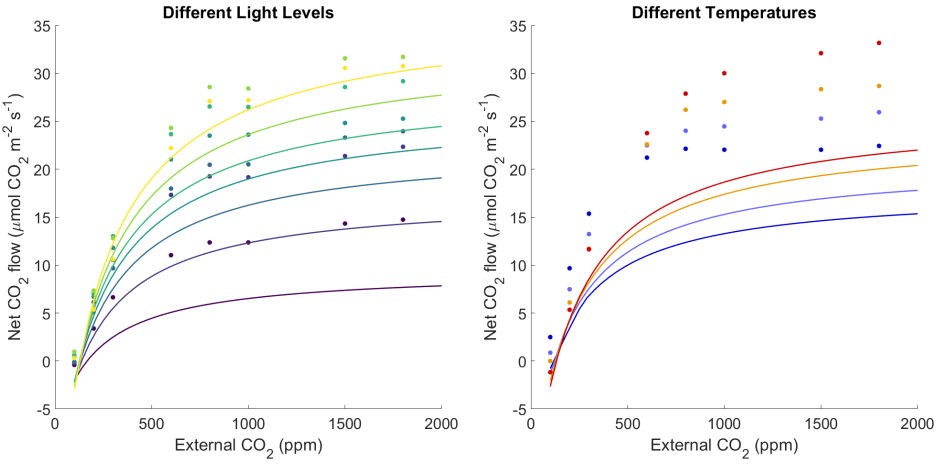


Figure 5: The dependence of CO_2 intake rate on temperature, light, and ambient CO_2 level, as predicted by the best-fit model (curves), and experimentally observed in [8]. **Left:** Dependence of CO_2 intake on ambient CO_2 level, for different light levels (and slight variations in temperature), brighter curves corresponding to higher levels of light. **Right:** Dependence of CO_2 intake on ambient CO_2 level, for different leaf temperatures (and a fixed light level of about $417\,\mathrm{W\,m^{-2}}$), redder curves corresponding to hotter temperatures. Both graphs show a good capture of all the qualitative trends seen in the data, but an overall underestimation of photosynthetic rate at lower light levels (which also causes an underestimation in the second graph).

temperature measurements. No adjustments to the model are made. The results of this test are given in Figure 6. During the colder season, very rapid variations in CO_2 are seen, and the model's predictions still qualitatively capture the behaviour but become rather poor about peak CO_2 levels and the amount of CO_2 present during the nighttime. On the other hand, despite seeing no training data from the warmer months, where the glasshouse CO_2 dynamics appear more strictly controlled, the predictions of the model are very good, with only the amplitude of very sharp CO_2 peaks being overestimated. This performance on unseen data further validates the model's predictive

ability.

Lastly, we consider how reasonable the calibrated parameter values are in the best-fit model, in particular the parameters that are separate to the Farquhar *et al.* photosynthesis model. Firstly, the prior allowed for the atmospheric CO_2 level (at the reference temperature $T^* = 11.5\,°C$) to range between $3.18 \times 10^{-4}\,kg\,m^{-3}$ and $5.04 \times 10^{-4}\,kg\,m^{-3}$, and the best-fit model's calibrated value is $5.00 \times 10^{-4}\,kg\,m^{-3}$. Atmospheric readings of CO_2 in the Southern Hemisphere (taken from the Cape Grim station operated by the Australian Bureau of Meteorology and CSIRO's Oceans and Atmosphere [1]) from a time close to the calibration data were recorded are 409.66 parts per million, which corresponds to approximately $6.48 \times 10^{-4}\,kg\,m^{-3}$ at an air density of $1.061\,06\,kg\,m^{-3}$ calculated [10] using the reference temperature, and Guyra's elevation and relative humidity in June. Although this is a significant discrepancy, concentrations reached by venting in the data do fall to around $5.00 \times 10^{-4}\,kg\,m^{-3}$ or slightly above, suggesting that the CO_2 readings in the dataset may have been taken using sensors that were not calibrated to Guyra's elevation of 1330 m above sea level (this was discussed as a possibility by the representative). The calibration is then compensating for this.

A calibrated parameter that is even more demonstrative is the surface area of the plants, which was allowed by the prior to vary over four orders of magnitude, between $5 \times 10^2\,m^2$ and $5 \times 10^6\,m^2$. The best model found using Bayesian calibration suggests a value of $A_{plants} = 3.698 \times 10^5\,m^2$. Using the surface area of the glasshouse $A_g \approx 5.05 \times 10^4\,m^2$, this corresponds to a leaf area index of $A_{plants}/A_g \approx 7.3$. This is larger than the range of two to four suggested by Costa's representative, but of the correct order of magnitude, despite the very large range of values permitted by the calibration process. The suggestion that dark respiration takes place at a significantly reduced rate at night ($\gamma_{night} \approx 0.203$) is surprising. Considering that plant material that does not receive light can still be active in dark respiration, and the model appears to be overestimating the photosynthetically active surface area but then downweighting the amount of dark respiration, it might be beneficial

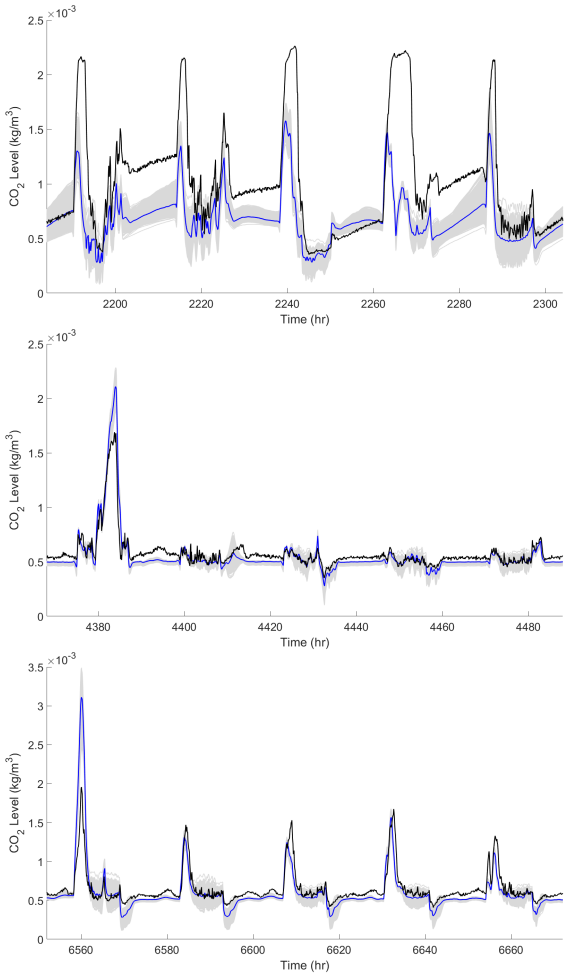


Figure 6: Testing the generalisability of the model ensemble by plotting their predictions for unseen time periods, selected to avoid cherry-picking. Each panel represents five days of glasshouse activity, from late September, late December, and late March, respectively from top to bottom. Although the fit is rather poor for the first time window selected, the model proves a mostly very good match to Costa's observations for CO_2 levels during the hotter months, despite not having seen any data from them during calibration.

in subsequent modelling to define separately the active and inactive surface areas.

3.4 Learnings

Using the model calibrated to the Guyra glasshouse tomatoes, their carbon uptake behaviours can be analysed to inform broad glasshouse operating procedure. Although the calibrated model is too long to easily reproduce here (model details are given in Appendix C, and best-fit parameters in Table 2), in Figures 7 and 8 we present a series of response curves and surfaces that show how the tomatoes' CO_2 uptake is affected by temperature, ambient CO_2 concentration, and temperature, as well as which of CO_2 or light is the limiting factor for photosynthesis rate. The most relevant data is for temperatures above 25 °C, given typical daily glasshouse operating temperatures. At such temperatures, the model predicts that photosynthesis is almost always light-limited, except in the case of high light and CO_2 concentrations lower than atmospheric. This suggests that the primary purpose of CO_2 injection should be to ensure that ambient CO_2 levels inside the glasshouse are not depleted during the day by photosynthesis, rather than spending CO_2 to chase more photosynthesis by increasing the ambient concentration far above the atmospheric baseline. Instead, given that the Guyra glasshouses depend on natural lighting, the model suggests that the more promising climate control for increased photosynthesis is to run the glasshouses a few degrees hotter, if feasible.

In Figure 9, we also consider how these conclusions extend beyond only the best-fit model, by using the ensemble of possible calibrations generated by our Bayesian inference approach. Figure 9a shows what proportion of the model ensemble predicts that CO_2 was the limiting factor for photosynthesis during the five days of glasshouse operation that were used for model calibration. Across the entire visualised time period, never more than 40% of models predicts CO_2 limitation, and where some models do, it is consistently in the middle of the day where the initial spike of injected CO_2 has been depleted.

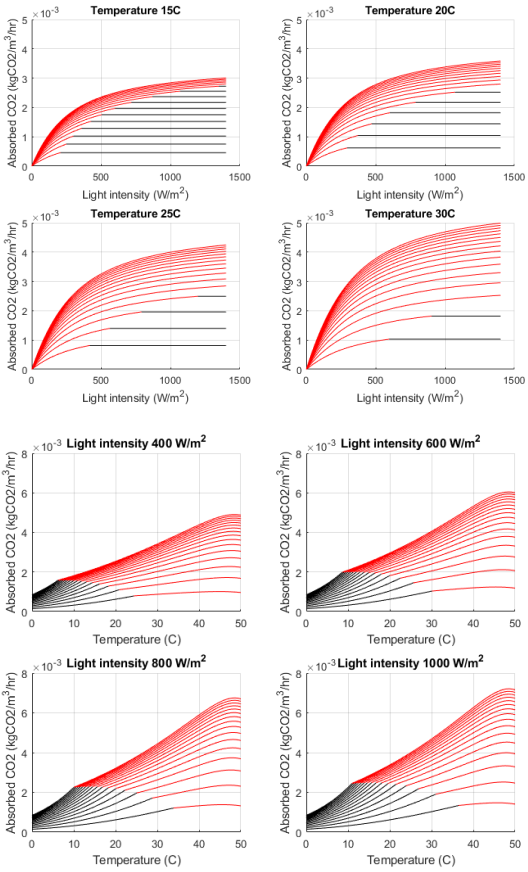


Figure 7: Response curves for the CO_2 uptake by plants, as predicted by the best-fit parameterisation of the Farquhar *et al.* model [12]. Shown is the CO_2 uptake rate versus light intensity at fixed temperatures (top four figures) and versus temperature at fixed light intensities (bottom four figures). The different curves from bottom to top in each figure correspond to glasshouse CO_2 concentrations of 0.3 g/m^3, increasing by 0.1 g/m^3 each line up to 1.8 g/m^3. Curve colour indicates whether photosynthesis is limited by CO_2 availability (black) or light availability (red).

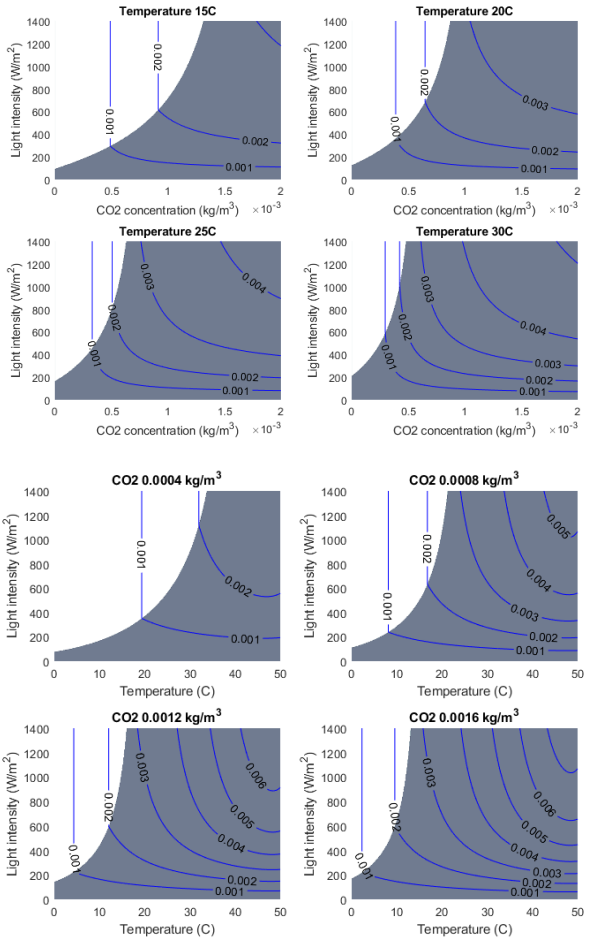


Figure 8: Contour plots of the CO_2 uptake by plants, as predicted by the best-fit parameterisation of the Farquhar *et al.* model [12]. CO_2 Uptake contours, labelled and separated by intervals of 10^{-3}kgCO2/m^3/hr are plotted against light intensity and CO_2 concentration at fixed temperatures (top four plots) and against light intensity and temperature at fixed CO_2 concentrations (bottom four plots). The light zones in these plots correspond to CO_2 limited regimes whilst the grey zones correspond to light limited regimes.

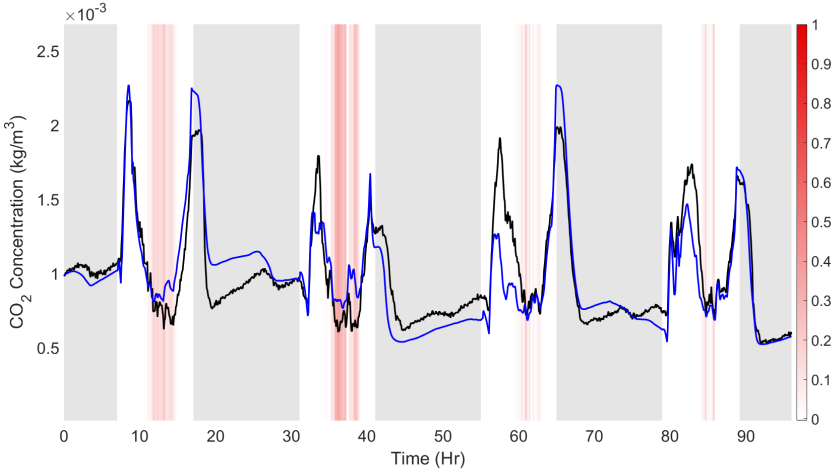


Figure 9: Graphical demonstration of the model ensemble's predictions of the factor that limits photosynthesis (light or CO_2) across the five days used for model calibration. The depth of red shading indicates the proportion of models that predict photosynthesis is limited by CO_2 at that moment in time. Nights are shaded in grey. In general, only a small proportion of the model ensemble predicts moments of CO_2 limitation, which consistently occur in the middle of the day.

Indeed, overlaying nighttime periods with CO_2 concentration suggest that the CO_2 injected to compensate for this depletion comes too late, often peaking at the onset of night. The same pattern in CO_2 dynamics was observed in the five days of data from late September (top panel in Figure 6), and a recommendation for the Guyra glasshouses is that regardless of whether or not the specific predictions of the model about light-limitation versus CO_2 -limitation are taken as gospel, it would be beneficial to ensure that CO_2 injection routines are switched off earlier to avoid large spikes of CO_2 at the end of the day where available light is decreased, and the injected CO_2 is simply lost to venting.

One caveat with the limitation-based view is that actually, as the Farquhar *et*

al. model describes (see Appendix C), an increased concentration of CO_2 can still slightly improve photosynthesis under the light-limited regime. The mechanism is that the relative concentrations of O_2 and CO_2 in the air around a plant affect the proportion of its Rubisco that is used for the photosynthetic Calvin cycle, as opposed to the wasteful photorespiratory cycle that takes in O_2 and releases CO_2 [12]. Given that our model is based upon the Farquhar *et al.* model and hence captures this effect, we extend our analysis beyond light-limited versus CO_2 -limited, and beyond the small windows of time used for model calibration and validation. To do so, we poll our model ensemble to predict how scaling the total amount of CO_2 injected into the glasshouse (but keeping the current injection strategy the same, for simplicity) affects the total carbon uptake by the plants across an entire year.

As detailed in Appendix C, altered levels of injection are tested by simulating the models in the ensemble across a year-long period and integrating the net CO_2 uptake term (photosynthesis minus photorespiration) across this period. Following the observations from Section 2 that suggest changing CO_2 injection levels has no significant effect on the temperature of the glasshouse, we use Costa's existing data for temperature in these simulations. We assume the same injection and venting strategy that is currently used, and so Costa's data for the rates of these two processes is taken as-is, except with a scaling applied to the former to represent the increased/decreased CO_2 injection being examined.

Figure 10 shows how the total carbon assimilation is affected by increased or decreased CO_2 injection, as predicted by the best-fit model and 100 members of the model ensemble selected at random. As might be expected, increasing CO_2 injection never harms total carbon gain, but does exhibit diminishing returns. If the glasshouse is operated with zero supplementation of CO_2 at all, the model ensemble predicts a mean relative decrease in carbon assimilation of 22.1% (best-fit model prediction: 24.2% reduction). If CO_2 supplementation is doubled (naively, without also adjusting CO_2 injection strategy), then the model ensemble predicts a mean relative gain in carbon assimilation of 11.2% (best-fit model prediction: 14.6% gain). These results suggest that CO_2

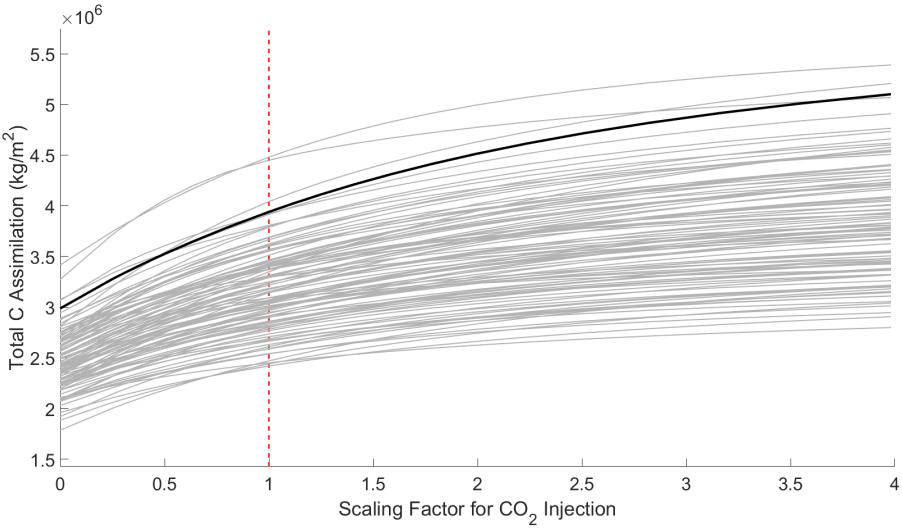


Figure 10: The effect of increased or decreased CO_2 injection (using the current strategy but with a scaling factor applied) on total photosynthesis achieved, as predicted by a hundred random members of the posterior ensemble (grey) and the best-fit model (black). The red dotted line indicates the current amount of CO_2 injection.

supplementation is indeed having a beneficial effect, but serves a secondary role in the glasshouse.

Whether or not further increasing CO_2 supply would be beneficial depends strongly on the cost of achieving this increased supply. In making such an evaluation, the curves in Figure 10 are likely best interpreted as a "worst-case" performance for an increased amount of CO_2 , owing to their calculation using an injection strategy that has not been adapted to the new amount of CO_2 being injected. Indeed, our observation above about how CO_2 is often still being injected late in the day and hence going wasted to venting suggests that CO_2 injection strategy is probably the better place to start from, before then re-evaluating potential photosynthetic gains from increasing the rate of

injection.

4 Modelling Tomato Yield

We have two approaches to conceptualising modelling tomato yield. The first we call the maturity model because the aim is to determine the yield at maturity (the mature tomatoes picked on a given day). This model involves determining fruit mass produced over the conditions from flowering to picking. The second model we call the stored value model because it takes the perspective that value is stored in the fruit at the moment CO_2 is converted to sugars. In this respect, the yield in the Stored value model is simply a measure of instantaneous CO_2 uptake which we can determine using the model calibrated in Section 3.

4.1 Maturity model

The final step of the model consists in combining the information obtained from temperature and CO_2 concentration to obtain an estimate of tomato yield. There are a number of very complex biophysical models for the growth and yield of tomato plants in the literature, most notably the TOMGROW model [18], and the Vanthoor model [37]. These models consider many of the biophysical processes that occur during the growth of a tomato plant, including both leaf and tomato growth, the generation of new tomatoes, and different stages in the growth of those tomatoes, and they perform well when fit to detailed experimental data either alone or combined (see e.g., [16, 2, 21]). However, the complexity of these models means that they have many parameters, model outputs that are not directly useful (and in a way that is difficult to reconcile with actual farming processes in the glasshouse, e.g., with how pruning is performed), and are ill-suited to modelling the generally steady-state nature of tomato yield over time in high-tech glasshouses. To better suit the way tomatoes are grown and harvested in this glasshouse (and to avoid overparameterization) we focus on modelling just the tomatoes that

will be harvested on a given day, over their history, rather than modelling the whole plant over its lifetime.

The Maturity model is a simple model that describes the growth of fruit mass m^j associated with fruit that flowers on day j. The time t is measured in discrete days and the yield at day $j + t$ where $t = 42$ days of growth is found through the daily addition of mass using the logistic growth recurrence relationship

$$\begin{cases} m^j(j+t+1) = m^j(j+t) + \beta(L,T,C)\, m^j(j+t)\left(1 - \dfrac{m^j(j+t)}{\tilde{K}}\right) & t \in [0,42] \\ m^j(j) = \tilde{m}_0, \end{cases}$$

where the function $\beta(L,T,C)$ is the growth rate of the fruit and is assumed to be related to photosynthesis and therefore to daily light L, temperature T and CO_2 levels C. Since the conditions L, T and C are indicative of daily averages, it does not make sense to use the finely calibrated model for CO_2 uptake presented in Section 3. Instead, two simple model forms for $\beta(L,T,C)$ are assumed. The first model assumes that photosynthesis is due only to daily light levels,

$$\beta_1(L,T,C) = \frac{\gamma L}{\alpha + L}.$$

The second model assumes a more comprehensive relationship between the various daily factors,

$$\beta_2(L,T,C) = \frac{\gamma LC}{(\alpha_L + L)(\alpha_C + C)}(-(T - T_0)^2 + c),$$

noting that the value of γ would be different to the above, and would require parameter estimation along with α_L, α_C, T_0 and c.

After fitting to available yield data (details in the Appendix), it is interesting that the two models provide very similar yield predictive distributions (both capturing the variation of yield data well). There is only a small discrepancy between the two models on the period between August and September. This

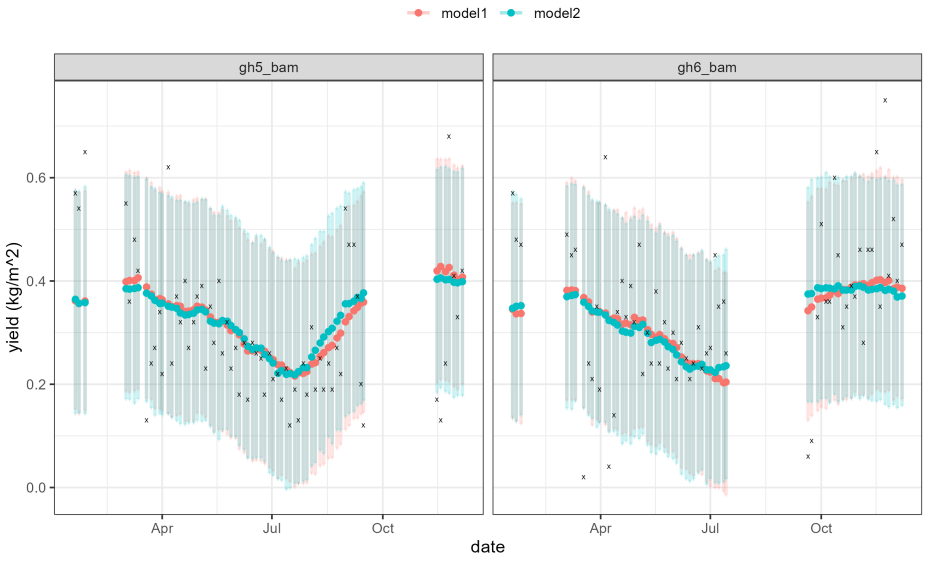


Figure 11: Posterior predictive distribution from the simple yield model with β_1 (model 1; light dependence only) and β_2 (model 2; light, CO_2 and temperature dependence) against observed yield (black crosses) for one variety of tomato across two greenhouses. Posterior median (coloured dots) and 90% credible intervals (errorbars) are shown. All observations of yield from December to December are shown where data are present and 42 days of climate observations are available.

could be taken as further evidence of the conclusion arrived at in Section 3 that light is the most important factor for plant carbon uptake, and hence yield.

It is important to observe that our model does not take into account the process of the flowering of the tomatoes, which is another factor which impacts yield. The expert practitioners who control glasshouse operations currently manipulate conditions within the glasshouse to ensure sufficient flowering based on observing the process in real time, so we assume here that this

aspect of the process is constant.

Furthermore, we assume a constant time period between flowering and harvesting of six weeks (42 days), which may not be accurate (particularly given that harvesting on each section of a glasshouse occurs twice per week). Similarly, the assumption that the number of tomatoes harvested per m^2 is constant and only the tomato's mass varies is not completely realistic. An improved model should therefore let the number of tomatoes vary according to the environmental conditions and to the harvest time, which depends on the growth stage of the tomato and the weekly picking schedule. This could be done using a structured population model of the kind analyzed in [25].

4.2 Stored value model

The premise of the stored value model is simple. Here we define the yield to be the instantaneous yield in gross tomato mass (both picked and unpicked). A metric for this yield is therefore the CO_2 uptake rate. That is, for each kg of CO_2 that is taken up by the crop a certain value is produced such that there is a some conversion factor r which converts CO_2 uptake in kg per hour into future revenue in dollars per hour. Likewise the CO_2 in the glasshouse is associated with a cost of production. That is, for every kg of CO_2 injected into the glasshouse, a cost is incurred such that it is possible to convert kg of CO_2 in the glasshouse to a value of c dollars per kilogram. Note that this cost per kilogram c is a net cost. For example, if the process that produces CO_2 also produces heat which otherwise has a known value then the net cost of CO_2 is reduced.

It is important to look at yield in terms of value because it is known that for the most part increasing CO_2 by injecting it into the glasshouse should almost certainly improve yield. This, however, is not the objective of the growers. The objective of the growers should be to ensure that profit is maximised. This is important because as we have shown in Section 3, increasing CO_2 delivers decreasing marginal yield but the cost of producing the CO_2 does not reduce at scale the same way.

In this model we shall make a simplifying assumption. We assume that CO_2 equilibrium concentrations, which are formed when CO_2 injection is balanced by net loss via the vents combined with net uptake by the crop, are reached rapidly. In this way, all the CO_2 that is produced at some moment in time and injected into the glasshouse is either lost out of the vents or is uptaken by the plants. We can therefore measure the instantaneous rate of profit (yield compensated for costs) generated by the glasshouse. If we denote p to be the profit generated by the glasshouse then $\dot{p} = r\dot{m}_{\text{uptake}} - c(\dot{m}_{\text{uptake}} + \dot{m}_{\text{vent}})$ is the rate profit is generated by the glasshouse. Under the simplifying assumption, then the rate of profit generation

$$\dot{P} = \Lambda \dot{m}_{\text{uptake}} - \dot{m}_{\text{vent}}, \tag{4}$$

using notation from Section 3 and setting $\dot{m}_{\text{vent}} = \dot{m}_{\text{vent-out}} - \dot{m}_{\text{vent-in}}$. Here we denote profits P in 'kg of CO_2'; the conversion to dollar profits is $p = cP$. The parameter $\Lambda = (r - c)/c$ is the 'value added' by the plant for converting CO_2 into sugars/tomatoes relative to the cost produced. We shall call the parameter Λ the 'plant mark-up'.

Our objective in deciding whether or not it is best to have an increase or decrease in CO_2 should come from a profits perspective. To this end, at each moment in time assess what CO_2 level will maximise $\dot{P}$. This is done by taking calibrated forms of $\dot{m}_{\text{uptake}}$ and $\dot{m}_{\text{vent}}$ generated by the model in Section 3. Therefore, the optimal CO_2 level from a profits perspective depends on the instantaneous solar radiation as well as temperature but also the parameter Λ which is determined by the cost of gas, operational costs, efficiency of CO_2 delivery system, the price of tomatoes etc. We assume that Λ is constant in time and optimise the CO_2 concentration at every 5 minute interval over 7 months of data (from December 6th to June 7th). Here we are able to determine three things; for a given plant mark-up Λ,

- What is the theoretical optimal profit that can be made over these 7 months?
- Is Costa group, on average, injecting too much or not enough CO_2? and

- How does this compare to the theoretical profits that Costa are realising at the moment? (How much improvement in efficiency can be generated through 'smart' CO_2 management?).

Whilst the results here are based on a simple model, they give estimates that may be used to justify further work. We also capped the optimal CO_2 concentrations at 0.002 kg per cubic meter at any moment in time.

To determine the total profit over the 7-month period, we simply aggregate profits over each five minute interval for which we have data. The factor Λ here is important. If Λ is small then there is a tight margin between CO_2 production costs and revenue to be made from converting this CO_2 into fruit. In this case, the optimal CO_2 that should be targetted is small since it is only profitable to utilise the high sensitivity of the plants to CO_2 at low CO_2 levels. On the other hand if Λ is large then even if the marginal return of photosythesis is low for high CO_2 concentrations, the revenue that the fruit generates more than compensates and so the optimal CO_2 levels are very high. It is uncertain to us what value Λ is on average.

Therefore for a wide range of values of Λ, we calculate the aggregate profits defined by Equation (4) over the 7-month period for the fully optimised CO_2 levels and divide this by the associated profits modelled from CO_2 levels measured from the glasshouse over the same period. Thus, if this ratio is 1.2 then this means that optimised CO_2 regulation could yield 20% greater profits. In Figure 12, we plot this theoretical profits gain versus the plant mark-up parameter Λ. It is interesting to see that this profit gain under an optimised system has a local minimum at around $\Lambda = 80$. One possible explanation for this is that $\Lambda = 80$ is already quite close to the range of the true values of Λ (the market value added to CO_2 by photosynthesis) and Costa are already operating their CO_2 injection in an attempt to optimise profits in this market (reducing the room for improvements by mathematical optimisation). Nevertheless, according to this model, even at its minimum, optimisation of CO_2 levels is likely to improve profits by at least 9% which is substantial.

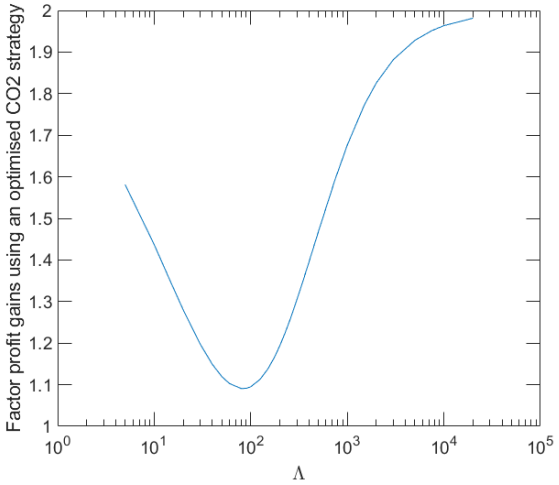


Figure 12: A plot of ratio of profits between optimised and measured CO_2 levels over the interval from December 6th to June 7th versus the plant mark-up parameter Λ. The parameter Λ describes the increase in value added by photosythesis and is determined by the market (and therefore may change significantly over time). This plot assumes that Λ is constant for the full 7-months.

Assuming that $\Lambda \approx 80$, we can easily determine if Costa should be injecting more CO_2. At each 5-minute interval over the 7-month sample period, in Figure 13 we show the difference between optimal glasshouse CO_2 and actual glasshouse CO_2. A simple histogram of the differences shows that there is no clear trend in under or over supply of CO_2 in aggregate. The average difference in optimal CO_2 and measured CO_2 levels is just 0.000025 kg per cubic meter of CO_2. That said, the histogram indicates that there are occasions where CO_2 could potentially be significantly increased for larger profits.

Of course, since choosing Λ to be at the minimum of increased profits we are choosing Λ such that we expect that current CO_2 management strategies are at their most efficient. Therefore, in Figure 14 the average increase in CO_2

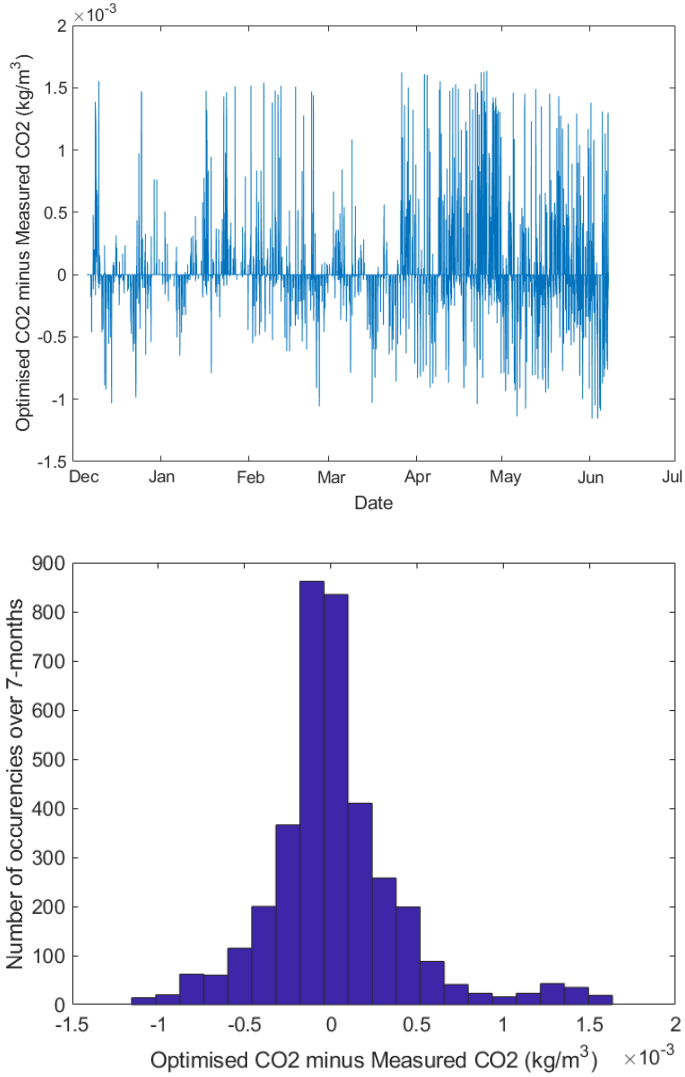


Figure 13: (Top) A plot of the difference in CO_2 as suggested by optimisation with that measured in the glasshouse over time. The optimisation was performed for profits with $\Lambda = 80$. (Bottom) A histogram of the frequencies of the differences plotted above.

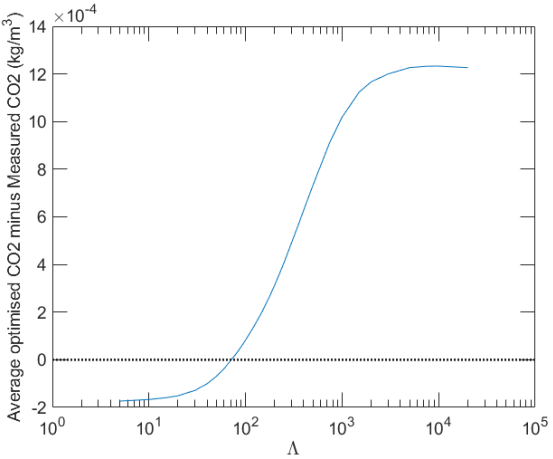


Figure 14: A plot of the average difference between optimised and measured CO_2 over the course of the interval from December 6th to June 7th as a function of Λ. For a given Λ, this plot indicates by how much CO_2 the glasshouse is wasting (if negative) or could use for better profits (if positive).

required over the current average CO_2 levels that would best optimise profits versus the parameter Λ to better show qualitatively how CO_2 should be increased or decreased in response to changing Λ (changing market conditions – for example, the cost of natural gas).

5 Conclusions

Production of commercial crops in greenhouses/glasshouses is an important contributor to food security, especially in regions where the availability of agricultural land is low [3]. Although this setting provides growers with protection against pests and a very high level of climate control, making best use of this control remains a topic of ongoing research, including for tomatoes [22]. There also exists a significant gap between the practical operation of glasshouses, and academic knowledge about how these operations

can be simulated and hence better understood and/or optimised (some reasons why are discussed in [14]). Using the MISG meeting as an opportunity, applied mathematicians spent an intensive week collaborating with a representative from such an operation to conduct research that stretches across this gap, and work towards answering specific questions of practical relevance to the commercial partner.

One matter of particular interest was the use of CO_2 supplementation to improve photosynthesis in the tomato plants. In particular, whether or not further supplementation might be expected to compromise glasshouse operating temperature, and the extent to which additional supplementation is predicted to increase yield (and especially profit). As has been detailed in this work, MISG subteams examined these questions from a number of directions, by modelling individual aspects of glasshouse operation in great detail (an approach recommended in [14]) and making direct use of operational data from Costa glasshouses for calibration. To respect uncertainty that might remain in model parameters following this calibration, a Bayesian inference approach was used.

Temperature within a glasshouse affects, for example, photosynthesis [12], plant development (including reproductive versus vegetative balance [17]) and conditions for glasshouse workers. In considering adjustments to operating strategy, including the possibility of increased CO_2 supplementation, it is therefore important to understand how different aspects of glasshouse operation affect its internal temperature. Using a physics-based model that achieved good model-data fit to glasshouse temperature data (Figures 1 and 2), we have determined that injected flue gas, and even the glasshouse's heating system have only minor effects on increasing temperature as compared to heat provided by the sun. Conductive losses of heat are also of minor importance. Instead, the bulk of heat loss in the glasshouse is via air exchange with the outside due to venting, and evaporative losses due to misting. The strength of their effects highlights the extra cost (in heating) incurred when these processes are used to control humidity, as opposed to temperature. The model's prediction that injected flue gas has a minor effect on temperature implies

freedom in CO_2 supplementation strategy, and supports the assumption of unchanged temperature where alternative strategies have been examined in this work.

Multiple models have been used to analyse glasshouse fruit production. Modelling yield allows Costa's data for the meaningful output of their glasshouses to be used directly and explicit representation of the time from flowering to harvest. Day-level data for CO_2 and temperature were found to make only slight difference to prediction of yield, suggesting that the current protocols for CO_2 and temperature control in the glasshouse are already good, such that light is the limiting factor for fruit development. This conclusion was also attained separately by a different subteam modelling CO_2 dynamics in the glasshouse, calibrated to experimentally obtained photosynthesis curves and fine-resolution CO_2 concentration data from within the glasshouse. As the Farquhar *et al.* [12] model for photosynthesis in C3 plants describes separately the potential rates of photosynthesis allowed by the levels of ambient CO_2 and incident light, it offers explicit prediction of which of these two factors restricts carbon uptake for growth. Using Bayesian calibration to generate an ensemble of possible parameter values for the model, we find that the great majority of these predict that light limits photosynthesis at almost all times. In some time periods of the data, CO_2 concentration shows significant dips around the middle of the day and this is where some models predict CO_2 -limited photosynthesis and further supplementation would be beneficial. The prominent recommendation for the industry partner is improved CO_2 control across the day so that the concentration better tracks light levels, which would also avoid the large amounts of CO_2 injected later in the day to compensate for midday dips, that does little to aid photosynthesis as light wanes, and is mostly lost to the evening venting.

The calibrated CO_2 model also informs questions of profit/loss. As CO_2 supplementation can still slightly increase net carbon uptake even under light-limited conditions, we calculated model predictions for total CO_2 uptake (as a proxy for growth/yield) when different amounts of CO_2 are supplied. Without adjustment of CO_2 injection timing/strategy, doubling CO_2 or eliminating

injection entirely only amounts to an expected 20% gain or loss, respectively, although how this affects profit depends upon current sales prices and the cost associated with CO_2 injection. Alternatively, optimal CO_2 injection strategies can be found by assigning a parameter to describe what "mark up" is achieved when injected CO_2 (with some associated cost) is taken up by plants (increasing eventual yield and revenue). Using the calibrated photosynthesis model and some simplifying assumptions, net "profit" (in the sense of revenue minus injection cost) gains using optimal injection strategies varied from 10% to 100%. Other operating costs, which are likely unaffected in any significant way by changes to CO_2 injection strategy, should be taken into account in interpreting these figures. Given the observation that light appears to be the most important limiting factor for glasshouse output, of particular note are the cases where CO_2 injection costs are not insignificant and the mark-up parameter takes on a relatively smaller value. In this case, the optimal strategy involves less, and more careful injection of CO_2 to target photosynthesis improvements only in conditions of very strong light.

The bulk of modelling work and analysis discussed here was carried out during the intensive week of in-person collaboration with the industry partner. As a result, there remains definite scope for expansion of the models considered, and further analysis. One important extension is the incorporation of humidity into the photosynthesis model, so that the interplay of the different climatic factors can be properly considered in informing operative strategy. Optimal control techniques [36] could also be applied to the CO_2 or yield models directly in order to determine optimum CO_2 injection protocols. The spatial model of heat and air flow that was constructed but not finalised during the workshop would also serve as an important tool to test the assumptions underlying the models we have used here, which all lean upon a well-mixed assumption.

Overall, this study demonstrates the usefulness of such interdisciplinary workshops and the gap between industry practice and academic knowledge. For example, the data-calibrated response curves and surfaces for photosynthesis (Figures 7 and 8) were suggested by Costa's representative as an important

output from the workshop for growers, even though such curves are extremely common in the scientific literature [8]. The access to extensive practical data provided by programs such as MISG has here ensured mathematical representations of the glasshouse system agree (to an extent) with actual operating parameters. It has also exposed the deficiencies in these data and/or the models constructed to try to fit to them, hence providing very specific directions of further inquiry that are relevant to industry.

References

[1] Australian Bureau of Meteorology and CSIRO Oceans & Atmosphere. Cape grim greenhouse gas data. https://capegrim.csiro.au/.

[2] L Bacci, P Battista, and B Rapi. Evaluation and adaptation of tomgro model to italian tomato protected crops. *New Zealand Journal of Crop and Horticultural Science*, 40(2):115–126, 2012.

[3] Sven Batke, Nathan Thomas, Nathalie Key, and Phil Morley. Protected and productive: How greenhouses should deliver uk food security. *Plants People Planet*, 8:626–638, 2026.

[4] Jorge A. Bolaños and Theodore C. Hsiao. Photosynthetic and respiratory characterization of field grown tomato. *Photosynth. Res.*, 28:21–32, 1991.

[5] James A. Bunce. Effects of humidity on photosynthesis. *J. Exp. Bot.*, 35:1245–1251, 1984.

[6] Amanda P. Cavanagh, Rebecca Slattery, and David S. Kubien. Temperature-induced changes in Arabidopsis Rubisco activity and isoform expression. *J. Exp. Bot.*, 74:651–663, 2022.

[7] Shuailiang Chen, Aolong Liu, Fei Tang, Pei Hou, Yanli Lu, and Pei Yuan. A review of environmental control strategies and models for modern agricultural greenhouses. *Sensors (Basel)*, 25:1388, 2025.

[8] Eun-Young Choi, Young-Ae Jeong, Seung-Hyun An, Dong-Cheol Jang, Dae-Hyun Kim, Dong-Soo Lee, Jin-Kyung Kwon, and Young-Hoe Woo. Comparison of measured and calculated carboxylation rate, electron transfer rate and photosynthesis rate response to different light intensity and leaf temperature in semi-closed greenhouse with carbon dioxide fertilization for tomato cultivation. *Journal of Bio-Environment Control*, 30(4):401–409, 2021.

[9] Michael Clerx, Martin Robinson, Ben Lambert, Chon Lok Lei, Sanmitra Ghosh, Gary R Mirams, and David J Gavaghan. Probabilistic inference on noisy time series (pints). *arXiv preprint arXiv:1812.07388*, 2018.

[10] Dominik Czernia and Bogna Szyk. Air density calculator. https://www.omnicalculator.com/physics/air-density.

[11] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Stat. Methodol.*, 68:411–436, 2006.

[12] G. D. Farquhar, S. von Caemmerer, and J. A. Berry. A biochemical model of photosynthetic co_2 assimilation in leaves of c_3 species. *Planta*, 149:78–90, 1980.

[13] Andrew M. Fraser and Harry L. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33:1134–1140, 1986.

[14] Christian Gary, James W. Jones, and Marc Tchamitchian. Crop modelling in horticulture: state of the art. *Sci. Hortic.*, 74:3–20, 1998.

[15] A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient Metropolis jumping rules. *Bayesian Statistics*, 5:599–607, 1996.

[16] Liyun Gong, Miao Yu, Shouyong Jiang, Vassilis Cutsuridis, Stefanos Kollias, and Simon Pearson. Studies of evolutionary algorithms for the reduced tomgro model calibration for modelling tomato yields. *Smart Agricultural Technology*, 1:100011, 2021.

[17] Maria José Gonzalo, Yi-Cheng Li, Kai-Yi Chen, David Gil, Teresa Montoro, Inmaculada Nájera, Carlos Baixauli, Antonio Granell, and Antonio José Monforte. Genetic control of reproductive traits in tomatoes under high temperature. *Front. Plant Sci.*, 11:326, 2020.

[18] James W Jones, Ehud Dayan, LH Allen, Herman Van Keulen, and Hugo Challa. A dynamic tomato growth and yield model (tomgro). *Transactions of the ASAE*, 34(2):663–0672, 1991.

[19] Matthew B. Kennel and Henry D. I. Abarbanel. False neighbors and false strands: A reliable minimum embedding dimension algorithm. *Physical Review E*, 66:026209, 2002.

[20] René LaCroix and Robert Kok. Simulation-based control of enclosed ecosystems - A case study: Determination of greenhouse heating setpoints. *Can. Agric. Eng.*, 41:175–183, 1999.

[21] Dingyi Lin, Ruihua Wei, and Lihong Xu. An integrated yield prediction model for greenhouse tomato. *Agronomy*, 9(12):873, 2019.

[22] Stef C. Maree, Pinglin Zhang, Bart M. van Marrewijk, Feije de Zwart, Monique Bijlaard, and Silke Hemming. Autonomous greenhouse cultivation of dwarf tomato: Performance evaluation of intelligent algorithms for multiple-sensor feedback. *Sensors*, 25:4321, 2025.

[23] Krishna Nemali. History of controlled environment horticulture: Greenhouses. *HortScience*, 57:239–246, 2022.

[24] Donald J. Nevins. Sugars: the origin in photosynthesis and subsequent biological interconversions. *Am. J. Clin. Nutr.*, 61:915S–921S, 1995.

[25] Benoît Perthame. *Transport equations in biology*. Springer Science & Business Media, 2006.

[26] F Rodríguez. Modelado y control jerárquico de crecimiento de cultivos en invernadero. *Universidad de Almería. España*, 2002.

[27] Zoran Savović.

[28] Stan Development Team. RStan: the R interface to Stan. R package version 2.26.13.

[29] Stan Development Team. Stan modeling language users guide and reference manual. Version 2.31.

[30] Yuanping Su and Lihong Xu. Setpoint decision support strategy and adaptive hybrid control of greenhouse climate: A simulation study. *Intl. J. Control Automat. Syst.*, 20:3740–3758, 2022.

[31] Floris Takens. Detecting strange attractors in turbulence. In David Rand and Lai-Sang Young, editors, *Dynamical Systems and Turbulence, Warwick 1980*, pages 366–381, Berlin, Heidelberg, 1981. Springer Berlin Heidelberg.

[32] Frank Tap. *Economics-based optimal control of greenhouse tomato crop production.* Wageningen University and Research, 2000.

[33] Marc Tchamitchian, Roger Martin-Clouaire, Jacques Lagier, Benoît Jeannequin, and Siméon Mercier. SERRISTE: A daily set point determination software for glasshouse tomato production. *Comput. Electron. Agric.*, 50:25–47, 2006.

[34] Engineering Toolbox. International standard atmosphere, 2005.

[35] Anke van der Ploeg and Ep Heuvelink. Influence of sub-optimal temperature on tomato growth and yield: a review. *J. Hortic. Sci. Biotech.*, 80:652–659, 2005.

[36] Gerrit van Straten, Gerard van Willigenberg, Eldert van Henten, and Rachel van Ooteghem. *Optimal Control of Greenhouse Cultivation.* CRC Press, Boca Raton, FL, 2010.

[37] BHE Vanthoor, PHB De Visser, Cecilia Stanghellini, and Eldert Jan Van Henten. A methodology for model-based greenhouse design: Part 2, description and validation of a tomato yield model. *Biosystems engineering*, 110(4):378–395, 2011.

[38] Sylvan H. Wittwer and William M. Robb. Carbon dioxide enrichment of greenhouse atmospheres for food crop production. *Economic Botany*, 18:34–56, 1964.

[39] Takenobu Yamasaki, Tomokazu Yamakawa, Yoshihiro Yamane, Hiroyuki Koike, Kazuhiko Satoh, and Sakae Katoh. Temperature acclimiation of photosynthesis and related changes in photosystem II electron transport in winter wheat. *Plant Physiol.*, 128:1087–1097, 2002.

[40] Xuemei Yu, Luqi Niu, Yuhui Zhang, Zijian Xu, Junwei Zhang, Shunhui Zhang, and Jianming Li. Vapour pressure deficit affects crop water productivity, yield, and quality in tomatoes. *Agric. Water Manag.*, 299:108879, 2024.

A Glasshouse Data

To effectively and usefully model the glasshouse system (both its climate and yield), we must have a thorough understanding of and strategically gather relevant data for analysis. The main reason for this is so that the models we produce operate on a realistic scale, can incorporate data collected by the glasshouse, and can be validated against their observed outcomes. To this end, we must understand what data are collected, how frequently, in what form, and how to reconcile these data so that they are comparable and can be incorporated into models.

We begin by summarising the broad categories of data that are routinely collected as part of glasshouse operations:

- external weather conditions, which are measured at a small weather station outside the glasshouse, and are completely outside the control of the operators;
- measured internal atmospheric conditions, which are measured at a limited number of (fixed) locations within each glasshouse. Glasshouse settings are manipulated in an attempt to control these measured internal conditions within ranges set by the expert glasshouse operators;
- glasshouse settings (and associated calculated values) which are the levers used to control measured internal atmospheric conditions, such as the opening of vents, and the pumping of CO2 and heat;
- yield measurements.

All measurements are collected through bespoke internal glasshouse software at high resolution (i.e., 5 minute intervals, except yield data). These data are reported on standard and interpretable scales for each variable; however these scales are often quite different and not directly comparable. For example, carbon dioxide levels are reported using the well-known units “parts per million" (i.e., mg/L), whereas carbon dioxide pumped into the glasshouse is reported in terms of mass per unit of glasshouse surface area per unit time (

Kg/m^2/hr). Therefore, to coherently use both these quantities in a model we must first translate so they are in the same units (in this case, $\mathsf{Kg/m^3}$ and $\mathsf{Kg/m^3/hr}$). Specific details follow in the sections below.

Variable	Units	min	max
Actual CO2 input	kg CO2/ha/hr	0	193.92
CO2 G1s1	ppm	137	1399.11
Glasshouse Temperature G1s1	°C	12.1	33.02
Radiation	W/m2	0	1295.28
Radiation CumSum	J/cm2	0	3426
Actual Water Pipe Temperature G1s1	°C	18.6	67
Atmosphere Temp	°C	-2.42	28.04
Actual Mist Input G1s1	% of 585 ml/hr/m2	0	99
Absolute humidity G1s1	g/m3	8.21	28.41
Relative Humidity G1s1	%	62.12	95
Absolute Humidity Atmosphere	g/m3	0	16.79
Humidity Deficeit Atmosphere	g/m3	0	24.28
Relative Humidity Atmosphere	%	0	100
Lee side open G1s1 Control 1	%	0	90
Lee side open G1s1 Control 2	%	0	90
Wind side open G1s1 Control 1	%	0	72.8
Wind side open G1s1 Control 2	%	0	70
Curtain G1s1	%	0	100
Air Exchange G1s1 Control 2	m3/(hr.m2)	0	145.04
Wind Speed	m/s	0	10.5
Wind direction degrees	°	0.05	359

Table 3: Variable quantities used.

A.1 External weather conditions

External weather conditions are a critical input to both glasshouse operations and all modelling of the glasshouse system. The glasshouse is located where it is, in Guyra, New South Wales, because of external weather conditions: Guyra is a relatively high altitude, inland location and as such does not reach the high-temperature extremes of most of Australia; and it is further North

than other locations that have similarly mild temperatures, resulting in more consistent year-round sunlight.

The external conditions are measured at a weather station external to the glasshouses, with measurements that are critical to this work including:

- **Outside Temperature**, measured in degrees Celcius;
- **Outside Relative Humidity**, as a percentage;
- **Outside Solar Radiation**, measured in Watts per square metre;
- **Wind speed**, in km per hour; and,
- **Wind direction**, in degrees (from North).

Additional measurements that were not used in this study include indicators for rainfall, frost, and storms. The internal glasshouse sensors also calculate the cumulative solar radiation for each day (in Joules per square metre).

While Guyra has relatively stable weather conditions compared to many places, there is still substantial variability over time, in all variables. As one might expect, temperature and solar radiation are substantially lower in winter, and wind and humidity vary substantially within and across days (see Figure 15 for indicative examples of how these data varied over 3 days in June and December).

A.2 Measured internal atmospheric conditions

In contrast to the external conditions, which are uncontrollable, the internal atmospheric conditions can be controlled, and indeed the glasshouse operators apply a substantial amount of effort and resources towards this goal, with the aim of producing conditions that optimise tomato growth and thus produce maximum yield.

For this work, the relevant measured internal atmospheric conditions include:

- **Internal glasshouse temperature**, measured in degrees Celsius;

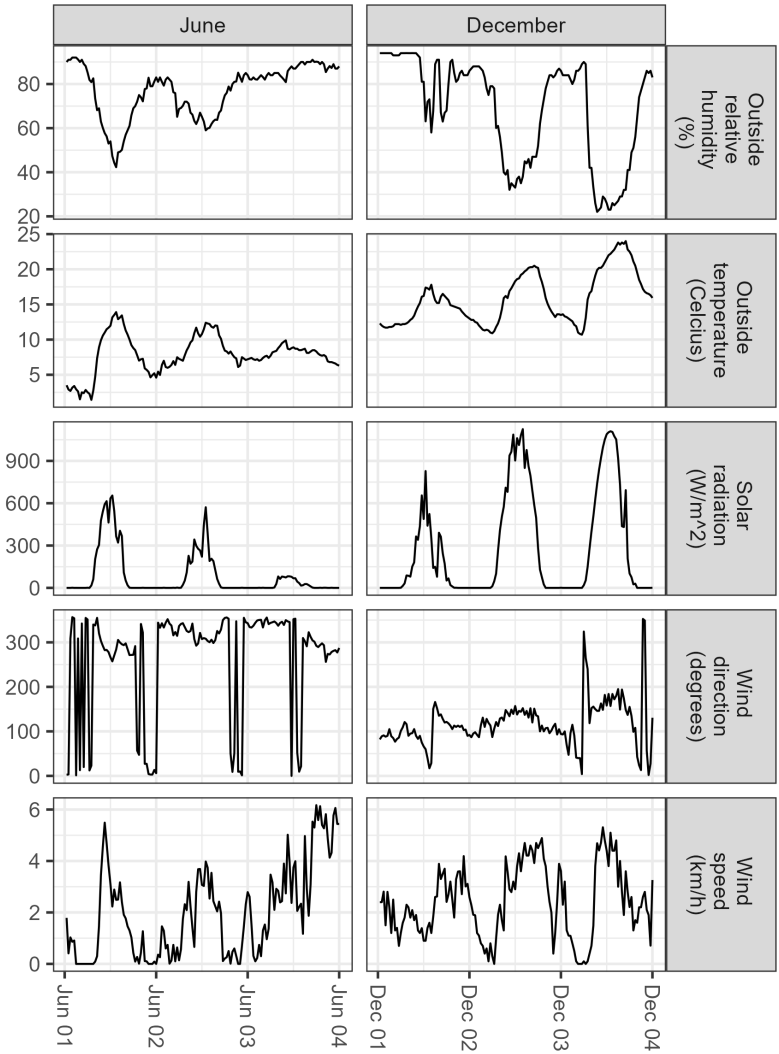


Figure 15: Example time-series of key external weather conditions measured at the glasshouse for the first three days of June (i.e., winter conditions), and December (summer conditions).

- **Internal glasshouse relative humidity**, as a percentage; and,
- **Internal glasshouse Carbon Dioxide concentration**, measured in parts per million.

Parts per million is a challenging unit to work with in models given that most other quantities are in more standard units; for convenience, we convert the internal glasshouse CO_2 concentration to units of Kilograms per cubic metre. One further difficultly with CO2 measurements is that the sensors must be calibrated, and it appears that these sensors within the glasshouse were calibrated at sea level. However, as Guyra has an elevation of c. 1,330m above sea level, these sensors do not take into account the necessary variations in pressure, and thus give readings approximately 15% lower than expected.

The closed nature of the glasshouse system combined with efforts to control the internal conditions lead the internal air temperature to be more uniform over the year than the external temperature (comparing Figure 16 to Figure 15). There is a complicated interplay between these internal measurements, in particular temperature and CO2 concentration, as both are impacted by exposure to external conditions (opening the vents; see the following subsection for details). We observe substantially higher internal CO2 concentrations in winter months, as in these months the vents are more likely to be closed to retain heat, and the energy curtains are more likely to be deployed (Figure 17).

A.2.1 Critical assumptions: mixing of internal glasshouse conditions

Measurements of the internal conditions are taken at small weather stations that are positioned at the head height of the growing vines in two locations between rows of tomato vines within each glasshouse. Thus, after the initial six week growing phase the height of the weather stations remains fixed. Because these stations are located at approximately the same height and away from the edges of the glasshouse, there is insufficient resolution in the data to produce models (that can be validated) which consider spatial variation in the internal glasshouse atmosphere. For that reason, we consider these measurements to

be entirely representative of the internal glasshouse atmosphere, effectively assuming that there is perfect and instantaneous mixing of internal conditions, with no variation through space (and, in particular, no difference between the bottom and the top of the glasshouse). It should also be noted that the growers have a strong desire to maintain a uniform atmosphere throughout the glasshouse and they adjust system parameters based on an assumption of uniformity.

This assumption of perfect, instantaneous mixing is not realistic, particularly given the mass of CO2, and that (warm) flue gas containing significant quantities of CO2 is pumped in to the glasshouse at the bottom and mixing with the outside air occurs through vents at the top of the glasshouse. Similarly, glasshouse heating occurs via hot water pumped through pipes at the bottom of the glasshouse, and loss to the external atmosphere will occur through the top and sides of the glasshouse. However, the glasshouse does have high-throughput internal fans and will ensure that mixing is faster than solely diffusion. Moreover, a model that includes space within the glasshouse would be catastrophically more complex than one that assumes uniform conditions, and we just do not have the data to validate such a complex model.

The thermal diffusivity of air is ≈ 0.07813 m^2/hr and the diffusivity constant for excess CO2 in air is a similar ≈ 0.0576 m^2/hr (both at 20°C and atmospheric pressure). With these values, using a one dimensional diffusion model in a 6m tall column, a fixed source temperature (or CO2 concentration) at the bottom of the glasshouse only raises the mean value by roughly 1/4 of the difference between the initial source and atmospheric values over 24 hours. Adding even a very small amount of advection in the positive vertical direction, e.g. to account for some buoyancy (due to temperature of injection) or motion induced by the fans, drastically changes this. Even 3.6 m/hr of vertical air movement (i.e. a very slow 1mm/s) leads to a near uniform distribution (equal to the source) in 2 hours. As such, it is clear that advection/convection are the main drivers of uniformity in the glasshouse, and the above estimates suggest that internal fans mix both temperature and CO2 throughout the glasshouse on a relatively short timescale.

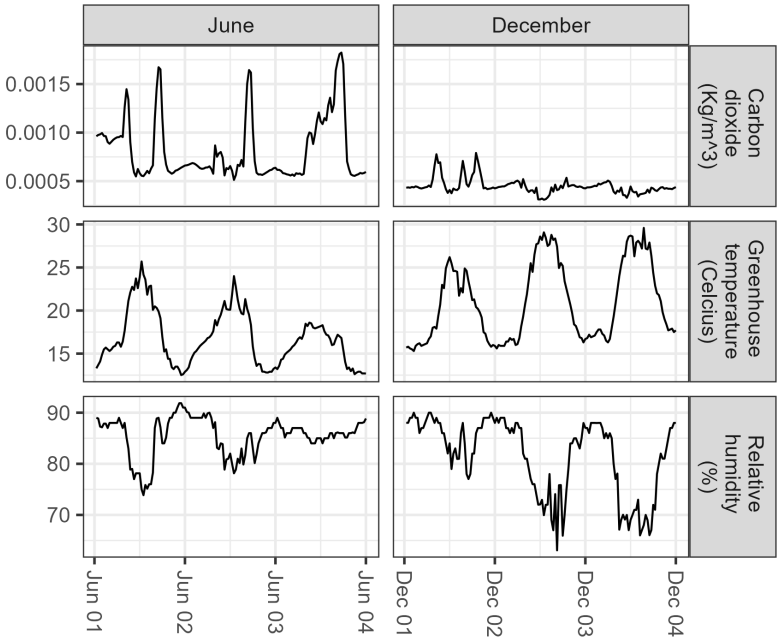


Figure 16: Example time-series of measured internal atmospheric conditions measured in one glasshouse for the first three days of June (i.e., winter conditions), and December (summer conditions).

A.3 Glasshouse settings

The glasshouse we model in this work is a modern system with a range of tools available to the expert operators, that they use to control the internal glasshouse atmosphere. A thorough understanding of the key inputs is necessary both for accurate modelling of the system (so that models can be used to provide advice around future glasshouse operations), and for validating models against historical data for which the operational conditions are known.

Before detailing the specific inputs, we emphasise that glasshouse operations

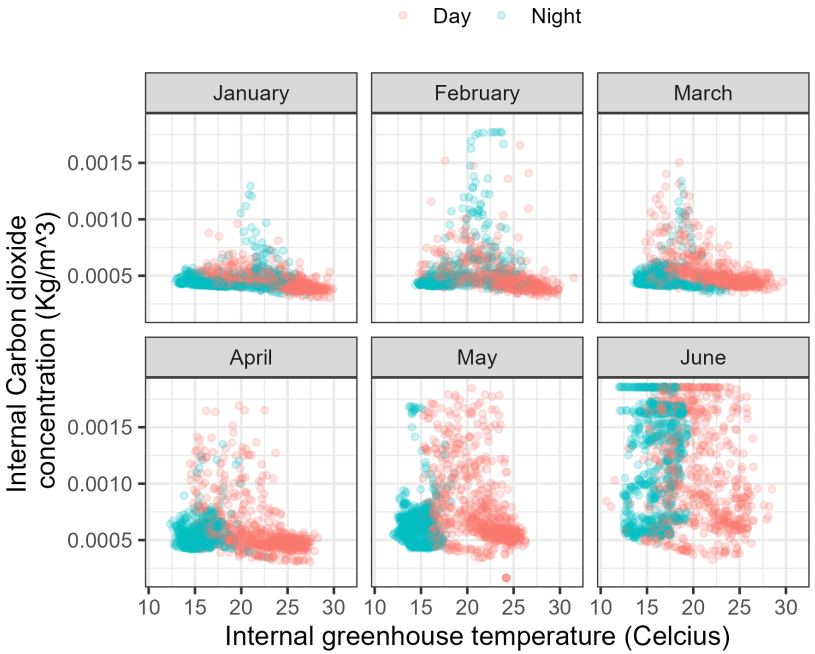


Figure 17: Relationship between glasshouse (internal) temperature and Carbon dioxide concentration within one glasshouse across January–June. This glasshouse was refreshed (all plants removed/re-planted) in July, so the displayed period demonstrates the standard operations of the glasshouse while the tomato plants are mature.

are a complex system with competing priorities. glasshouse settings are not managed with the goal of maintaining perfectly constant internal conditions across a year. Instead, they are managed in order to provide ideal conditions for growth (given the operators expert understanding of how this occurs) while minimising cost. Further, in some cases operational decisions are made to ensure consistency, e.g., glasshouse heating (by pumping hot water) is typically not turned on and off in short succession, rather it is operated during periods where it should be maintained across a number of hours. Many settings are controlled automatically by bespoke glasshouse operating software, so rather than manually controlling every specific condition, operators specify a target range for a measured quantity (temperature or CO2 concentration) and the system automatically adjusts controls in an attempt to achieve that target. However, as this work seeks to explicitly model the physical processes taking place in the glasshouse system, we describe and use the specific glasshouse system controls rather than operator specified target outcomes.

Key control inputs for this modelling work (Figure 18) are:

- Heating by pumping hot water through pipes in the glasshouse, with the water temperature measured in degrees Celcius. This contributes to modelling of the internal temperature.
- Misting, whereby water mist is sprayed in the glasshouse incrementally to control temperature and humidity, measured in $\mathrm{mL/hr/m^2}$ and converted to $\mathrm{Kg/hr/m^3}$.
- The energy curtain, which can be extended across the top of the glasshouse (its operation is measured as a percentage, but in practice it is almost always either fully extended or fully closed), in order to retain heat.
- The operation of vents at the top of the glasshouse, which are identified as wind-side or lee-side (this changes depending on the current wind direction), and are measured as a percentage. The vents and the energy curtain contribute to air exchange with the external atmosphere, which

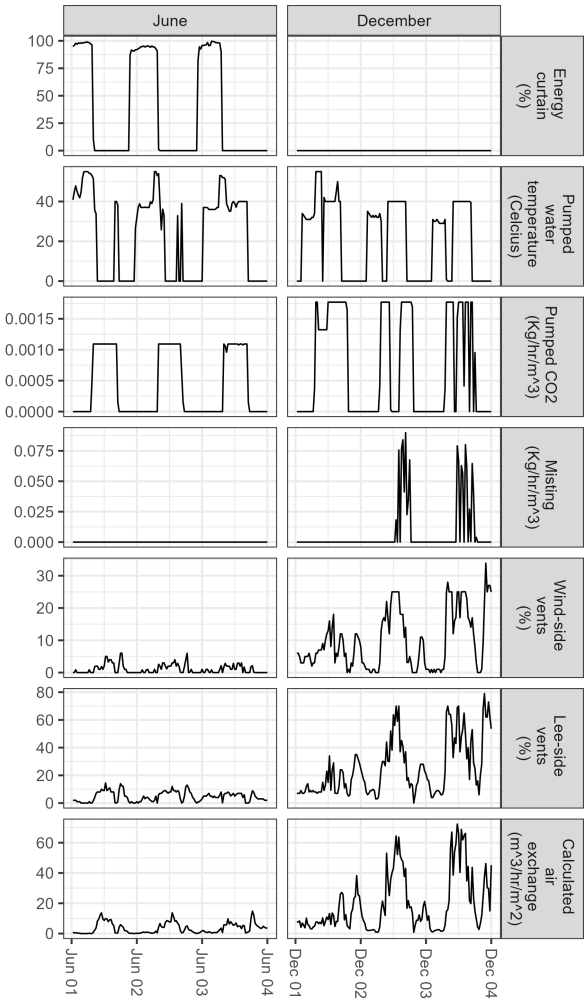


Figure 18: Example time-series of key glasshouse control settings in one glasshouse for the first three days of June (i.e., winter conditions), and December (summer conditions). Note that the calculated air exchange rate is a function of the wind-side and lee-side vent openings (and wind speed).

is critical to both temperature and CO2 models and detailed further in section A.3.1 below.

- Pumping CO2 directly into the glasshouse, measured in Kg/HA/hr which we convert to $Kg/hr/m^3$. This is done to maximise the value of incoming solar radiation as the tomatoes grow when they have access to sufficient sunlight and CO2. This impacts modelling of CO2 in the glasshouse, but also potentially has an impact on temperature, as the pumped CO2 may not be the same temperature as the air in the glasshouse.

A.3.1 Critical assumption: air exchange calculation

An additional critical assumption we make in this work is in accepting as written the glasshouse manufacturer's calculation of the air exchange ratio, as a function of vent opening, wind speed and the status of the energy curtain. This air exchange ratio is a key ingredient in the modelling of both temperature and CO2, as it governs the amount of interchange with the outside air. We do not know how the parameters in this calculation were estimated, nor the physical rationale for its form. The calculation (Figure 19) operates by:

- Total calculated air exchange is the sum of three components: leek, lee-side air exchange, and wind-side air exchange. A multiplicative correction (10%) is also applied.
- Leek is predicted by a linear regression, with wind speed as the only predictor.
- Lee-side air exchange is the product of wind speed, a lee-side wind factor, and a lee-side conduction factor.
- Wind-side air exchange is the product of wind speed, a wind-side wind factor, and a wind-side conduction factor.
- Both lee-side and wind-side wind factors are, in effect, dependent only

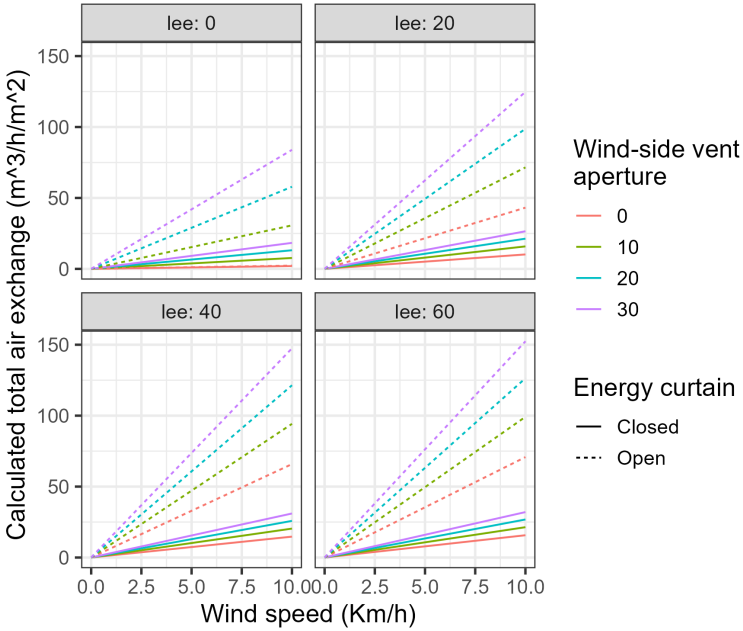


Figure 19: Visualisation of the calculation of air exchange rate (provided by the glasshouse manufacturer) based on wind speed, energy curtain usage, and the wind- and lee-side vent apertures.

on if the energy curtain is open or closed.

- The lee-side conduction factor is a cubic function of the lee-side vent opening aperture (as a percent).
- The wind-side conduction factor is a quadratic function of the wind-side vent opening aperture (as a percent).

Based on actual glasshouse operation, the majority of air exchange rates were relatively low: the median air exchange rate (from December to June in an operational glasshouse) was 7.41 $\mathrm{m}^3/\mathrm{h}/\mathrm{m}^2$, with 95% of observations in that time period having air exchange rate below 40.31 $\mathrm{m}^3/\mathrm{h}/\mathrm{m}^2$. However,

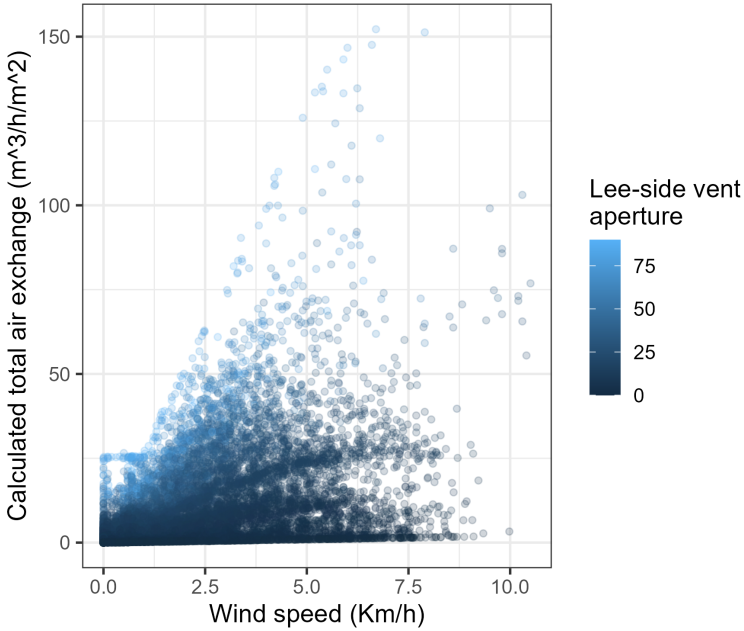


Figure 20: Observed wind speeds and associated calculated air exchange rates From December to June the following year in an operational glasshouse.

in some instances when there were higher wind speeds and open vents the exchange rate reached above 150 $\mathrm{m}^3/\mathrm{h}/\mathrm{m}^2$ (Figure 20).

A.4 Yield measurements

The glasshouse (overall) is divided into a number of smaller, discrete glasshouses, each of which is has independent internal conditions and controls. Tomatoes in each glasshouse are grown for approximately 48 weeks of the year, with schedules rotated to maintain constant production (i.e., only one glasshouse will be in the early growth stage at a given time). Within each glasshouse, multiple species of tomato may be planted. Harvesting is scheduled so that each plant is harvested twice a week, removing those tomatoes that are ripe.

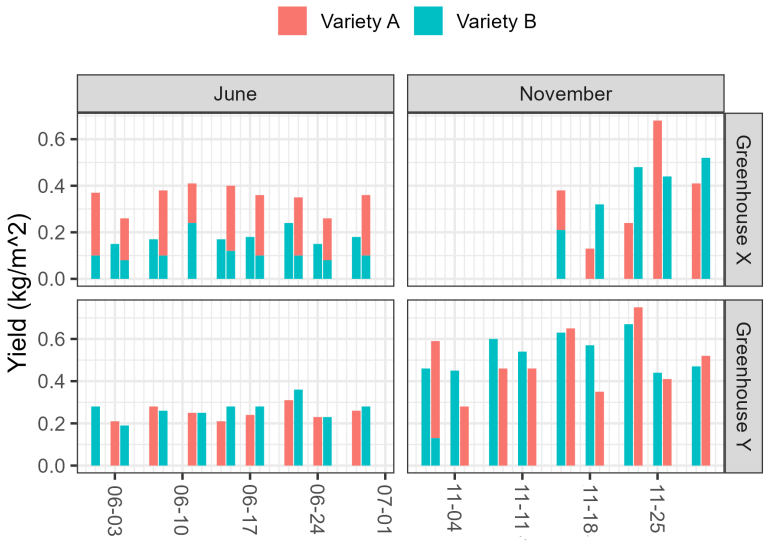


Figure 21: Example daily records of tomato yield across two glasshouses for June and November. Two breeds of tomato were grown in these glasshouses.

Yield, then, is reported in Kg of tomatoes per metre squared (of glasshouse area) in discrete, daily observations, with many days having no yield in a given glasshouse because there was no harvesting (Figure 21). The yield on a given day is impacted by the conditions in the glasshouse over the past 6-8 weeks, and also by the harvesting process itself. As such, yield is a quantity that is very different in form to the continuous measurements of atmospheric conditions (both internal and external) and glasshouse operations, and thus requires a different modelling approach.

A critical point is that we do not have data on harvesting effort, as such it is possible that a difference in yield between two days is driven by differences in harvesting effort rather than differences in tomato growth (caused by the conditions in the glasshouse over the lifetime of the tomatoes being harvested). Therefore, we must necessarily assume that harvesting effort is constant across

days (or, at least, across weeks when yield over weeks is smoothed).

A.5 Additional data considerations

When reading this modelling work, it is necessary to also keep in mind that:

- All measurements come from sensors that may be inaccurate or suffer from technical failures, and may not fully capture every possible driver of glasshouse conditions. One factor on which data are not available is the movements and operations of workers within the glasshouse, which may have impacts on the conditions either directly (i.e., due to body heat) or indirectly (air exchange with the outside environment when someone enters or leaves the glasshouse).
- We focus this modelling on the steady state operations of the glasshouse, when the tomato plants have achieved their operational length and are being regularly harvested. At other times conditions may be managed differently.

B Temperature Model Details

B.1 Solar radiation

We assume that the heat influx due to solar radiation is given by:

$$Q_{rad}(t) = \tau A_g I(t). \tag{5}$$

Here $0 < \tau < 1$ represents the fraction of incoming solar radiation that contributes to glasshouse heating (assumed here to be constant), A_g is the base area of the glasshouse, and $I(t)$ is the intensity of solar radiation as it changes with time; for which historical data exists.

B.2 Heating system

The heating system acts to raise the internal energy of the glasshouse air via thermal conduction. More specifically, when applied, running hot water is sent through pipes that approximately span the glasshouse base area and act to dissipate the heat to the glasshouse air. For simplicity, we assume this conductive heat transfer is modelled by Newton heating:

$$Q_{heat}(t) = \kappa_{heat}\sigma_{heat}(t)\big(T_{heat}(t) - T(t)\big). \tag{6}$$

The constant κ_{heat} determines the rate of heat transfer from the pipes to the air, $\sigma_{heat}(t)$ is a function which changes in time like a switch between 1 when the heating system is applied and is zero otherwise, and $T_{heat}(t)$ is the temperature of the water entering the pipes.

B.3 Injection of flue gas

We model the heat influx due to flue gas injection with:

$$Q_{heat}(t) = \rho C_p \dot{V}_{flue}(t)\big(T_{flue} - T(t)\big). \tag{7}$$

Equation (7) assumes that flue gas enters the glasshouse according to the volume influx $\dot{V}(t)$ (volume of gas per unit time), with a mass density ρ_{flue}, heat capacity $C_p^{(flue)}$, and temperature T_{flue}. We assume that an equal volume of glasshouse air is passively displaced through the vents (which are open during flue gas injection) to equilibrate the pressure in the glasshouse to the outside atmospheric pressure.

B.4 Ventilation

We assume that the heat flux due to venting takes the form:

$$Q_{vents}(t) = -k_{vents}\rho C_p \dot{V}_{vents}(t)\big(T(t) - T_{atm}(t)\big). \tag{8}$$

Equation (8) accounts for the active exchange of air between the glasshouse and outside environment due to wind. We multiply the reported volume

exchange rate, $\dot{V}_{vents}(t)$, by a fraction $0 < k_{vents} < 1$ to account for possible lack of mixing between the air near the vents (at the roof) and the air near the plants (where the temperature is measured).

B.5 Conductive heat loss

We model heat conduction to the surroundings via Newton cooling:

$$Q_{cond}(t) = -\kappa_{cond}\big(T(t) - T_{atm}(t)\big). \tag{9}$$

For simplicity we assume that the passive conductive heat loss acts to equilibrate the glasshouse temperature to that of the surrounding air. The rate of heat conduction is determined by the constant κ_{cond}.

B.6 Misting

Costa Group use a misting system to cool and maintain appropriate humidity levels in their glasshouses. When applied, heat is removed from the glasshouse air to vaporise the mist. We model this heat loss by:

$$Q_{mist}(t) = -\kappa_{mist} f_{mist}(t), \tag{10}$$

where $0 \leqslant f_{mist}(t) \leqslant 1$ is the open fraction of the misting valve, and κ_{mist} measures the efficacy of misting to remove heat from the glasshouse air.

B.7 Model statement

Substituting the heat flux functional forms (5)-(10) into the energy balance (1), and dividing by $\rho V_g C_p$, gives the following equation for glasshouse air temperature:

$$\begin{aligned}\frac{dT}{dt} = {} & k_{rad} I(t) + k_{heat}\sigma_{heat}(t)\big(T_{heat}(t) - T\big) + \left(\frac{\dot{V}_{flue}(t)}{V_g}\right)\big(T_{flue} - T\big) \\ & - \left[k_{vents}\left(\frac{\dot{V}_{vents}(t)}{V_g}\right) + k_{cond}\right]\big(T - T_{atm}(t)\big) - k_{mist} f_{mist}(t).\end{aligned} \tag{11}$$

where we have introduced the following (dimensional) parameter groupings for notational clarity:

$$k_{rad} := \frac{\tau A_g}{\rho V_g C_p}, \quad k_{heat} := \frac{\kappa_{heat}}{\rho V_g C_p}, \quad k_{cond} := \frac{\kappa_{cond}}{\rho V_g C_p}, \quad k_{mist} := \frac{\kappa_{mist}}{\rho V_g C_p}. \quad (12)$$

The constant k_{vents} remains unchanged from its definition via equation (8).

B.8 Parameter inference

We use data, provided by Costa Group, to calibrate the mathematical model (11). Specifically, we use Bayesian Inference, implemented via the Python package PINTS [9], to obtain estimates for the parameters: k_{rad}, k_{heat}, k_{cond}, k_{mist} and k_{vents}.

To avoid identifiability issues with k_{vents} and k_{cond} (since they both act to equilibrate T and T_{atm}), we begin by fitting the model to a time interval in which the vents are closed. In doing so, we obtain estimates for k_{cond} and k_{heat}. We then estimate the remaining parameters by fitting the full model to a 5-day interval. In this computation we take the marginal posterior distributions for the reduced model fit of k_{cond} and k_{heat} as the new prior distributions.

B.8.1 Estimates for conduction and heating

We first consider the model behaviour with $k_{rad} = \dot{V}_{flue} = \dot{V}_{vents} = f_{mist} = 0$. This regime corresponds to a time interval in which there is no radiation (i.e. it is night), there is no injection of flue gas, the vents are closed, and misting is not applied. An example of such a time interval in the data is: 21:30 on 19th February to 01:10 on 20th February. In this setting, equation (11) reduces to:

$$\frac{dT}{dt} = k_{heat}\sigma_{heat}(t)\big(T_{heat}(t) - T\big) - k_{cond}\big(T - T_{atm}(t)\big), \quad (13)$$

which reflects how the temperature changes only via heating and conductive heat loss to the surroundings.

Smoothing spline interpolation: To solve equation (13) it is necessary to evaluate the time-dependent input functions $\sigma_{\text{heat}}(t)$, $T_{\text{heat}}(t)$ and $T_{\text{atm}}(t)$, at arbitrary times within the time interval of interest. Since the data only specifies these inputs at discrete time points, we use interpolation to specify the values of these quantities in continuous time. Specifically, we use a smoothing spline of degree 3, which we implement via the SciPy function: *scipy.interpolate.UnivariateSpline*.

Plots of the measured values of $\sigma_{\text{heat}}(t)$, $T_{\text{heat}}(t)$ and $T_{\text{atm}}(t)$ together with their smoothing spline approximations are shown in Figure 22. We note that although the indicated value of $T_{\text{heat}}(t)$ is zero for $t \lesssim 0.3$, this is simply the default value reported by the measurement instrument when no heating is applied. Mathematically, we see that $\sigma_{\text{heat}}(t) = 0$ for $t \lesssim 0.3$, so that the heating system does not act non-physically to cool the glasshouse to the indicated 0°C in our model.

Bayesian inference: We approach parameter calibration using the Bayesian inference paradigm. Full details of this approach are not given here, but may be found in for example [15]. In short, this approach uses Bayes' theorem,

$$p(\theta|\mathbf{d}) = \frac{p(\theta)p(\mathbf{d}|\theta)}{p(\mathbf{d})}, \tag{14}$$

to change the order of statistical conditioning relating the data, $\mathbf{d}$, to the model parameters, θ. Put more simply, it provides a means to go from the model telling us what we might expect to observe as data, to the data telling us what we believe the model (more specifically, the values of its parameters) might be.

As a statistical method, Bayesian inference works in terms of probability distributions, converting some prior beliefs about the model parameters expressed as a distribution, $p(\theta)$, into a new set of data-informed, *posterior* beliefs about them, $p(\theta|\mathbf{d})$. The term on the denominator of equation (14) is

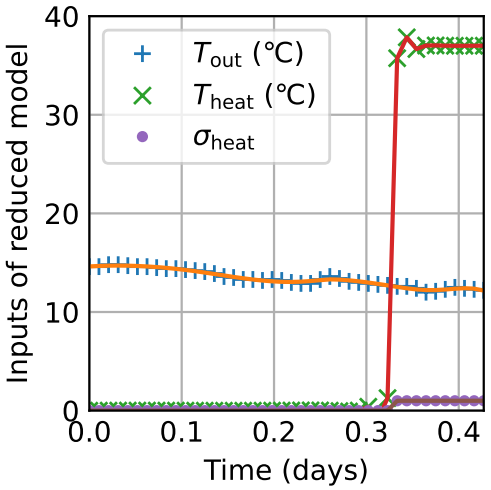


Figure 22: Reduced model inputs and their associated smoothing spline interpolations. Time $\mathfrak{t} = 0$ corresponds to 21:30 on 19th February.

called the model evidence, but is simply the normalising constant, calculated as an integral of the numerator over the entire output space. Computational approaches to Bayesian inference avoid this integral by using methods that work with ratios of equation (14).

The remaining term is known as the *likelihood*, $\mathfrak{p}(\mathbf{d}|\boldsymbol{\theta})$, describing the probability of observing some set of data given the output of the model for some value of its parameters. Here we work with deterministic models of form $\mathbf{y} = \mathcal{M}(\boldsymbol{\theta})$, and so the likelihood essentially acts as a way to penalise models that do not agree with the data (more formally, it encodes some assumptions about error in measuring the data being fit to, as well as non-modelled factors affecting the data). We make the simple but typical choice of a Gaussian likelihood with each datapoint treated independently, which takes the mathematical

form

$$p(\mathbf{d}|\theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(\mathcal{M}_i(\theta) - d_i)^2\right). \tag{15}$$

Here each d_i is one of the N points of data, and $\mathcal{M}_i(\theta)$ is the corresponding prediction of the model for that data point. Note that the likelihood is also denoted as depending on the variance, σ^2, which controls how strongly the likelihood penalises deviations from the data. This dependence is made explicit because we may also estimate σ^2 through the inference process, simply by treating it as an additional parameter, appending it to θ in equation (14). One can readily see that the likelihood is maximised by minimising the exponential term, at least if σ is fixed. In this case, the maximum likelihood value for the parameters corresponds to fitting them using least-squares estimation.

Here, our vector of parameters is $\theta = (k_{heat}, k_{cond})$. We choose a prior distribution that has all parameters independent and uniformly distributed over a large compact domain, reflecting a lack of information on the values of these parameters prior to performing inference. Model parameters k_{heat} and k_{cond} are distributed uniformly over $[0, 100]$, and σ is uniformly distributed over $(0, 100]$.

We estimate the posterior distribution numerically via Markov Chain Monte Carlo (MCMC) sampling. Specifically, we use a Metropolis-Hastings algorithm with adaptive covariance called *Haario Bardenet ACMC*. The three chains are initiated at $(k_{cond}, k_{heat}, \sigma^2) = q(3, 2, 1)$ for $q = 0.5, 1, 2$. We compute 10000 iterations and discard the first 2000 iterations as burn-in samples.

Results: The posterior distribution for the model parameters (including the noise parameter σ) is shown in Figure 23. We note that each of the marginal distributions (shown on top of each column) is localised and approximately normally distributed, showing that the parameters are practically identifiable. Indeed the results suggest the following estimates for the ODE model parameter values:

$$k_{heat} = 1.24 \pm 0.22\,\mathrm{day}^{-1}, \qquad k_{cond} = 1.85 \pm 0.09\,\mathrm{day}^{-1}, \tag{16}$$

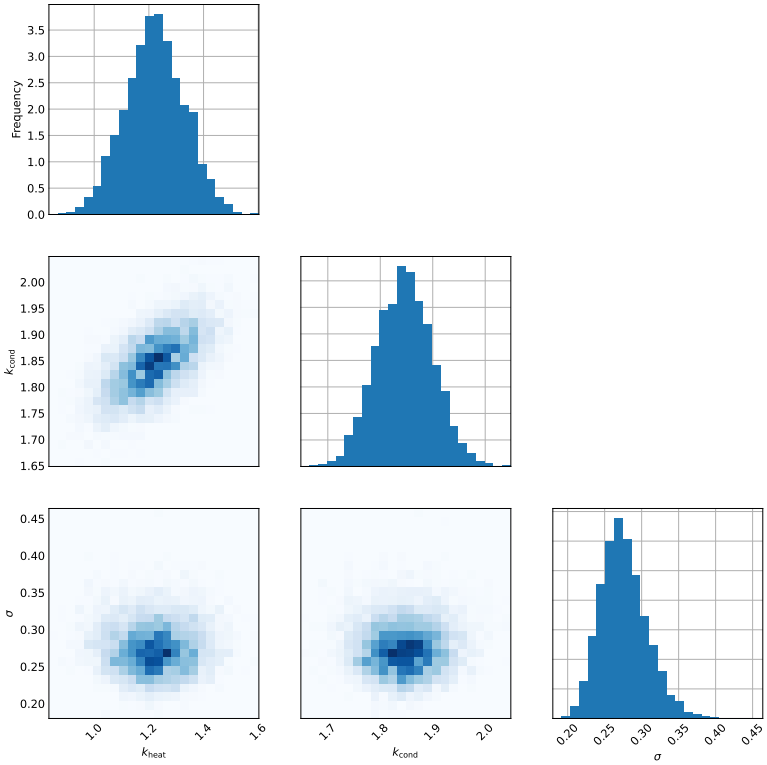


Figure 23: Summary plots of the posterior distribution for the reduced model (13).

and an estimate of $\sigma = 0.28 \pm 0.08\,^{\circ}\mathrm{C}$ for noise. The values here indicate the mean and uncertainties indicate one standard deviation.

We compare the solution $T(t)$ for the mean parameter values with the measured temperature data in Figure 24. The model agrees well with the data, with only one data point falling outside the 2σ envelope shaded in orange. In particular, we note that the model is capable of reproducing the rapid increase in temperature upon application of the heating system.

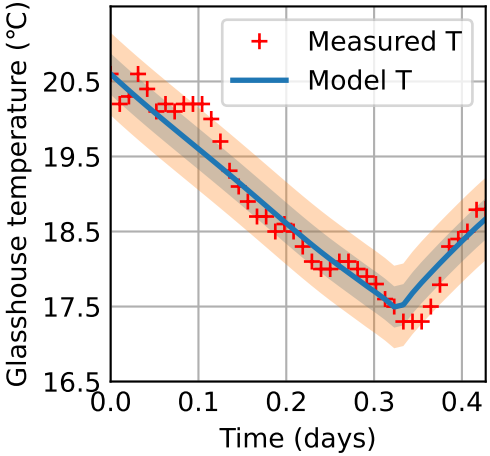


Figure 24: Fit of the reduced model to measured values of glasshouse temperature. The shaded regions indicate the predicted noise; the inner dark region is the 1σ envelope and the outer lighter region is the 2σ envelope. Time $t = 0$ corresponds to 21:30 on 19th February.

B.8.2 Estimates for the remaining parameters

We fit the full model (11) to a 5-day interval of data that starts from 0:00 on 23rd January. We choose such a multi-day time interval to ensure that the parameter estimates are robust to daily fluctuations that may occur (e.g. stochasticity in the weather conditions, glasshouse activity, etc.).

Plots of the input functions of the model are shown together with their smoothing spline interpolations in Figure 25. The plots show that the input quantities vary considerably in smoothness; the solar radiation intensity, $I(t)$, (normalised) ventilation air exchange rate $\dot{V}_{vents}(t)/V_g$, and misting valve open fraction, $f(t)$, each exhibit considerable fluctuation during certain time intervals. The fluctuations in $I(t)$ are presumably due to cloud cover in the final three days. The ventilation air exchange rate varies rapidly with time in a stochastic manner due to transient wind conditions. The open fraction of

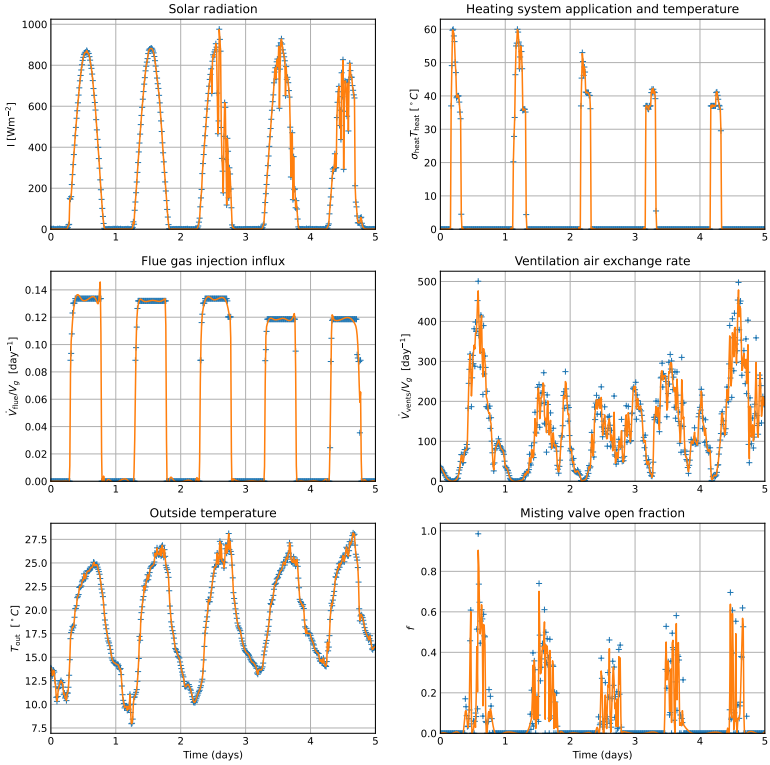


Figure 25: Data values and smoothing spline interpolations for the time-dependent inputs of the full model (11) for a 5-day interval beginning at 0:00 on 23rd of January.

the misting valve, $f(t)$, shows similar fluctuations but these are likely caused by operators applying the misting system in short bursts.

We use the results of the reduced model sampling to inform the prior distribution of the parameter values of the full model. Specifically, we assume that the prior distributions for heating and conduction are independent from each other and given by the normal distributions $k_{heat} \sim \mathcal{N}(1.24, 0.22^2)$ and $k_{cond} \sim \mathcal{N}(1.85, 0.09^2)$ respectively. The remaining parameters are supplied uninformative priors that are uniform distributions over: $k_{rad} \in [0, 10]$, $k_{vents} \in [0, 10]$, $k_{mist} \in [0, 5000]$ and $\sigma \in (0, 10]$. We choose a larger interval for the misting parameter, k_{mist}, because the magnitude of the valve open fraction, $0 \leqslant f(t) \leqslant 1$, is much smaller than the data values of solar radiation and ventilation air exchange rates (the former of which can peak at $I(t) \sim 850 \mathrm{Wm}^{-2}$). We assume that $T_{flue} = 50°\mathrm{C}$ for all simulations.

Summary plots of the posterior distribution are given in Figure 26. All marginal distributions are localised with a bell-like shape, which indicates that the model parameters are practically identifiable. The results supply the following parameter estimates:

$$\begin{aligned} k_{rad} &= 0.65 \pm 0.12\,°\mathrm{C\ day}^{-1}(\mathrm{W/m}^2)^{-1}, & k_{heat} &= 1.38 \pm 0.11\,\mathrm{day}^{-1}, \\ k_{vents} &= 0.66 \pm 0.16, & k_{cond} &= 1.87 \pm 0.09\,\mathrm{day}^{-1}, \\ k_{mist} &= 870 \pm 206\,°\mathrm{C\ day}^{-1} & \sigma &= 1.84 \pm 0.11\,°\mathrm{C}. \end{aligned} \tag{17}$$

We further note that the pairwise plot of the parameters k_{rad} and k_{vents} suggests that these parameters may be positively correlated. This raises the possibility of reducing the model by assuming an algebraic relation between k_{rad} and k_{vents} if model expediency is desired.

A plot comparing the measured temperature data to the solution of model (11) when the parameter values are set to the mean of the posterior distribution are given in Figure 27. The model agrees well with the measured data, with nearly all data points falling within the 2σ envelope. We note that the rapid fluctuations of the model output are due to the rapid variation in the inputs, as shown in Figure 25, as opposed to overfitting.

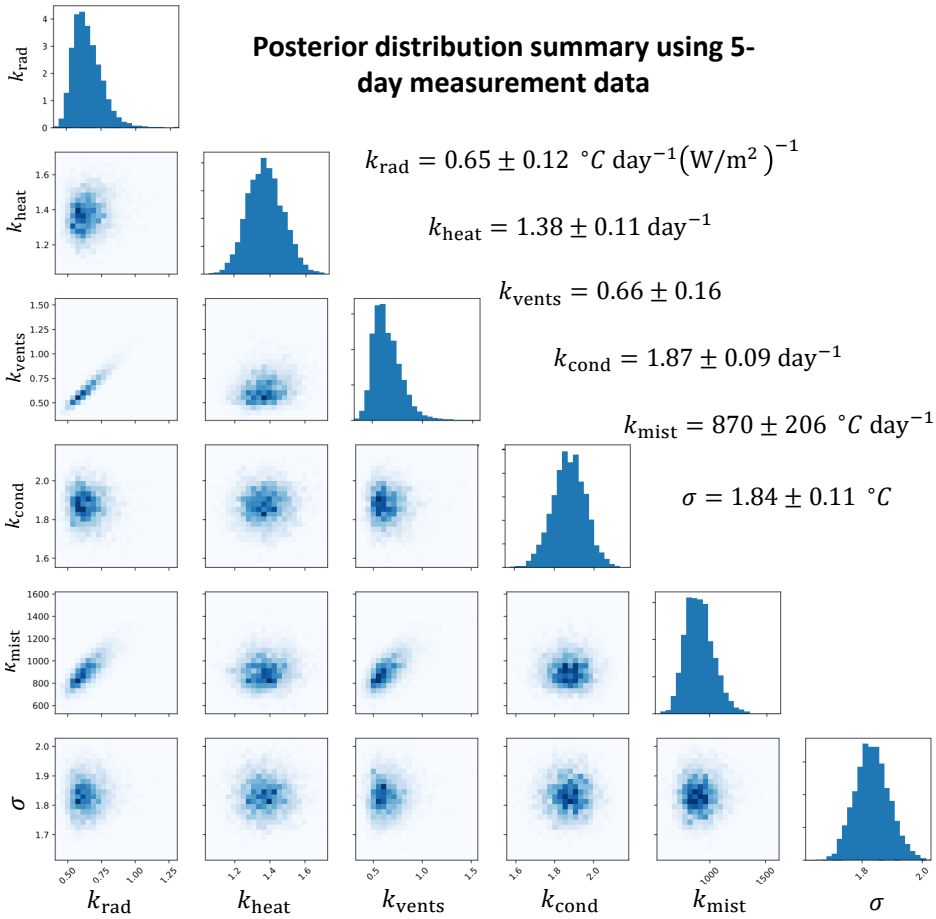


Figure 26: The posterior distribution for the parameters of the full model (11) when fit to 5-day data starting at 0:00 on 23rd January. Marginal distributions for each of the parameters are shown on the top of each column. Pairwise plots are shown below.

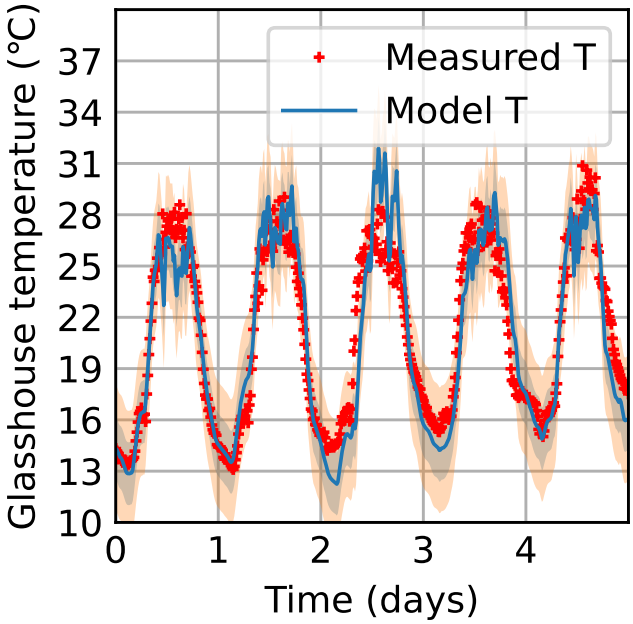


Figure 27: Plot of the solution to the full model (11) for the mode of the posterior distribution against the measured temperature values. The darker shaded region indicates the envelope for 1σ noise while the lighter region corresponds to 2σ noise.

B.9 Contributions to changing glasshouse temperature

We can use the parameter estimates (17) to quantify the relative contribution of the heat fluxes in changing the glasshouse temperature, $T(t)$. To do so we estimate typical values of the individual terms in equation (11) by using the following reference values:

$$\begin{aligned} I^{(\mathrm{ref})} &= 500\,\mathrm{W m^{-2}}, & \left(T_{\mathrm{heat}}^{(\mathrm{ref})} - T\right)^{(\mathrm{ref})} &= 25\,^{\circ}\mathrm{C}, \\ \left(\frac{\dot{V}_{\mathrm{flue}}}{V_g}\right)^{(\mathrm{ref})} &= 0.13\,\mathrm{day^{-1}}, & \left(T_{\mathrm{flue}} - T\right)^{(\mathrm{ref})} &= 25\,^{\circ}\mathrm{C}, \end{aligned} \tag{18}$$

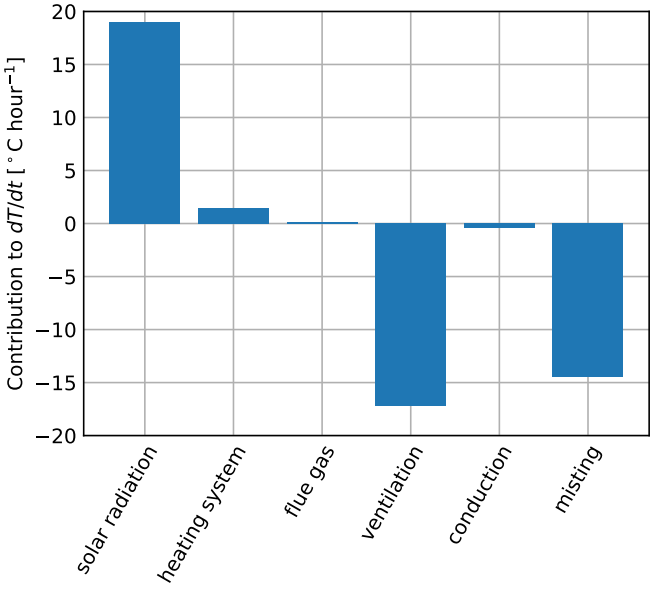


Figure 28: Typical impact of the heat sources/sinks on changing glasshouse temperature. These are the magnitudes of the terms on the right side of equation (11) using the parameter estimates (17) and reference values (18).

$$\left(\frac{\dot{V}_{\text{vents}}}{V_g}\right)^{(\text{ref})} = 200\,\text{day}^{-1}, \qquad \left(T - T_{\text{atm}}\right)^{(\text{ref})} = 5\,^{\circ}\text{C},$$
$$f_{\text{mist}}^{(\text{ref})} = 0.4.$$

The reference values (18) represent typical values of the input quantities during the daytime. The values for heating, flue gas and misting assume that these processes are applied.

A bar plot of the contributions to $\frac{dT}{dt}$ by each heat flux is shown in Figure 28. The results indicate, unsurprisingly, that solar radiation is the primary driver of temperature change in the glasshouse during the day. Ventilation and misting are the major heat sinks that oppose the increase in temperature due to solar radiation. The heating system and conduction to the external en-

vironment provide smaller, second-order contributions to temperature change. Finally, the injection of flue gas provides the smallest contribution to $\frac{dT}{dt}$ at $+0.14°C\ hour^{-1}$.

B.10 Effect of CO2 pumping temperature

One of the important questions for Costa Group is whether modifying the temperature at which the flue gas (containing CO_2) is pumped into the glasshouse has a significant impact on the temperature dynamics. Now that we have calibrated our mathematical model (11) to measured temperature data, we are in a position to quantitatively address this issue.

The rate of temperature increase due to flue gas injection is represented in our model by the term:

$$\tilde{Q}_{flue}(t) = \left(\frac{\dot{V}_{flue}(t)}{V_g}\right)(T_{flue} - T(t)). \tag{19}$$

The model fitting indicates that flue gas injection has a negligible impact on glasshouse temperature if the gas enters the glasshouse at the temperature $T_{flue} = 50°C$. Indeed, as shown in Figure 28, flue gas injection represents the smallest heat source by a considerable margin. Hence, it is clear that the flue gas temperature must be raised substantially in order to impact the glasshouse temperature dynamics. A quantity of interest is the value of T_{flue} required for the impact of flue gas injection to equal that of the heating system, which is $1.34°C\ hour^{-1}$. Setting $\tilde{Q}_{flue}$ to this value and assuming a typical glasshouse temperature of $T = 25°C$ and ventilation influx $\dot{V}_{flue}/V_g = 0.13\ day^{-1}$ (consistent with the reference values (18)) shows that the required flue gas temperature is:

$$T^{\star}_{flue} \approx 272°C. \tag{20}$$

This temperature is almost certainly too high to be safely injected into the glasshouse. We conclude that modifying the temperature of the flue gas

will induce a negligible impact on glasshouse temperature for the currently implemented volume influx of flue gas.

It remains an open question whether pumping in the flue gas at a lower temperature will allow it to be retained in the glasshouse for longer (especially when the vents are open) so that a greater proportion of the CO_2 is absorbed by the plants. This would require a detailed study of the dynamics of convection in the glasshouse in competition with the operation of fans (which are intended to create a uniform environment) and the air exchange through venting (which is used primarily to control the temperature). The Rayleigh number for air (at room temperature and using the 1.5m between rows as a characteristic length) has a magnitude of $3.6 \times 10^8 \times \Delta T$, where ΔT is the temperature difference in degrees Celsius, which indicates that convection is significant. A preliminary finite element computation of a two dimensional model was conducted, but far more work is required to develop and calibrate this model to produce actionable advice on the degree to which the temperature of the flue gas influences the retaining of CO_2. The assumption that fans in the glasshouse mix the air in the glasshouse on a relatively fast timescale also needs to be teased out as part of this analysis.

C CO_2 Model Details

C.1 Modelling CO2 dynamics

The CO_2 dynamics inside the glasshouse are primarily determined by three different effects: CO_2 injection, mixing of air with the external atmosphere due to venting, and the actions of the plants themselves (which both uptake CO_2 via photosynthesis, and release CO_2 via photorespiration and dark respiration). Mathematically, this is expressed via a mass balance equation,

$$\frac{d[V_g C]}{dt} = \mathcal{J}_{\text{inject}} + \mathcal{J}_{\text{vent}} + \mathcal{J}_{\text{plants}}. \tag{21}$$

Here V_g is the total volume of the glasshouse (treated as unchanging), and C is the concentration in $\text{kg}\,\text{m}^{-3}$ of CO_2 in the glasshouse atmosphere, such that $V \times C$ is the total mass of CO_2 inside the glasshouse. Each $\mathcal{J}$ term represents the flux of CO_2 (in $\text{kg}\,\text{h}^{-1}$) in or out of the glasshouse atmosphere, due to the listed effect. We now discuss the modelling of each effect in turn.

C.1.1 Injection

CO_2 is supplied to the glasshouse via pumping, as dictated by a control process that acts in response to sensor readings of the climatic conditions inside. The amount injected varies with time, both because the injection process is switched on or off, and the intensity of pumping varies. As such, a model operating on shorter time-scales (for example, hourly) must capture this variation. In seeking an optimal glasshouse strategy, this injection flux could be treated as a control variable, or as a deterministic function of other climatic variables that encodes some "growing strategy". Extension to strategy and control theory requires a model that we trust, however, and so our first step is to construct and validate such a model. This validation involves comparing model predictions to data, and as such we use Costa's readings for the injected gas to define $\mathcal{J}_{\text{inject}}$ as a function of time directly (as defined by $\dot{V}_{\text{flue}}(t)$).

C.1.2 Venting

Opening the glasshouse vents to allow air exchange with the surrounding atmosphere is for the purposes of reducing temperature or humidity, but this air exchange also serves to drive CO_2 levels within the glasshouse to equalise with the outside air. Under the assumptions discussed in Section A, namely that concentrations are the same throughout the glasshouse and that the air exchange data is correct, this term takes on a relatively simple form, at least for our initial data-fit purposes. Using the air exchange rate in $\mathrm{m^3\,h^{-1}}$, $\dot{V}_{\mathrm{vent}}$, the flux of CO_2 is given by the CO_2 coming in from the outside air, minus the CO_2 leaving with the glasshouse air,

$$\mathcal{J}_{\mathrm{vent}} = \dot{V}_{\mathrm{vent}}\, C_{\mathrm{atm}} - \dot{V}_{\mathrm{vent}}\, C = \dot{V}_{\mathrm{vent}}\left(C_{\mathrm{atm}} - C\right).$$

Here the atmospheric concentration of CO_2 could be treated as time-dependent, or constant as a simplification. Time-dependence of C_{atm} comes about through two factors that act on different timescales. On the short timescale, temperature variations cause minor fluctuations in the value of C_{atm} because here this has units $\mathrm{kg\,m^{-3}}$, and the density of CO_2 in the air is temperature-dependent. On longer timescales, atmospheric CO_2 levels vary due to seasonal changes in plant photosynthetic activity, and CO_2 emissions due to human activity. Atmospheric CO_2 levels are apparently not measured by Costa, such that modelling these variations may be important.

Defining some baseline temperature T^* (here we use the mean atmospheric temperature from the data used for CO_2 model fitting), we have

$$C_{\mathrm{atm}} = \frac{273.15 + T^*}{273.15 + T_{\mathrm{atm}}} C^*_{\mathrm{atm}},$$

where the fraction is the density-based correction in terms of external temperature T_{atm}, and C^*_{atm} is the atmospheric concentration of CO_2 at $T = T^*$. For the range of temperatures observed in the dataset, the density correction term varies from 0.93 to 1.06, to provide a sense of its importance (minor but perhaps not discountable). C^*_{atm} is still given as a time-dependent quantity

due to the possibility of seasonal variations. Fitting to some sea-level CO_2 data suggested a linear upwards trend plus a sinusoidal variation was an appropriate model, however this approach does not necessarily hold for a specific location at a different elevation.

C.1.3 Plant activity

Easily the most complex component of the CO_2 balance in the glasshouse is the activity of the plants themselves. Even making several simplifying assumptions including the aforementioned spatial uniformity and also a constant amount of photosynthetically active plant matter (which is perhaps less implausible than it might originally sound, due to Costa's careful control of leaf number), the rates of photosynthesis, photorespiration, and dark (not dependent on light) respiration are non-trivially dependent on the climatic conditions in the glasshouse. However, existing modelling efforts in this space should provide useful here. Although parameters of a plant CO_2 transfer model could be expected to vary from species to species, the fundamental process of C3 photosynthesis should be similar enough that taking a well-constructed model, even one fit to a different plant species, is a sufficiently good baseline.

The Farquhar *et al.* model [12] provides a thorough description of all of photosynthesis, photorespiration, and dark respiration. Underlying dynamics of molecules involved in chemical reactions are discussed and inform the model, but are not explicitly modelled themselves. However, the availability of electrons and carbon together determine the photosynthetic rate, and so it captures the most important features we would like to model (however, the impact of humidity is not explicitly capture). The model represents the CO_2 flux due to plants as

$$\mathcal{J}_{\text{plants}} = \mathcal{J}_{\text{photo}} + \mathcal{J}_{\text{dark}},$$

where the first term represents photosynthesis and photorespiration (which occur proportionally), and the second term represents dark respiration. The first term acts at a rate that is the slower of the rate allowed by light, $\mathcal{R}_{\text{light}}$

and that allowed by CO_2 , $\mathcal{R}_{CO2}$,

$$\mathcal{J}_{\text{photo}} = A_{\text{plants}} \min\left(\mathcal{R}_{\text{light}}, \mathcal{R}_{CO2}\right)\left(1 - \frac{1}{2}\phi\right). \tag{22}$$

Here A_{plants} is the area of (active) plant material, as the rates $\mathcal{R}$ are specified by Farquhar *et al.* as per unit area. A further constant of proportionality, ρ_{photo} specifies what proportion of the active plant material is chlorophyll. After dividing the fluxes by the volume of the glasshouse (as suggested by equation (21)), this term becomes a sort of glasshouse density of active plant matter. The constant of proportionality between photosynthesis and photorespiration, ϕ depends on the balance of concentrations of CO_2 (C) and O_2 (O) surrounding the plant,

$$\phi = 0.21\,\frac{O}{K_o}\frac{K_c}{C},$$

where the constant is taken from the Farquhar *et al.* work (though its temperature dependence is not modelled here), and K_o and K_c are Michaelis–Menten constants.

The rate of photosynthesis allowed by light is modelled by Farquhar *et al.* as

$$\mathcal{R}_{\text{light}} = \frac{\beta}{4(1+\phi)} \cdot \frac{1}{2}\left(I + I_{\text{sat}} + Z - \sqrt{(I + I_{\text{sat}} + Z)^2 - 4I\,I_{\text{sat}}}\right),$$

where β describes the conversion from irradiance I in the measured units ($W\,m^{-2}$) into a rate of (potential) CO_2 production, I_{sat} is the saturation irradiance, and Z is a curvature parameter. The denominator represents the availability of a photosynthesis reaction component, $NADP^+$, and so depends on the CO_2/O_2 balance of the plant. The final term is a saturating function that increases with I up to a value I_{sat}. One can observe that for $Z = 0$, it represents a linear increase up to I_{sat} and then no further increase, and values $Z > 0$ act to smooth out this curve.

The final component of the light-dependence of photosynthesis is that the maximum rate with which the plant can utilise light is temperature-dependent

(taking on a modified Arrhenius form),

$$I_{sat}(T) = I_{sat\text{-}max} \frac{H \exp\left(\frac{E_I S}{RH} + \frac{E_I}{H} \ln\left(\frac{H}{E_I} - 1\right)\right)}{H - E_I} \frac{\exp\left(-\frac{E_I}{R(T+273.15)}\right)}{1 + \exp\left(\frac{S}{R} - \frac{H}{R(T+273.15)}\right)},$$

where S, H and E_I are terms related to the energies of the overall electron-sourcing reaction, and R is the gas constant. The complex constant in front of the temperature-dependent term simply scales the function so that I_{sat} takes a specified maximum value, $I_{sat\text{-}max}$, so as to ease parameterisation. This maximal value is attained at $T = \frac{H}{S + R\ln(H/E_I - 1)} - 273.15 = 31.02\,°C$ for the baseline parameter values in Farquhar *et al.* (but these are not parameters specifically for tomato).

The rate of photosynthesis allowed by CO_2 is modelled by Farquhar *et al.* as

$$\mathcal{R}_{CO2} = r_{cmax} \frac{C}{C + K_c\left(1 + \frac{O}{K_o}\right)},$$

where the definitions of constants are as given in the definition for ϕ, above. The maximal rate is defined by r_{cmax}, which is temperature dependent in Arrhenius fashion, and indeed the Michaelis–Menten constants also vary with temperature in the same manner, and each parameter may be expressed in terms of its value at the baseline temperature,

$$r_{cmax}(T) = r^*_{cmax} \exp\left(\frac{E_c}{R}\left(\frac{1}{T^* + 273.15} - \frac{1}{T + 273.15}\right)\right)$$
$$K_c(T) = K^*_c \exp\left(\frac{E_{Kc}}{R}\left(\frac{1}{T^* + 273.15} - \frac{1}{T + 273.15}\right)\right)$$
$$K_o(T) = K^*_o \exp\left(\frac{E_{Ko}}{R}\left(\frac{1}{T^* + 273.15} - \frac{1}{T + 273.15}\right)\right).$$

Each parameter has its own activation energy, E, that determines the strength of its temperature dependence.

The final term is dark respiration, which is treated by Farquhar *et al.* [12] as a rate per amount of plant (here treated as constant) that depends only on temperature. Here we incorporate an additional point of flexibility in that day/night status also affects the rate of respiration,

$$\mathcal{J}_{\text{dark}} = \begin{cases} A_{\text{plants}}\, r_d^*\, \exp\left(\frac{E_{Rd}}{R}\left(\frac{1}{T^*+273.15} - \frac{1}{T+273.15}\right)\right) & I \geqslant I_{\text{day}} \\ A_{\text{plants}}\, \gamma_{\text{night}} r_d^*\, \exp\left(\frac{E_{Rd}}{R}\left(\frac{1}{T^*+273.15} - \frac{1}{T+273.15}\right)\right) & I < I_{\text{day}} \end{cases}$$

Here A_{plants} is again the area of active plant material, r_d^* is the rate per unit area of respiration at $T = T^*$, E_{Rd} controls the temperature dependence of the respiration rate, and γ_{night} is the multiplicative factor describing the difference in nighttime and daytime respiratory behaviour in plants. This description, along with the other temperature dependence relations (excepting the more complex one for light), assumes the temperature does not exceed the value at which a drop-off effect begins to occur. As this value is roughly around 40 °C for plants, this is likely to be a reasonable assumption for the conditions inside the glasshouse at Guyra. More specifically, temperatures above 35 °C are rarely seen in the data and where they do occur they occur as brief spikes, presumably reflecting quick temperature correction by venting.

C.2 Model Calibration and Validation

Inference Process

For almost all of the parameters discussed above, values are available (albeit in different units) in the Farquhar *et al.* [12] paper, although many parameter values are sourced from experimental observations of spinach. As a result, we infer parameters in response to Costa's data. As discussed in the main text, after first fitting to Costa's data alone and observing unphysiological results when the model was then used to predict, we additionally used the data published for tomatoes under controlled conditions in [8] in the inference process.

We use the Bayesian inference paradigm already described in Section B.8.1, generating samples from the posterior as defined by equation (14) and again

using a likelihood of form (15). For each parameter allowed to vary, k_i, we define a prior that is uniform in *scaling* around some baseline value,

$$k_i = k_{i(\text{base})} 10^{\theta_i}. \tag{23}$$

These baseline values are selected as described below. The parameters defining the scaling factors θ_i are each given an independent, uniform prior, centred on a value of zero (corresponding to no scaling applied, $10^0 = 1$). The range allowed for the scaling is selected in a parameter-dependent manner, in attempt to avoid scaling parameters expected to be sensitive by overly large amounts. The joint prior can be expressed mathematically as a density function in terms of the indicator function, χ_c, that takes a value of one when the condition c is satisfied and a value of zero otherwise,

$$p(\theta) = \prod_{i=1}^{N_p} \frac{1}{2\theta_{\max,i}} \chi_{\theta_i \leqslant \theta_{\max,i}} \chi_{\theta_i \geqslant -\theta_{\max,i}} \tag{24}$$

For both setting parameter baseline values and later inferring an ensemble of models fit to Costa's data, sequential Monte Carlo [11] is used to generate samples from the posterior (14). Sequential Monte Carlo is a technique that takes some number of samples from the prior, termed particles, and gradually evolves their locations in the parameter space so that they represent samples of a sequence of intermediary distributions that eventually reaches the posterior. This approach can aid sampling by gradually introducing the complexity of the sampling problem, and allows multiple regions of high density to be simultaneously explored (consider the related intuition of annealing schedules that use a gradually cooling temperature to achieve a combination of global and local exploration).

Baseline parameter values were set by first calibrating the Farquhar *et al.* model to the data published in [8]. This data represents measured photosynthesis rates for different levels of C, I and T, collected from the figures depicting photosynthesis-versus-CO_2 curves using a plot digitiser. The parameters r_{cmax}, E_c, K_c, E_{Kc}, K_o, E_{Ko}, β, $I_{sat\text{-}max}$, E_I, Z, r_d^*, E_d, S and H were

varied over large ranges centred on their values published in [12], to maximise the ability to find reasonable fits to the Choi *et al.* data. Out of the many potential fits to the photosynthesis data returned by Bayesian inference, the set with values overall closest to the published values in [12] (the θ with the smallest 1-norm) were used as the baseline calibration. Once this baseline was found, those parameters were varied over a smaller range (centred on the new baseline) in calibrating to Costa's data.

Regarding other parameters separate to the photosynthesis component of the model, the amount of active plant matter (A_{plants} was set to a baseline value that gave reasonable predictions when other parameters were at baseline, and γ_{night} was set rather arbitrarily (both of these parameters are allowed to vary over a reasonably large range). In terms of fixed parameters, the glasshouse volume was provided by Costa (V_g), baseline atmospheric CO_2 density (C^*_{atm}) was set to a lower value such that minimal CO_2 levels in the data were at its upper range, ambient O_2 concentration (constant) was set to the value in Farquhar *et al.* [12], T^* was set using mean external temperature and I_{day} was set to a small, non-zero value to ensure that minor readings during nighttime did not incorrectly register as day.

The baseline parameters, and the ranges over which they were allowed to vary for subsequent calibration to Costa's data, are all given in Table 4. Again, photosynthesis parameters were allowed to vary only over small ranges, to avoid losing the physiological realism achieved by first calibrating to the Choi *et al.* [8] experimental data for CO_2 intake. Inference was performed using 5000 particles, resulting in an ensemble of 5000 plausible models that were all calibrated to five days of Costa data. The success of calibration is discussed in the main document, but here we present some additional analysis of the results of the Bayesian calibration process.

Posterior Analysis

In Figure 29, we present the probabilities that each scaling parameter takes on a certain value across the set of candidate models that we obtain by

Variable Parameters			
Param.	**Meaning**	**Base Value**	θ_{max}
α_{inject}	Scaling factor for CO_2 injection data	1 *(no dim.)*	0.3
α_{vent}	Scaling factor for vent air exchange data	1 *(no dim.)*	0.3
C^*_{atm}	Atmospheric CO_2 level (at $T = T^*$)	$4 \times 10^{-4}\,kg\,m^{-3}$	0.1
A_{plants}	Surface area of active plant matter	$5 \times 10^{4}\,m^2$	2
r_{cmax}	CO_2-saturated photosynthesis rate	$7.60 \times 10^{-3}\,kg\,m^{-2}\,h^{-1}$	0.2
E_c	r_{cmax} activation energy (T-dependence)	$4.974 \times 10^{4}\,J\,mol^{-1}$	0.2
K_c	Michaelis–Menten constant for CO_2	$6.114 \times 10^{-4}\,kg\,m^{-3}$	0.2
E_{Kc}	K_c activation energy (T-dependence)	$7.009 \times 10^{4}\,J\,mol^{-1}$	0.2
K_o	Michaelis–Menten constant for O_2	$0.261\,kg\,m^{-3}$	0.2
E_{Ko}	K_o activation energy (T-dependence)	$4.943 \times 10^{4}\,J\,mol^{-1}$	0.2
β	Potential CO_2 production from light	$1.121 \times 10^{-4}\,kg\,W^{-1}\,h^{-1}$	0.2
$I_{sat\text{-}max}$	Light level of photosynthesis saturation	$280.85\,W\,m^{-2}$	0.2
E_I	$I_{sat\text{-}max}$ activation energy (T-dependence)	$1.670 \times 10^{4}\,J\,mol^{-1}$	0.2
Z	Curvature parameter for light saturation	$32.7\,W\,m^{-2}$	1
r^*_d	Respiration rate at baseline temperature	$1.038 \times 10^{-4}\,kg\,m^{-2}\,h^{-1}$	0.2
E_d	r_d activation energy (T-dependence)	$3.812 \times 10^{4}\,J\,mol^{-1}$	0.2
γ_{night}	Increase in nighttime respiration	2 *(no dim.)*	1
Fixed Parameters			
V_g	Volume of the glasshouse	$3.28 \times 10^{5}\,m^3$	-
T^*	Baseline temperature	11.5 °C	-
R	Universal gas constant	$8.314\,J\,mol^{-1}\,K^{-1}$	-
O	Concentration of O_2 inside glasshouse	$0.2577\,kg\,m^{-3}$	-
S	Entropy for light saturation "reaction"	$690.27\,J\,K^{-1}\,mol^{-1}$	-
H	Enthalpy for light saturation "reaction"	$2.277 \times 10^{5}\,J\,mol^{-1}$	-
I_{day}	Threshold irradiance delineating daytime	$5\,W\,m^{-2}$	-

Table 4: Baseline parameter values used to calibrate the CO_2 model to Costa's data, after first fitting to published photosynthesis data. Parameters listed as variable have their values inferred from fitting to the data using the quoted value as a baseline, as described in Section C.2. The value of θ_{max} specifies to what extent that parameter is allowed to vary, as per (23) and (24).

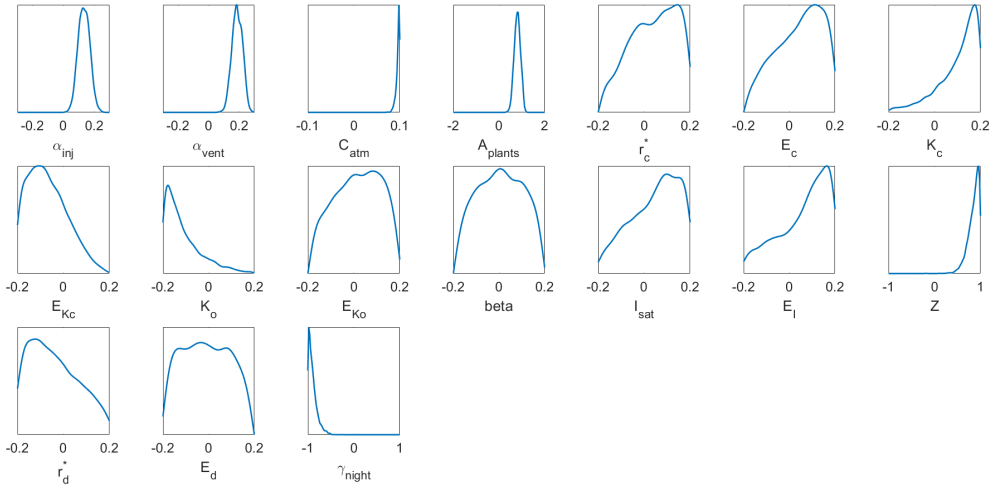


Figure 29: Marginal distributions for the scaling factor multiplying each varied parameter in the CO_2 model. Each density function is plotted over the range of its prior, in order to make the effect of the model-data fit process more clear. Sharp spikes represent parameters well-identified by the data, whereas those that take diffuse ranges are potentially less critical to the model-data fit process.

calibration. Note that the distributions between all pairs of parameters, as can be visualised in bivariate scatter plots, were also examined to search for obvious two-parameter interdependencies, but little worthy of comment was observed. A simplified way to interpret the results in Figure 29 is that if the probability mass is concentrated in a small region (tall, thin peak), then this suggests that it is very important that the parameter take on about that value in order to achieve successful calibration. If probability mass is distributed rather evenly throughout the whole of the plotted range, then this indicates that the parameter may be unimportant, and its value is not learned well from the calibration data.

The parameters α_{inject} and α_{vent} are well-constrained by the model-data fit process, but both are suggested to take on values larger than 1, suggesting

a discrepancy between Costa's measurements for these processes, and the model's beliefs regarding their strength (specific scalings for the best-fit parameter are given in Table 2 in the main document). We suspect this discrepancy arises due to a combination of deficiencies in the data for these processes (in particular venting relies upon an empirical calculation to predict amount of air exchange), and deficiencies in the model combined with the restrictive prior applied to the parameters. It might be informative to compare the model calibrations discussed here, to a separate set where α_{inject} and α_{vent} are constrained to a value of one (no leeway allowed, meaning Costa's injection and venting data is interpreted verbatim).

The value of C_{atm} is also seen to be very strongly localised, and to the edge of the range allowed by the prior placed on this variable. Typically, this would suggest that the prior needs to be adjusted, but in this case the restrictive prior was specifically chosen to ensure that C_{atm} would not take on values so high that venting could not reduce CO_2 to the levels observed in the first five days of data. Preliminary calibration efforts indicated that if C_{atm} was allowed to take on larger values, the fitting process would set its value too high, seeking a level that matched the average CO_2 levels across the night and in the process essentially abandoning any attempt to properly fit the data around sunset and subsequent increase in CO_2 due to dark respiration overnight.

Interestingly, the parameter fitting process makes stronger conclusions about the values of auxilliary parameters describing photosynthesis CO_2 dependence (in particular the Michaelis–Menten constants K_c and K_o), than it does the base rate, r_c^*. While the rich variation in temperature and CO_2 levels inside the glasshouse does potentially provide enough information to effectively constrain these parameters, especially across only five days of data it is not expected that these parameters are being uniquely identified together. Performing a more computationally expensive calibration that uses more of Costa's data might help with this.

The other most interesting distributions among the parameter marginals

are those for Z and γ_{night}. The curvature parameter for the photosynthesis-irradiance relationship (Z) is pushed to the maximal end of its range, suggesting that a better fit to the data would be achieved by even further delaying the onset of saturation for increasing light levels. Although this might help with the fit to the data, allowing its value to increase further seems liable to produce photosynthesis-irradiance curves predicted by the model that saturate too late and are hence somewhat unphysiological.

The other parameter bunched up against the bounds of its prior is γ_{night}, and interestingly its value is pushed to the lower bound, such that $\gamma_{night} \approx 0.2$. Although some difference in respiratory activity during the day and night is physiologically expected (reflecting for example changes in the relative levels of maintenance-driven and growth-driven respiration), this extreme value perhaps suggests that the respiration level that fits the daytime data has been overestimated, and γ_{night} is being set to a value much lower than one in order to compensate when it comes to fitting the nighttime data.

Calculations of Total Photosynthetic Achievement

As described in the main document, we evaluate the effects of changing the CO_2 injection strategy by multiplying all rates of injection in the data by a constant (denoted κ). This is a simplified representation of changing the injection strategy, but without a fully-integrated model combining CO_2, temperature and humidity, it is difficult to formulate or evaluate more sophisticated modifications of injection strategy. For example, decisions about venting will also depend on temperature and humidity, and these decisions will then also modify the optimal injection strategy.

Dark respiration represents the utilisation of stored carbon for plant function and growth, and so release of CO_2 through this process does not represent a loss of potential growth (as opposed to photorespiration, which is a wasteful competition for rubisco by O_2 , and so does limit net plant carbon uptake through photosynthesis). Following this thinking, in calculating the total amount of photosynthetic activity over some time period $[0, T]$, we ignore

dark respiration and simply integrate the rate from equation (22) in the main document,

$$P_{\text{total}} = A_{\text{plants}} \int_0^T \min\left(\mathcal{R}_{\text{light}}(t), \mathcal{R}_{\text{CO2}}(t)\right) \left(1 - \frac{1}{2}\phi(t)\right) \, dt. \tag{25}$$

Given the good temporal resolution with which the data is collected and model solution, this integral is evaluated numerically via the trapezoidal rule.

Figure 10 is calculated by evaluating equation (25) for the representative set of models from the calibrated population, after simulating the glasshouse CO_2 levels using the model's predictions and Costa's injection/venting data together. The injection rates are all multiplied by κ, reflecting the change in strategy.

C.3 Mass balance equation

There are three components to the rate of change of greenhouse CO_2 mass:

- Air exchange with the outside atmosphere;
- Injection of CO_2;
- Uptake of CO_2 by the tomato plants.

For the mass balance of CO_2 in the glasshouse, if C [kg-m^{-3}] is concentration of CO_2, then the mass balance is

$$\frac{d(V_g C)}{dt} = f_{\text{vent in}} - f_{\text{vent out}} + f_{\text{inject}} - f_{\text{plants}}, \tag{26}$$

where V_g [m^3] is the total glasshouse volume. During venting, the CO_2 entering the glasshouse is $f_{\text{vent in}} = C_a \dot{V}_{\text{in}}$ and that exiting the glasshouse is $f_{\text{vent out}} = C\dot{V}_{\text{out}}$, where C_a is the ambient CO_2 concentration and $\dot{V}_{\text{in}}$, $\dot{V}_{\text{out}}$ are respectively the volume flow rates [m^3-sec^{-1}] into and out of the vents. As, by assumption, glasshouse environment equilibration is fast compared

to timescales of modelling interest, pressure fluctuations are neglected and $\dot{V}_{in} = \dot{V}_{out} = \dot{V}$.

For injected CO_2, $f_{inject} = C_f \dot{V}_f$, where C_f is the CO_2 concentration of the injected flue gas and $\dot{V}_f$ is the volume flow rate of the injected flue gas.

How to mathematically describe the amount of CO_2 adsorbed (or respired) by plants, f_{plants} [kg-sec^{-1}], is the key unknown to complete modelling of the CO_2 mass balance inside the glasshouse. Before addressing accurately approximating a form for f_{plants}, we non-dimensionalize and extract the structure of equation 26 that does not depend on the exact form of f_{plants}.

To non-dimensionalize, multiply equation 26 by $T_s/(C_a V_g)$, where T_s is a characteristic time scale that we leave unspecified for the moment. The product $C_a V_g$ sets the scale for the mass of CO_2 in the glasshouse: $C_a V_g$ is the amount of CO_2 in the glasshouse *in the absence of normal growing activity*, i.e. the kilograms of CO_2 in the glasshouse at ambient concentration. Activities of injection, absorbtion, and venting change the amount of CO_2 with respect to the ambient concentration. With this multiplication equation 26 becomes

$$\frac{d\frac{C}{C_a}}{d\frac{t}{T_s}} = \frac{(-C + C_a)\dot{V}}{\frac{C_a V_g}{T_s}} + \frac{C_f \dot{V}_f}{\frac{C_a V_g}{T_s}} - \frac{f_{plants}}{\frac{C_a V_g}{T_s}}. \tag{27}$$

Noting that the air exchange rate measured by Costa is $\dot{V}/V_g$, we can similarly call $\dot{V}_f/V_g$ the flue gas exchange rate. Defining $C' = C/C_a$, $t' = t/T_s$ and dropping the primes, equation 27 becomes

$$\frac{dC}{dt} = (1 - C)\frac{\dot{V}T_s}{V_g} + \frac{C_f}{C_a}\frac{\dot{V}_f T_s}{V_g} - \frac{f_{plants}}{\frac{C_a V_g}{T_s}}. \tag{28}$$

Now comes a choice of time scale T_s. A natural choice of time scale is the inverse of the air exchange rate: $T_s = V_g/\dot{V}$. This has the advantage of removing a parameter from equation 28; however, other choices, such as 12 or 24 hours may be informative too. For now we chose the inverse air exchange rate for T_s so that the non-dimensionalized form of the mass balance equation

becomes

$$\frac{dC}{dt} = (1 - C) + \Gamma_{in} + \Gamma_{pl}, \tag{29}$$

where

$$\Gamma_{in} = \frac{C_f \dot{V}_f}{C_a \dot{V}} \qquad \Gamma_{ab} = \frac{f_{plants}}{C_a \dot{V}}. \tag{30}$$

Γ_{in} is the ratio of the flue gas and air exchange ratios weighted by the excess concentration of the flue gas. When $\Gamma_{in} - \Gamma_{ab} = 0$, the fluxes of injected and absorbed CO_2 are in balance, and all the costly injected CO_2 is productively absorbed by plants. In reality $\Gamma_{in} > \Gamma_{ab}$ to overcome losses from venting. A key modelling question is how closely can we bring the fluxes of injected and absorbed CO_2 into balance within the constraints of operating the glasshouse?

Before discussing mathematical forms for f_{plants}, we next make a qualitative analysis of the ODE equation 29.

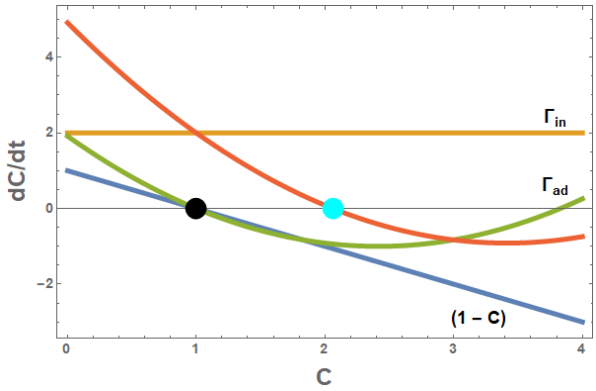
C.4 Qualitative analysis and expectations

A qualitative analysis of an ODE seeks to uncover the major structure of the equation or set of equations: fixed points, their stability, and how the location and stability of fixed points changes with parameters. The state- or phase-space of equation 29 is $(dC/dt, C)$, as depicted schematically in figure 30(a) which shows separately each term on the right-hand side of equation 29, as well as their sum.

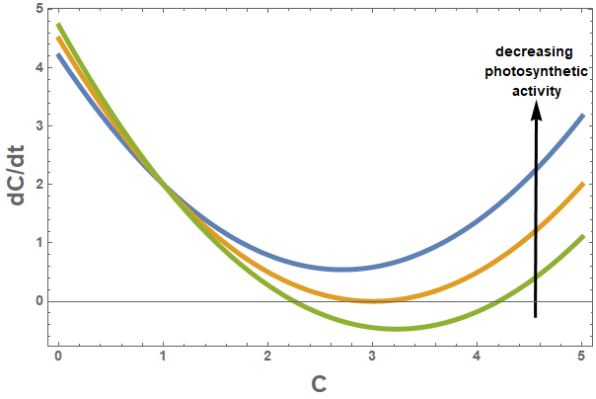
Fixed points are points in phase space where $dC/dt = 0$. Setting $dC/dt = 0$ in equation 29 and solving for C, fixed points occur at concentrations

$$C^* = 1 + \Gamma_{in} + \Gamma_{pl}. \tag{31}$$

At night, when $\Gamma_{in} = 0$, the fixed point is around the black dot in figure 30(a): concentration moves towards ambient or $C^* = 1 + \Gamma_{pl}$ if there is respiration from the plants. Stability of a fixed point is determined by the slope of the line of the function dC/dt at C^*. A stable fixed point has a negative slope,



(a)



(b)

Figure 30: Qualitative analysis of CO_2 mass balance equation 29. (a) Colored lines are flux terms in equation 29 with the red line their sum for certain ranges of parameter values. Black dot is a possible night time stable fixed point, and the cyan dot is a possible day time stable fixed point. (b) For some parameter ranges there is no stable day time fixed point. Here $|\Gamma_{\text{pl}}^{\text{min}}|$ decreases photosynthetic effect. At some point, though $|\Gamma_{\text{pl}}^{\text{min}}|$ is still well above zero, a day time fixed point ceases to exist (blue curve).

and an unstable fixed point has a positive slope. This can be seen by noting that for a negative slope a small increase in C from C^* means that $dC/dt < 0$ and the system pushes the concentration back towards C^*, while for a small decrease in C from C^*, $dC/dt > 0$ and again the system pushes C back towards C^*. At an unstable fixed point, on the other hand, small fluctuations around C^* are amplified by the system and C moves away from an unstable fixed point.

During the day we can take Γ_{in} as approximately a constant, the gold line in figure 30(a). As plants, during the day, adsorb CO_2, $dC/dt < 0$. As it is also physically realistic that photosynthetic activity falls off at both high and low extremes of CO_2 concentration, for the purposes of this section we will assume that Γ_{pl} is a wide parabola,

$$\Gamma_{pl} = -|\Gamma_{pl}^{min}| + \lambda(C - \bar{C})^2, \tag{32}$$

where Γ_{pl}^{min}, λ, and $\bar{C}$ are parameters that in principle depend on temperature and perhaps other variables, but that in this section we will manipulate to uncover fixed points and their changes of stability. In fact, our conclusions in this section only depend on Γ_{pl} having a minimum. As we have scaled concentration to be the excess (higher or lower) concentration from ambient, likewise Γ_{pl} is the excess flux beyond that at ambient concentration, and Γ_{pl} should go through zero at $C = 1$. A brief calculation shows that enforcing this constraint can be accomplished by setting $\bar{C} = 1 + \sqrt{|\Gamma_{pl}^{min}|/\lambda}$.

The results are in figure 30. For night time there is a stable fixed point (black dot) at ambient $C = 1$ or or a little above accounting for respiration as a small constant. During the day, the red curve in figure 30(a) shows the sum of the three fluxes for some range of parameter values. The fixed point moves to the cyan dot. For some ranges of parameters the glasshouse, according to equation 29 cycles between day and night fixed points. However, for other parameter ranges there is no day time fixed point. Figure 30(b) shows a parameter variation reducing photosynthetic activity (reducing $|\Gamma_{pl}^{min}|$ but keeping its value well above zero) and moving the system from having a

fixed point (green curve) to not having a fixed point (blue curve). Without a stable attractor it is unclear, without further analysis, what the behavior of equation 29 will be.

In the next section we interrogate the data to see if there are or are not stable attractors in the glasshouse: how much of the structure of equation 29 is reflected in the glasshouse data?

C.5 Glasshouse CO2 data is chaotic

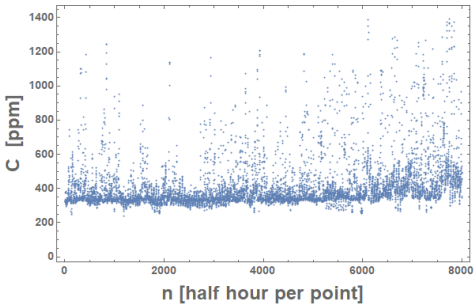
Independent of any equation we can reconstruct the true phase-space from the CO_2 time-series data using lagged embedding coordinates [31]. If C_k is the kth data point, we can form a d-dimensional vector from the time-series

$$(C_{k-\tau}, C_{k-2\tau}, \ldots, C_{k-d\tau}), \tag{33}$$

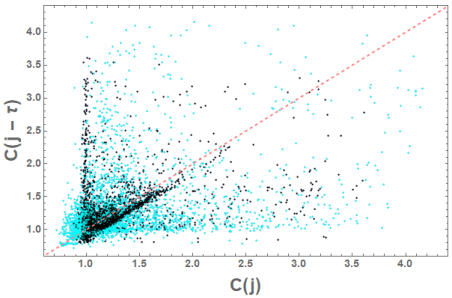
where τ is the delay. The embedding dimension d and the delay τ are chosen so as to optimally unfold the dynamics in the phase-space, which just means to, to the extent possible with noise in the data, represent the essential dynamical system in a visually obvious way. Taken's theorem [31] guarantees that the reconstructed phase-space is equivalent to the true phase-space, up to a change of coordinates.

Figure 31a shows the CO_2 time-series data from Costa glasshouse 1 in units of parts per million [ppm]; points are measured every half hour, and cover about six months growing season from December through June. While the global average atmospheric CO_2 concentration is 414 pppm, the most common value through the noticeable fuzzy of points is $C_a = 335$ ppm, which we take as the effective ambient CO_2 concentration at the site.

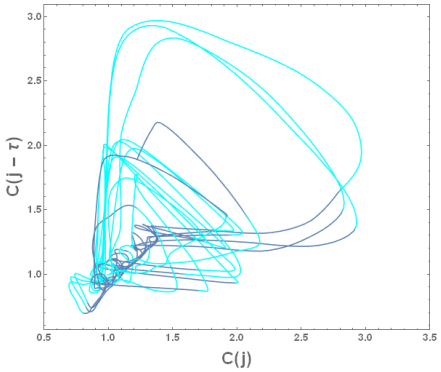
Using the false-nearest-neighbors method [19], the embedding dimension could be 2 or 3, depending on the noise tolerance. We choose 2 for simplicity. Using the first minimum of the mutual information of the delayed time-series [13], the delay is $\tau = 4$, or in real time two hours. Figure 31b shows the reconstructed CO_2 phase-space using the full data set in figure 31a. The dashed pink line



(a) CO_2 time-series data (measured every half hour) from midnight 6th December to 16:00 21st July.



(b) Lagged coordinate reconstruction of CO_2 phase space from all data in panel (a) scaled to ambient concentration. Pink dashed line is diagonal.



(c) Smoothed lines replace data points for an arbitrarily chosen 11 days (points 6500 to 7048) from the data in panel (a); suggests the smooth trajectories that would be obtained by continuous measurement.

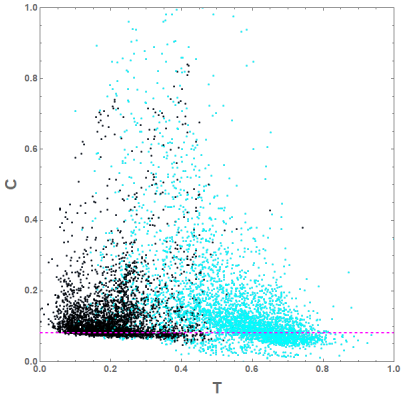
Figure 31: CO_2 data (half hour per point) and the phase space reconstructed from data. Lag is $\tau = 4$ points or 2 hours. Colors indicate day (cyan) or night (black) based on the coincidentally measured light intensity in the glasshouse. Fractal dimension of the attractor in (b) is 1.8 ± 0.1.

is a diagonal, along which any fixed points should show up. Note that below the diagonal $(C_j - C_{j-\tau})/\tau > 0$, i.e. where dC/dt is increasing, and above the diagonal is where dC/dt is decreasing. The data points are colored cyan for day time and black for night time. Day/night was determined by the coincidentally measured light intensity in the glasshouse. An examination of the light intensity time-series says that a value of 20 W/m^2 reasonably divides day from night. Figure 31c shows a typical ten days of data with the points joined by continuous fitted lines.

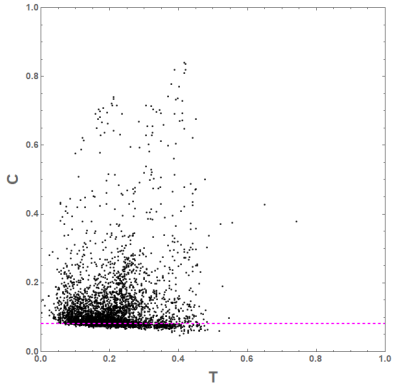
A standard question to ask in this type of nonlinear time-series analysis is, what is the dimension of the attractor in the phase-space?

As we hypothesize the CO_2 concentration-temperature dynamical system is the key interaction to understand glasshouse dynamics, it is useful to see how that system functions currently in the glasshouse. Figure 32 shows the concentration-temperature data at half hour intervals over approximately six months (see figure caption for exact dates). The concentration and temperature data has been put on a common scale with variable $X = \dfrac{(X_{raw} \min(X_{raw}))}{(\max(X_{raw}) - \min(X_{raw}))}$ which scales both C and T to a range $[0, 1]$. $(\min(C_{raw}), \max(C_{raw})) = (240, 1394)$ ppm, and $(\min(T_{raw}), \max(T_{raw})) = (12.1, 33.0)$ °C. The dashed magenta line indicates $C_a = 0.08$, and the colors indicate day (cyan) and night (black) times. Panel (a) shows the full data set, while panels (b) and (c) separately plot the night and day data respectively.

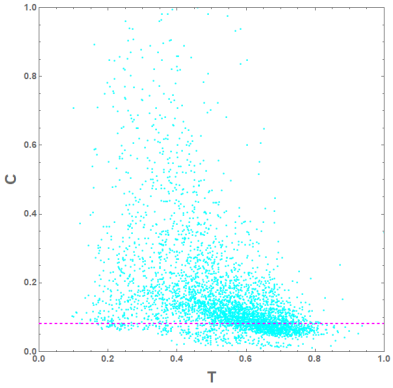
Several aspects of the data are noteworthy. The data for day and night have a remarkably similar appearance. Day or night, most of the time CO_2 concentration is not too far from ambient: 76% of the points are $\leqslant 2C_a$. Concentration near ambient depends on temperature: the denser part of the point clouds of night or day show C decreasing approximately linearly with increasing T and with an increased slope at higher day time temperatures. However, the most striking aspect of the data are the large and frequent, though irregularly timed, excursions of concentration above ambient (over $12C_a$). In the CO_2 model equation the large excursions correspond to the



(a) Concentration vs.Temperature from glasshouse data.



(b) Data in panel (a) only for night.



(c) Data in panel (a) only for day.

Figure 32: Concentration-Temperature dynamical system from glasshouse data from midnight 6th December to 16:00 21st July, points at half hour intervals. Colors indicate day (cyan) or night (black) based on the coincidentally measured light intensity in the glasshouse with 20 W-m^{-2} the boundary light intensity between day and night. Dashed magenta line indicates ambient CO_2 concentration.

fixed point losing stability. While a day time loss of stability can be explained by lowering plant uptake of CO_2 sufficiently, at night the only avenue to losing stability is a plant respiration function with an positive slope. It seems simplest to suppose that both day and night irregularly timed respiratory eruptions from the plants destabilizes the glasshouse system. That the reconstructed data in figure 31 is a fractal suggests that any such respiratory eruptions are not random events, rather that there is more mathematical structure and dynamics behind them. It also suggests the importance of the transpiration function modelling and in particular modelling what might trigger any such respiratory eruptions.

D Modelling Yield Details

A natural approach would be to consider a tomato yield proportional to the photosynthesis, i.e. to the term f_{absorb} in equation (26), as this corresponds to the main way in which tomato plants gain energy. However, there are several challenges related to the mathematical modelling of CO_2 adsorption that we were only partially able to solve, as already explained in section 3. Furthermore, tomatoes grow at a time scale of several weeks, while the models describes in the previous sections referred to a time scale of hours. In order to overcome these obstacles, we have developed a simple model from scratch independently to the previous analysis of temperature and CO_2 .

Let us assume that a single tomato that flowers on day j with mass m_0, and its mass $m^j(t)$ grows following the logistic equation

$$\begin{cases} \dfrac{dm^j}{dt} = \beta(L, T, C)\, m^j \left(1 - \dfrac{m^j}{\tilde{K}}\right) & t \in [j, j+42] \\ m^j(j) = \tilde{m}_0 \end{cases} \tag{34}$$

where the growth rate β in principle depends on the light radiation L, on the temperature T and on the CO_2 concentration C. We then assume that tomatoes are harvested after 6 weeks, i.e. at $t = j + 42$ days. If the number of tomatoes harvested per m^2 is N and all the tomatoes follow the

same growth law given by (34), then the tomato mass harvested per m^2 is $M(j+42) = N\,m^j(42)$. In other words, the harvest at a given day depends on the conditions of the previous 42 days. The total mass per m^2 (of harvested tomato fruit) is a quantity we observe (and is the main quantity of interest); to reorganise, the harvest on day τ is thus

$$M(\tau) = N m^{\tau-42}(42).$$

Observe that we can multiply equation (34) by N (which we assume constant) to obtain

$$\begin{cases} \dfrac{d(Nm^j)}{dt} = \beta(L,T,C)\,Nm^j\left(1-\dfrac{Nm^j}{K}\right) \\ m^j(j) = N\tilde{m}_0 =: m_0 \end{cases} \tag{35}$$

Therefore, the important parameters for the model are $K = N\tilde{K}$ and m_0.

Note that as yield data are discrete, it is helpful to instead think of this model as a difference equation, i.e.,

$$\begin{cases} m^j(j+t+1) = m^j(j+t) + \beta(L,T,C)\,m^j(j+t)\left(1-\dfrac{m^j(j+t)}{\tilde{K}}\right) & t \in [0,42] \\ m^j(j) = \tilde{m}_0 \end{cases} \tag{36}$$

We now need to find a suitable expression for the growth rate β. Temperature and CO_2 concentration are controlled and therefore do not vary considerably around the year. On the other hand, light radiation is not controlled and empirical evidence suggests that this is the main source of yield variations (see figure 33; observe that yield variations are delayed with respect to light variations, justifying our modelling approach). Greenhouse experiments in the literature suggest ([8]) that photosynthetic rates increase with light intensity, before levelling off, suggesting a Michaelis-Menten shaped functional form would be appropriate for this relationship. We therefore begin by modelling β with a Michaelis-Menten terms dependent on light radiation:

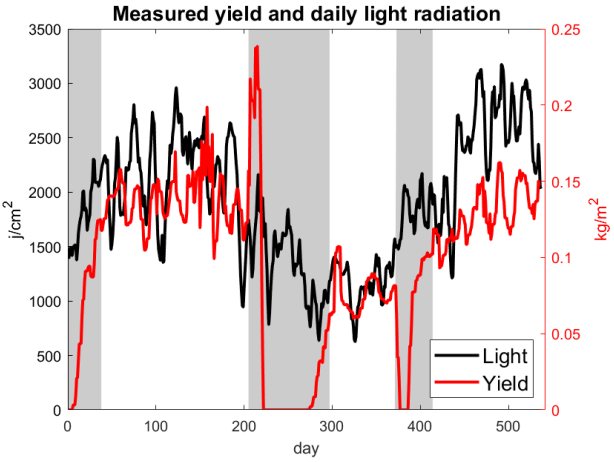


Figure 33: Tomato yield and total daily radiation. Both data have been averaged over a week in order to smooth daily variations. The first period with 0 yield correspond to the period in which the glasshouse was empty. Some technical problems prevented direct measurement of light from day 378 to day 406, therefore we have used the data of the same period measured in a different year; the same problem affected the yield data, causing an apparent decrease of the yield.

$$\beta_1(L, T, C) = \frac{\gamma L}{\alpha + L} \tag{37}$$

More relevant are models that also depend on temperature and CO2 levels. We anticipate (supported by the literature) temperature and CO_2 are likely to also impacts growth of tomatoes (and thus their yield), and so we consider a variant of β in which they are included. Based on the experimental results of [8], we select a Michaelis-Menten shaped functional form for CO_2 and a quadratic shape for temperature, giving:

$$\beta_2(L, T, C) = \frac{\gamma LC}{(\alpha_L + L)(\alpha_C + C)}(-(T - T_0)^2 + c), \tag{38}$$

Parameter	Prior	Notes
K	Exponential(1)	Represents the carrying capacity in kg/m^2.
γ	Exponential(1)	Prior chosen as positive and regularised towards zero.
α (also α_L,α_C)	Exponential(1)	Prior chosen as positive and regularised towards zero.
T_0 and c	Normal(1.5, 0.3)	Chosen based on known relationships
σ	Half Cauchy(0, 2.5)	A standard prior for error terms.

Table 5: Yield model parameters and prior distributions.

noting that the value of γ would be different to the above, and would require parameter estimation along with α_L, α_C, T_0 and c.

D.1 Parameter estimation and indicative model performance

While this yield model is quite simple and thus relatively preliminary, we demonstrate its indicative performance by performing Bayesian inference on the model parameters (as in other sections). We estimate tomato yield on (discrete) harvest days for two different varieties of tomato (with 131 and 157 observed days of yield, respectively); a separate set of parameters was estimated for each variety on the assumption that they may respond differently to climate inputs (and as their yields are not expected to be the same). Note that we fit the discrete version of the model, and use daily total solar radiation, daily mean internal temperature, and daily average CO_2 (where the average is weighted by solar radiation; i.e., CO_2 measurements when there is a large amount of solar radiation get more weight), and each of these measurements is scaled by dividing by the annual mean (giving an annual scaled mean of 1). We add Gaussian noise to account for potential variation due to factors not included in the model (including, in particular, harvesting effort), and initialise the model with initial tomato mass $m_0 = 0.01 kg/m^2$.

Therefore, for variety v harvested on day T we have

$$Y_v(T) \sim N(M(T), \sigma^2).$$

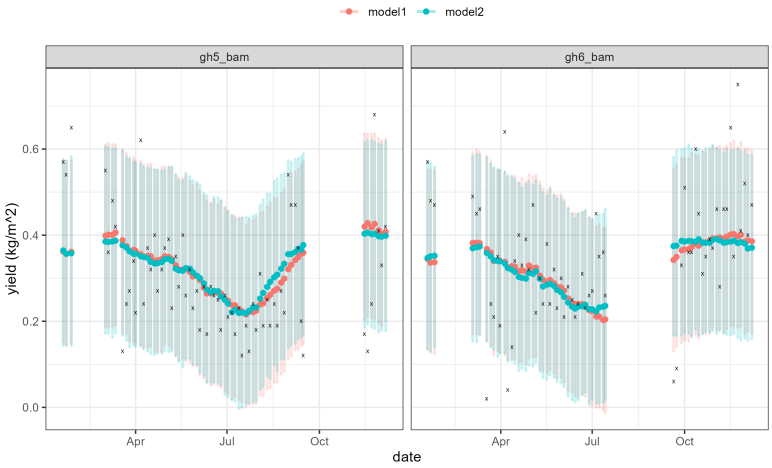


Figure 34: Posterior predictive distribution from simple yield model with β_1 (model 1; light dependence only) and β_2 (model 2; light, CO_2 and temperature dependence) against observed yield (black crosses) for one variety of tomato across two greenhouses. Posterior median (coloured dots) and 90% credible intervals (errorbars) are shown. All observations of yield from December to the following December are show where data are present and 42 days of climate observations are available.

Inference was performed using Hamiltonian Monte-Carlo using the Stan probabilistic programming language [29], accessed using the rstan library ([28]) in R.

Figures 34 and 38 compare the results of both models (model 1 driven only by light radiation, and model 2 driven by light, CO_2 and temperature) with the measured yield for two varieties of tomato across two glasshouses. Both models produce generally similar predictions, with some subtle variations, suggesting that light radiation is the primary driver of yield but that CO_2 has some ancillary impact. These numerical results show that even this most simple model is capable of reproducing the qualitative behaviour of an established tomatoes cultivation, despite many simplifying assumptions.

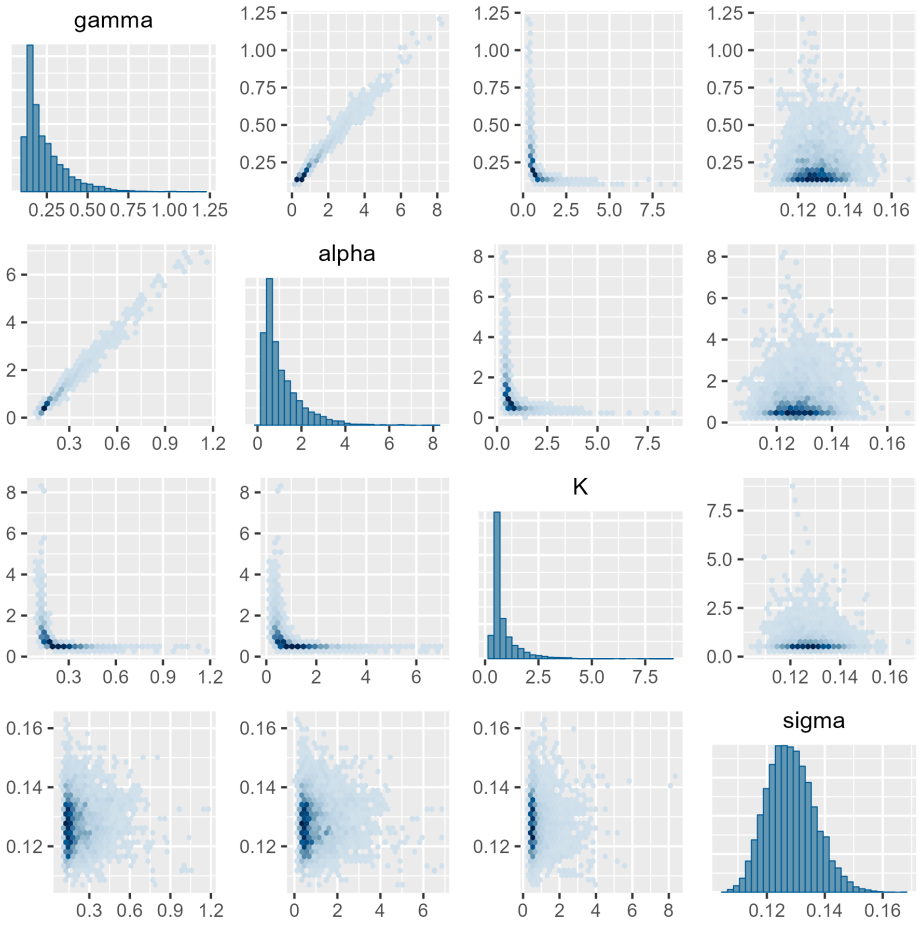


Figure 35: Posterior distribution for parameters in the simple yield model with β_1 (light dependence only) (corresponding to 34).

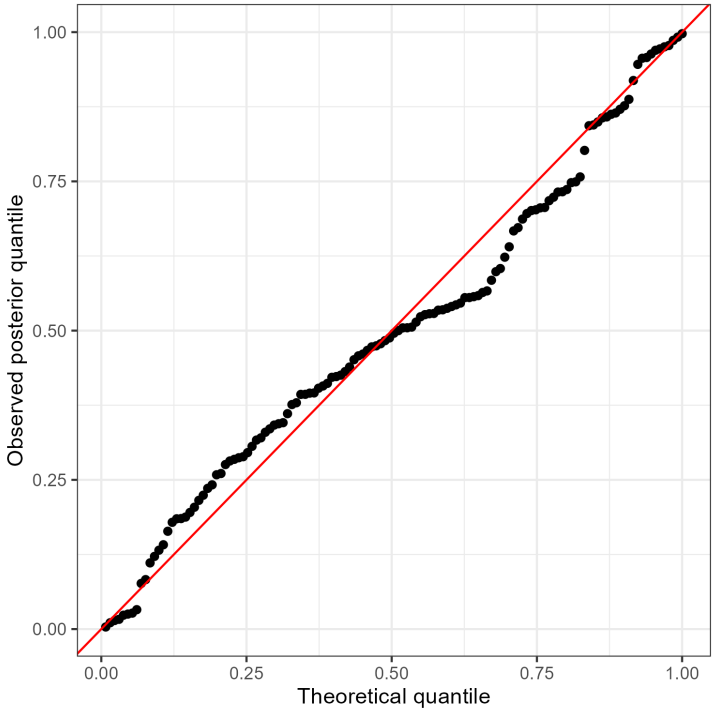


Figure 36: Calibration plot for posterior predictive distribution for the simple yield model with β_1 (light dependence only) (corresponding to 34). Observed quantiles indicate where in the posterior distribution the true observed value lies; a perfectly calibrated model would produce a straight line.

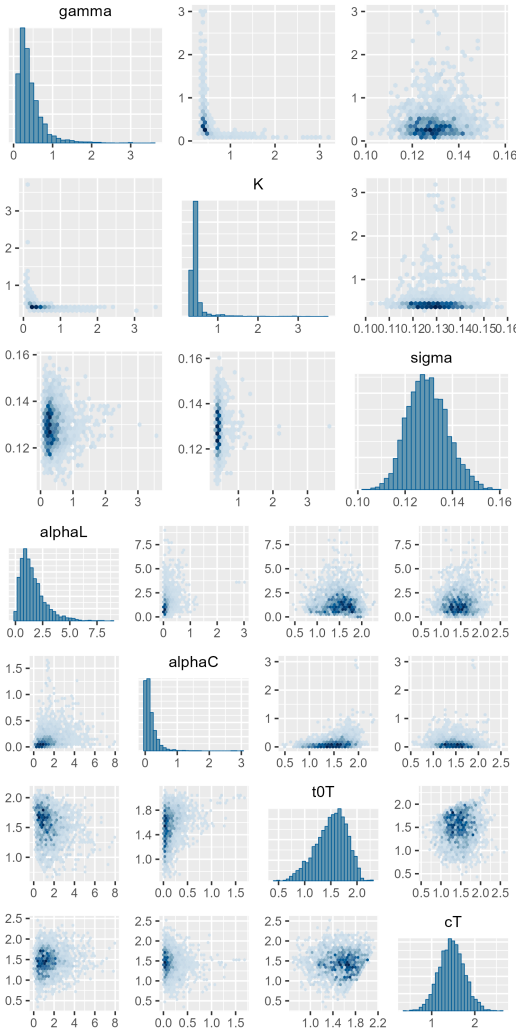


Figure 37: Posterior distribution for parameters in the simple yield model with β_2 (light, CO_2 and temperature dependence) (corresponding to 34).

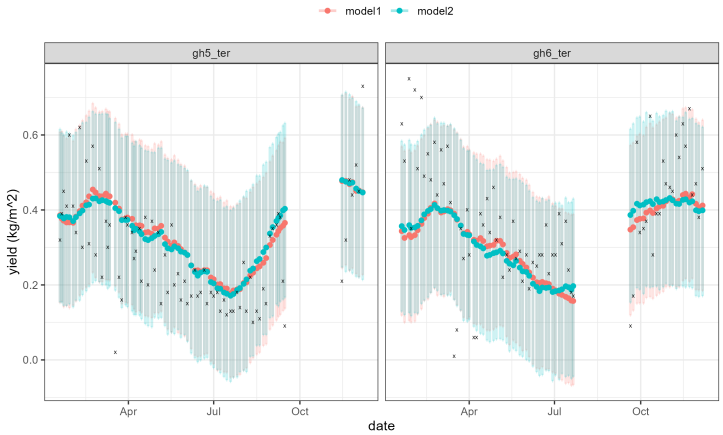


Figure 38: Posterior predictive distribution from simple yield model with β_1 (model 1; light dependence only) and β_2 (model 2; light, CO_2 and temperature dependence) against observed yield (black crosses) for the other variety of tomato across two greenhouses. Posterior median (coloured dots) and 90% credible intervals (errorbars) are shown. All observations of yield from December to the following December are show where data are present and 42 days of climate observations are available.

Most observations fall within the 90% posterior predictive intervals, and the predictions are well calibrated (Figures 36 and 40). Posterior parameter estimates corresponding to these 4 figures appear in Figures 35, 39, 37 and 41.

In the posterior distributions, we observe substantial posterior correlations (particularly between γ and α in model 1 and γ and α_L in model 2). The noise component is estimated to be quite large (relative to the scale of the observed yield), which is not surprising given that our model does not include sampling effort which may account for substantial variations in harvesting. Both models produce similar marginal estimates for σ (Tables 6 and 7). α_L in model 2 is analogous to α in model 1, and both parameters have posterior means

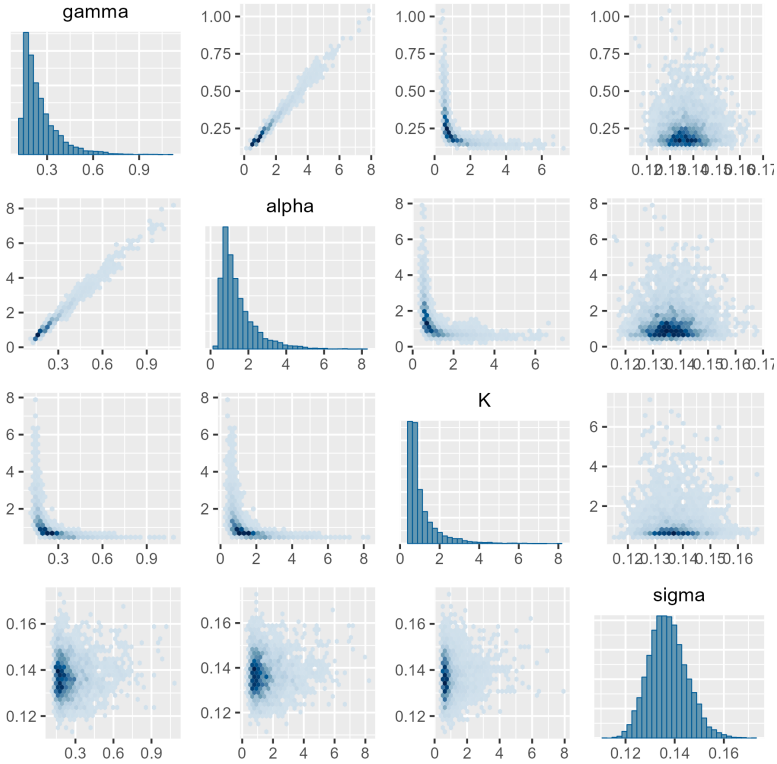


Figure 39: Posterior distribution for parameters in the simple yield model with β_1 (light dependence only) (corresponding to 38).

consistently above 1 (though with large standard deviations); in contrast, α_C is estimated to be quite small. For the parameters governing the temperature relationship in model 2, the marginal posterior distributions are very similar to the priors. K and γ are most variable in the posterior distributions; they are clearly different to the priors, but have a complex relationship indicative of possible identifiability challenges. In future work it would be useful to build more informative priors on K based on historical yield data.

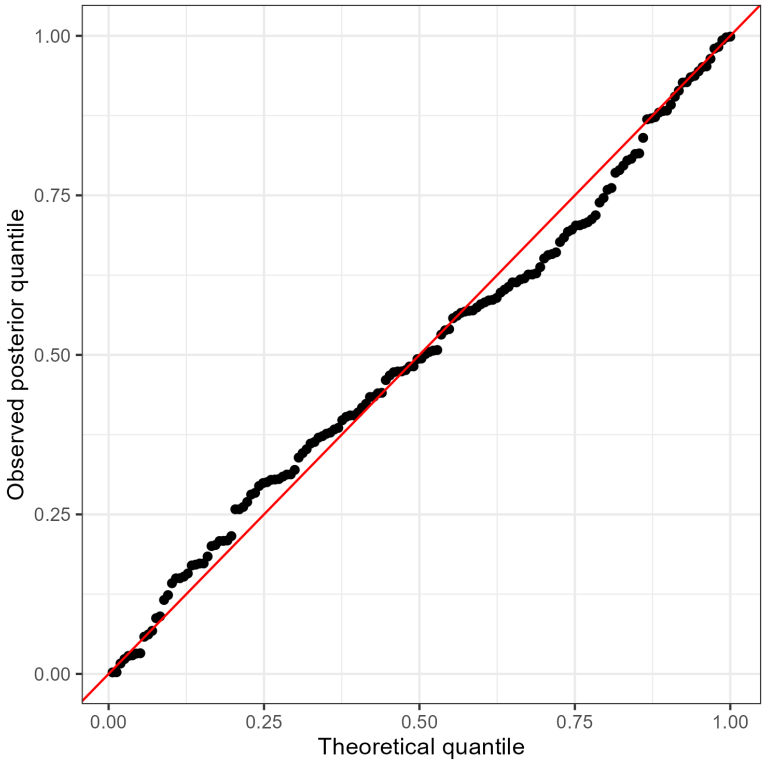


Figure 40: Calibration plot for posterior predictive distribution for the simple yield model with β_1 (light dependence only) (corresponding to 38). Observed quantiles indicate where in the posterior distribution the true observed value lies; a perfectly calibrated model would produce a straight line.

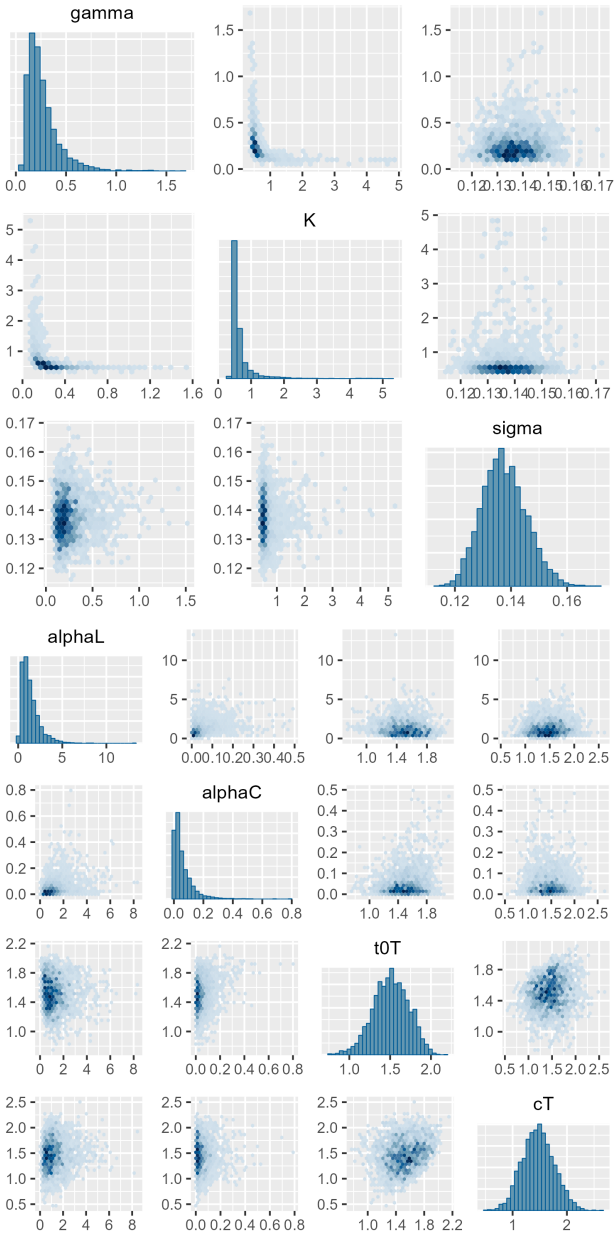


Figure 41: Posterior distribution for parameters in the simple yield model with β_2 (light, CO_2 and temperature dependence) (corresponding to 38).

	Mean	SD	2.5%	97.5%
Model 1				
γ	0.24	0.13	0.12	0.59
K	0.85	0.67	0.39	2.82
σ	0.13	0.01	0.11	0.15
α	1.10	0.91	0.25	3.57
Model 2				
γ	0.43	0.33	0.10	1.37
K	0.47	0.22	0.36	1.06
σ	0.13	0.01	0.11	0.15
α_L	1.59	1.15	0.18	4.62
α_C	0.16	0.20	0.00	0.66
T_0	1.51	0.29	0.86	1.99
c	1.46	0.29	0.89	2.05

Table 6: Marginal posterior summary for both models for one variety of tomato.

	Mean	SD	2.5%	97.5%
Model 1				
γ	0.26	0.12	0.14	0.59
K	1.07	0.78	0.48	3.31
σ	0.14	0.01	0.12	0.15
α	1.46	0.96	0.48	4.11
Model 2				
γ	0.27	0.17	0.09	0.70
K	0.67	0.43	0.42	1.82
σ	0.14	0.01	0.12	0.15
α_L	1.38	1.02	0.24	4.04
α_C	0.06	0.07	0.00	0.24
T_0	1.51	0.22	1.06	1.90
c	1.47	0.29	0.93	2.05

Table 7: Marginal posterior summary for both models for the other variety of tomato.

Author addresses

1. **Brodie Lawson**, School of Mathematical Sciences, Queensland University of Technology, Australia;
ARC Centre of Excellence for Plant Success in Nature and Agriculture, Australia
mailto:b.lawson@qut.edu.au
orcid:https://orcid.org/0000-0003-1317-5988

2. **Robert Cope**, School of Science & Technology, The University of New England, Armidale NSW 2350, Australia.
mailto:robert.cope@une.edu.au
orcid:https://orcid.org/0000-0002-2048-858X

3. **Brendan Harding**, School of Mathematics and Statistics, Victoria University of Wellington, New Zealand.
mailto:brendan.harding@vuw.ac.nz
orcid:https://orcid.org/0000-0002-6755-9998

4. **Jessica Crawshaw**, Oxford University
mailto:jessica.crawshaw@maths.ox.ac.uk
orcid:https://orcid.org/0000-0003-0797-1144

5. **Mark McGuinness**, School of Mathematics and Statistics, Victoria University of Wellington, New Zealand.
mailto:Mark.McGuinness@vuw.ac.nz
orcid:https://orcid.org/0000-0003-1860-6177

6. **Tal Kanety**, Costa Group Tomato Category
mailto:tal.kanety@costagroup.com.au

7. **Keith Chambers**, Oxford University
mailto:keith.chambers@maths.ox.ac.uk
orcid:https://orcid.org/0000-0001-6878-8007

8. **Keith Chan**, School of Mathematics, Monash University, Clayton VIC 3800, Australia.

mailto:keith.chan@monash.edu
orcid:https://orcid.org/0009-0005-8050-7171

9. **Guy Metcalfe**, Faculty of Science, Engineering & Technology, Swinburne University of Technology, AUSTRALIA
mailto:gmetcalfe@swin.edu.au
orcid:https://orcid.org/0000-0003-4679-8663

10. **David Morselli**, Department of Mathematical Sciences "G. L. Lagrange", Politecnico di Torino, ITALY;
Department of Mathematics, School of Science, Computing and Engineering Technologies, Swinburne University of Technology, AUSTRALIA;
Department of Mathematics "G. Peano", Università di Torino, ITALY.
mailto:david.morselli@polito.it
orcid:https://orcid.org/0000-0002-9321-2701

11. **Mark Flegg**, School of Mathematics, Monash University, Clayton VIC 3800, AUSTRALIA.
mailto:Mark.Flegg@monash.edu
orcid:https://orcid.org/0000-0002-4697-4789