LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR GEOWISSENSCHAFTEN

DEPARTMENT FÜR GEOGRAPHIE

MASTERARBEIT

Umweltsysteme und Nachhaltigkeit
Monitoring, Modellierung und Management

Wissenschaftliche Arbeit zur Erlangung des akademischen Grades Master of Science

„Tomatenertrag unter verschiedenen Beschattungsgraden in einem agrivoltaischen Gewächshaus in Südspanien - Analyse und Kalibrierung des reduced state-variable Tomato Growth Model (TOMGRO).“

‘Tomato yield under different shading levels in an agrivoltaic greenhouse in southern Spain - Analysis and calibration of the reduced state-variable Tomato Growth Model (TOMGRO).’

Verfasser:
Julian Kornas
julian.kornas@campus.lmu.de

Betreuer:
Prof. Dr. Ralf Ludwig
Mitbetreuung:
Anna Kujawa

Datum: 17.07.2024

Zusammenfassung

Gewächshäuser in Südspanien bieten ein hohes Potential für agrivoltaische Lösungen. Sie versprechen mit einer effizienteren Landnutzung ökologische Vorteile und ökonomische Gewinne im Vergleich zu konventionellen Gewächshausbetreibern. Dennoch gibt es noch ein unzureichendes Verständnis über die Auswirkungen von verringertem Sonnenlicht auf das Pflanzenwachstum. Aus diesem Grund müssen Modelle entwickelt und kalibriert werden, die sowohl die Pflanzenproduktion als auch die Stromerzeugung integrieren, um den lukrativsten und nachhaltigsten Kombinierungsansatz zu identifizieren. Die vorliegende Masterarbeit untersucht den Tomatenertrag unter verschiedenen Beschattungszonen (0%, 30%, 50%) durch auf dem Dach installierte PV-Dummies in einem typischen „raspa y amagado“ Gewächshaus im Süden Spaniens. Dies beinhaltet die Messung, Analyse und Vergleich der Erträge in den unterschiedlichen Zonen. Außerdem wurde das „Reduced State-Variable Tomato Growth Model” von Jones et al. (1999) als biophysikalischer Modellansatz kalibriert und getestet, um das Wachstum von Tomatenpflanzen unter dynamischen Bedingungen im Gewächshaus zu simulieren. Das Modell soll dann die Erträge unter den verschiedenen Beschattungsbedingungen möglichst genau vorhersagen können. Um dies zu ermöglichen, erfolgte eine Sensitivitätsanalyse mittels der „Extended Fourier Amplitude Sensitivity Test“ Methode und eine Kalibrierung mit Hilfe des „Particle Swarm Optimization“ Algorithmus. Die Bewertung der Modellqualität wird durch verschiedene Qualitätskriterien wie beispielsweise den „Root Mean Squared Error“ (RMSE) und „Mean Absolute Error” (MAE) etc. festgestellt. Die Ergebnisse zeigen eine gute Übereinstimmung zwischen den simulierten und gemessenen Erträgen in allen Zonen, wobei beispielsweise der RMSE zwischen 9,15 und 9,84 liegt. Die Anwendung des PSO-Algorithmus erwies sich als geeignet zur Kalibrierung des TOMGRO-Modells. Die Arbeit kommt zu dem Schluss, dass die Nutzung und Anpassung präziser Pflanzenmodelle zur Simulation und Vorhersage des Tomatenertrags in Gewächshäusern sehr vorteilhaft ist. Nach Anpassung mit einer zusätzlichen Validierungsperiode wird für zukünftige Arbeiten empfohlen, das angepasste Pflanzenwachstumsmodell in ein integriertes agrivoltaisches Gewächshausmodell zu implementieren, um den wirtschaftlichsten Ansatz zwischen Strom- und Pflanzenproduktion zu ermitteln.

Abstract

Greenhouses in southern Spain offer great potential for agrivoltaic solutions. With more efficient land use, they promise ecological benefits and economic gains compared to conventional greenhouse operators. However, there is still insufficient understanding of the effects of reduced sunlight on plant growth. For this reason, models that integrate both crop production and electricity generation need to be developed and calibrated to identify the most lucrative and sustainable combination approach. This master thesis investigates the tomato yield under different shading zones (0%, 30%, 50%) by roof-mounted PV dummies in a typical 'raspa y amagado' greenhouse in the south of Spain. This includes measuring, analyzing and comparing the yields in the different zones. In addition, the 'Reduced State-Variable Tomato Growth Model' by Jones et al. (1999) was calibrated and tested as a biophysical modeling approach to simulate the growth of tomato plants under dynamic greenhouse conditions. The model should then be able to predict yields as accurately as possible under different shading conditions. To make this possible, a sensitivity analysis was carried out using the 'Extended Fourier Amplitude Sensitivity Test' method and calibration using the 'Particle Swarm Optimization' algorithm. The evaluation of the model quality is determined by various quality criteria such as the 'Root Mean Squared Error' (RMSE) and 'Mean Absolute Error' (MAE) etc.. The results show good agreement between the simulated and measured yields in all zones, with the RMSE, for example, lying between 9.15 and 9.84. The application of the PSO algorithm proved to be suitable for calibrating the TOMGRO model. The work concludes that the use and adaptation of accurate crop models to simulate and predict tomato yield in greenhouses is very beneficial. After adaptation with an additional validation period, it is recommended for future work to implement the adapted crop growth model into an integrated agrivoltaic greenhouse model to determine the most economical approach between electricity and crop production.

Contents

List of Figures

List of Tables

List of Equations

Acronyms

ARV - Average relative variance

DLR - German Aerospace Center

eFAST - Extended Fourier Amplitude Sensitivity Test

FSPM - Functional Structural Plant Model

GHI - global horizontal irradiance

MAE - Mean Absolute Error

PAR - photosynthetic active radiation

PDF - probability density function

PPFD - Photosynthetic Photon Flux Density

PSO - Particle Swarm Optimization

PV - photovoltaic

R^2 - Coefficient of determination

RH - relative humidity

RMSE - Root Mean Squared Errors

r-RMSE – relative RMSE

Reduced TOMGRO - Reduced State-Variable Tomato Growth Model

SALib - Sensitivity Analysis Library

Si – Sensitivity index

STi - Total effect index

Td - Daytime temperature

WM - Dry weight of the ripe tomato fruits

%SEP - Percent standard error of the prediction

Z0 - control zone or Zone 0

Z30 - Zone 30

Z50 - Zone 50

1 Introduction

The term agrivoltaic describes the combination of agriculture and photovoltaic (PV). Since years it's a topic of rising interest and research on the topic is rapidly increasing. On the one hand experimental test sites are set up at different locations all over the word, while on the other hand, modelling approaches are developed and tested, all in order to find the most economically rewarding combination of crop and PV yield. This master thesis is going to zoom into this very broad topic and focus on crop yield production and modelling for an agrivoltaic greenhouse experiment conducted in Southern Spain.

This first chapter serves as a general introduction and motivation for the topic of agrivoltaic with a focus on greenhouses and the region of Southern Spain.

1.1 The potential of agrivoltaic solutions for greenhouses in Southern Spain

Food production has changed dramatically in recent decades. The growth of the world's population, the current human-induced global warming, the increase in prosperity and the globalization of agriculture have led to an increased demand for food, which in turn has led to a more intensive use of resources such as land, energy and water. This intensive agriculture not only puts a strain on the environment through higher greenhouse gas emissions and the loss of natural habitats, but also leads to a heavy burden on limited water resources. In addition, rising energy prices can increase the cost of farming and have a direct impact on crop prices. This complex, interrelated system and its interactions and challenges is described as the global 'Food-Energy-Water Nexus' (Smajgl et al., 2016). This system requires an integrated approach to enable the sustainable use of limited resources in the future. In this context, agrivoltaic solutions offer a promising potential to realize this (Dinesh and Pearce, 2016; Smajgl et al., 2016).

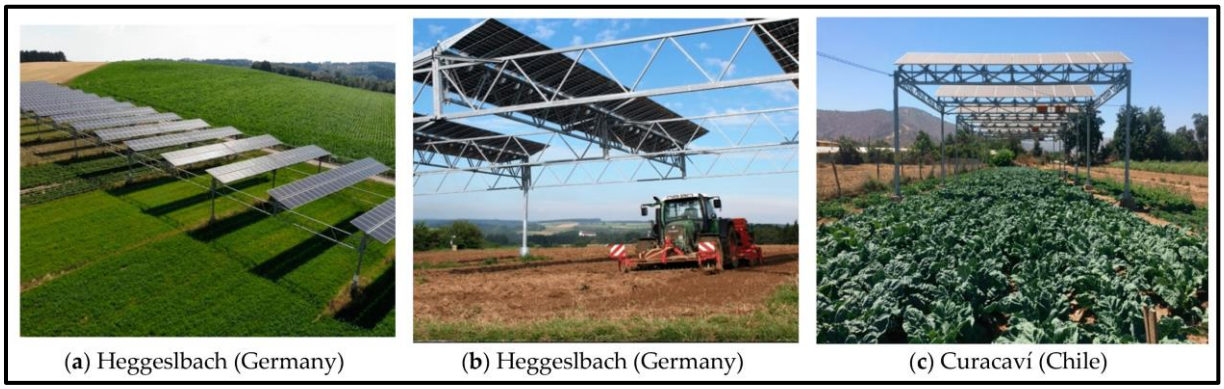

Figure 1: Experimental agrivoltaic systems installed by Fraunhofer ISE in Germany (a,b) and Chile (c) (Fraunhofer ISE, n.d.).

Figure 1 shows two open-field agrivoltaic installations implemented by Fraunhofer ISE, the German test site in Heggelsbach, already in operation since 2016, and the Chilenian test site in Curacavi, installed in 2016 (Fraunhofer ISE, n.d.). Agrivoltaic is not only limited to open-field agriculture. Nowadays, there are several

different approaches for agrivioltaic, e.g. the combination with farming, aquaculture or horticulture (greenhouses). Especially greenhouses offer a promising potential, due to the already existing infrastructure. The so-called 'Mar de Plastico' ('Sea of plastic') in the province of Southern Spain (fig. 2), for example, is a network of plastic greenhouses, which has expanded rapidly in recent years and now covers some 33,000 hectares (fig. 3). This infrastructure is ideal for agrivoltaic greenhouse solutions, as photovoltaic (PV) concepts could be installed on the existing greenhouse framework. In addition to the existing infrastructure, the province of Almería is one of the sunniest places in Europe (fig. 4), with local high irradiation levels, which makes greenhouse cultivation possible at any time of the year (AEMET, 2024). However, plant development in greenhouses highly depends on irradiance distribution among other

Figure 2: Bird's eye view of the Mar de plástico in Almería (Economía 3, n.d.).

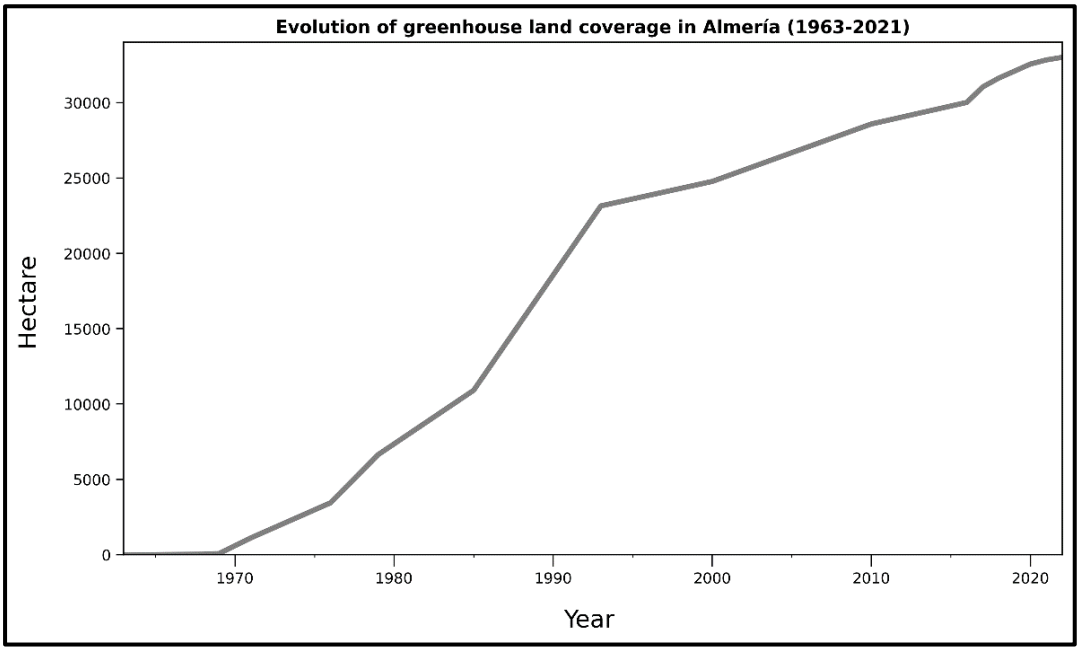

Figure 3: Evolution of greenhouse land coverage in the province of Almeria (based on data of (Instituto de Estadística y Cartografía de Andalucía, 2021; Junta de Andalucia, 2021; Junta de Andalucía, 2020)).

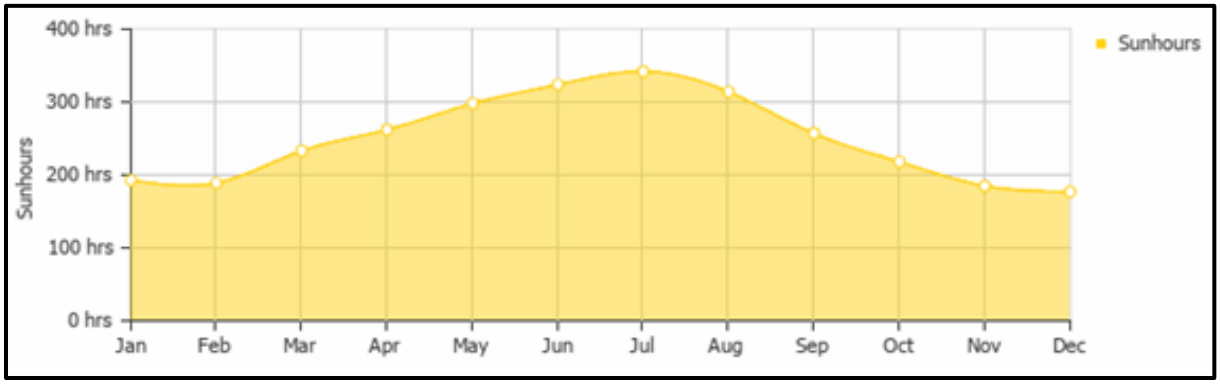

Figure 4: The monthly total of sunhours over the year in Almería (Andalucía), Spain (World Weather & Climate Information, 2024).

factors, so that too little or too much light can both harm plants (Dueck et al., 2012). In Southern Spain, for example, the irradiation levels are so high that several times throughout the year the light transmission of the greenhouse plastic covers is actively reduced by the application of chalk paint (fig. 5). This procedure is commonly known as 'blanqueo' (Valera et al., 2016). Light management is therefore necessary for optimal plant growth. Agrivoltaic concepts can actively support this light management of the grower. The shading provided by PV panels and the absorption of sunlight hitting the roof can create a more stable microclimatic environment for the plants. In addition to this ecological advantage and the land-saving aspect, the combination of plant cultivation and solar power generation also results in numerous economic benefits for the individual farmer, as agrivoltaic farms generally outperform conventional agriculture economically (Dinesh and Pearce, 2016). Overall, this innovative solution of greenhouse cultivation and simultaneous renewable energy production thus promises a more sustainable future, both locally for Southern Spain and globally, ensuring food and energy needs are met.

Figure 5: Applying white chalk paint to the roofs of a 'raspa y amagado' greenhouse.

1.2 Objectives

In order to optimally utilize or plan a agrivoltaic greenhouse system so that plants do not suffer damage, or in other words to identify the most lucrative method of combining cultivation and electricity generation to get the most economic yield from both, integrated agrivoltaic yield models need to be developed, as the influence of decreased irradiance is not yet fully understood and critical threshold for healthy plant development is unknown (Willockx et al., 2022).

This work focuses exclusively on the component of plant growth. For its modeling, a biophysical model-based approach for tomatoes was selected, which can be integrated into the overall agrivoltaic yield model after successful calibration and validation. The tomato growth model is calibrated and evaluated using measurement data from a real-scale tomato greenhouse in Southern Spain with different shading levels by PV dummies on the roof.

The research objectives of this work are, on the one hand, the analysis of tomato yield in different shading zones in an agrivoltaic plastic greenhouse. And on the other hand, this work aims to examine how well the selected tomato growth model is suitable as a tool for the simulation of tomato yield with different shading. Various measures of quality are used to interpret how well the simulation approximates the actual measured values for the different shading conditions. In this way, statements can be made about the quality of the tomato growth model and the selected calibration approach, which are crucial for further follow-up simulations with different shading variants as input parameters.

2 State of the art in agrivoltaic greenhouse crop modeling with a focus on tomatoes

In order to be able to classify the relevance and need of this work, the current state of research in the modeling of agrivoltaic greenhouse crops with a focus on tomatoes, which are influenced in their natural growth by different shading, is explained below.

First of all, previous pilot projects have shown that the design of agrivoltaic solutions must be tailored to the site-specific climatic conditions, such as variations in control strategies, climate parameters and greenhouse designs. Furthermore, their results are limited to the investigated cases and not directly transferable to more Southern locations such as Almeria (Katsikogiannis et al., 2022; Kempkes et al., 2018; Willockx et al., 2022). In Southern Europe, experimental studies have also been carried out to investigate the effects of shading on the yield and quality of greenhouse crops, using various PV configurations with different artificial shading geometries and shading ratios (Aroca-Delgado et al., 2019; Cossu et al., 2018; Cossu et al., 2020; Ezzaeri et al., 2018; López-Díaz et al., 2020; Ureña-Sánchez et al., 2012). In a so-called 'Venlo' greenhouse with a linear arrangement of PV dummies, it was found that yield production was delayed towards the end of the season with a shading ratio of over 30 %. In the study by Ureña-Sánchez et al. (2012), no reduction in tomato yield was determined with a PV coverage of 10 % arranged in a checkerboard pattern for a 'raspa y amagado' plastic greenhouse typical in Southern Spain. However, no modeling was carried out in these studies, so that a Lack on agrivoltaic studies with crop models for greenhouse yield simulation exists.

Regarding modeling of greenhouse crop yield, there are numerous studies on greenhouse crop models, but not in combination with an agrivoltaic greenhouse or with different shading by PV panels. Numerous models for the growth of greenhouse crops have been developed in the past. The reduced TOMGRO model was designed by Jones et al. (1999) and explains plant behavior under dynamic greenhouse conditions. Heuvelink (1999) developed the TOMSIM model, which focuses on the light absorption of the plant canopy. Cohen and Gijzen (1998) developed the universal growth model HORTISIM, which was useful for environmental control and greenhouse management. Zhao et al. (2019) developed SIMPLE, which is a simple but general crop model that can be used for various crops. And also Vanthoor by Vanthoor et al. (2011) is a well-structured explanatory model and uses the buffer theory. A comparison of growth models and model fusions with machine learning approaches, as in the study by Gong et al. (2023), to estimate and simulate greenhouse tomato yield has been conducted (Fink et al., 2023; Gong et al., 2023; Heuvelink and Bertin, 1994; Lin et al., 2019). The results showed that the reduced TOMGRO model has always been applicable for the greenhouse yield estimation. Consequently, numerous studies have tested the reduced TOMGRO model and investigated different sensitivity methods and calibration algorithms at real greenhouse sites in Southern Spain, Italy, central Mexico and northern Europe (Bacci et al., 2012; Gong et al., 2021; Ramírez et al., 2004; Varella et

al., 2010; Vazquez-Cruz et al., 2014). The best results were obtained with the Extended Fourier Amplitude Sensitivity Test (eFAST) and the Particle Swarm Optimization (PSO) calibration algorithm. For this reason, it was decided to use these methods in this work. The studies indicated that the differences in optimal control strategies, local climate parameters and greenhouse structures mean that Tomgro requires more accurate parameter identification and calibration tailored to existing greenhouse conditions, which will also be taken care of in the presented work.

This work presents the first application of a crop growth model for tomato yield prediction in an agrivoltaic greenhouse.

3 Material and methods

In this chapter, the material and the methods are explained and the experimental setup on the one hand and the methodology of model development on the other are clarified.

3.1 Experimental setup

In this subchapter, the study area, the period of the growing season and the construction of the greenhouse are presented in order to better understand the conditions described in 3.1.4. In addition, the monitoring devices and the information about the tomato plant are explained.

3.1.1 Study area and time frame

The experiment is performed in a so-called ‘raspa y amagado’ plastic greenhouse in the eastern region of Almería, Southern Spain, which are simple constructions covered completely by light transmitting plastic materials (fig. 7). The experiment took place during the winter growing season from September to March. The tomato plants in the greenhouse were transplanted on 16.09.2023. The plants were then removed from the greenhouse on 18.03.2024 (cf. fig. 6). Accordingly, the data set contains information of this time period, resembling one typical crop cycle.

Figure 6: Pictures from different points in the growing period. They show the planted tomatoes in the first weeks (left) up to the harvested mature fruits in the right picture.

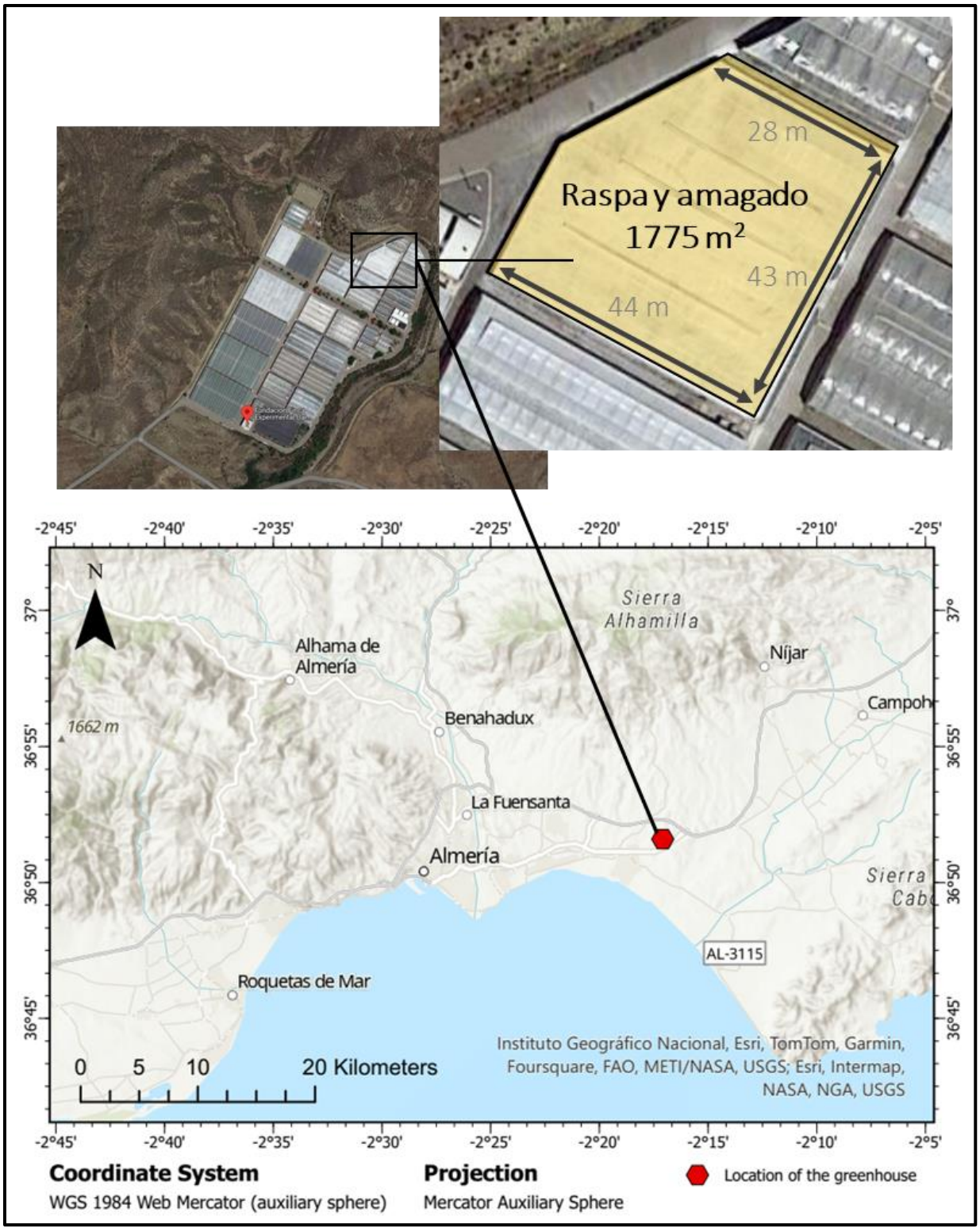

Figure 7: Study area and location of the greenhouse.

3.1.2 Greenhouse construction

The greenhouse is owned by Fundación Finca experimental UAL-Anecoop, with a total area of 1775m². The experiment was carried out in the facilities of the UAL-ANECOOP foundation while the foundation performed the greenhouse operation and maintenance. Raspa y amagado greenhouses are the most common greenhouse type in the region (Valera et al., 2016). The plastic cover is a multilayered polyethylene with

200µm thickness and 89% light transmission between 400–700nm. To achieve a shading effect of PV modules, black and opaque plastic panels (100 x 170 cm) are attached to the greenhouse roof. The panels are arranged in a checkerboard pattern to increase light homogeneity on an annual average. The greenhouse is divided in three zones with different artificial shading levels or cover ratio: 0% (control zone), 30% and 50% (fig. 8 and 9). The position of the test zones was determined by Kujawa et al. (2023) in order to exclude boundary effects and overlaps with shadow throughout the year. The 30% and 50% zones are distributed diagonally, the control zone is located in the southwestern quarter. This enables a direct comparison between the zones so that the differences between the shaded zones and the usual unshaded control zone can be analyzed. There is no physical separation of the zones in the experimental greenhouse.

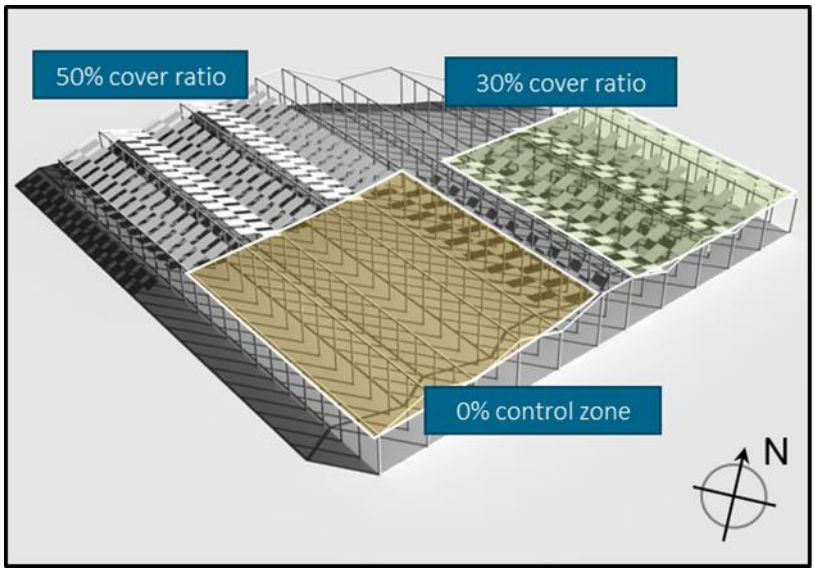

Figure 8: Virtual copy of the experimental greenhouse with three different zones and a remaining area that is not observed. The virtual copy was generated by (Kujawa et al., 2023).

Figure 9: Pictures from inside the greenhouse. a. shows zone 0 as control zone, b. shows zone 30 and c. shows zone 50 with 50 % coverage or shading.

3.1.3 Microclimatic measuring and monitoring devices and data processing

In the three test zones, global horizontal irradiance (GHI), temperature and relative humidity are measured continuously at a height of 2.8 meters, which is the maximum canopy height of the tomato plants during the cultivation, with a time resolution of one

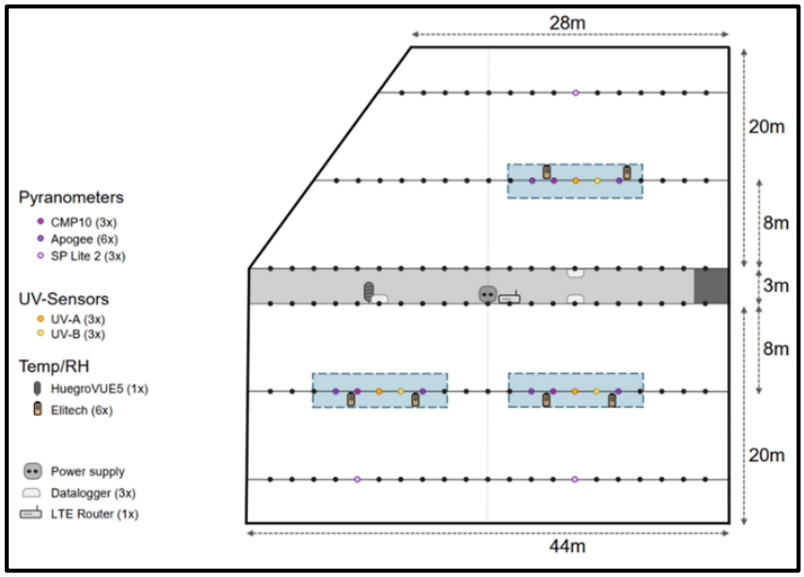

Figure 10: Locations of the monitoring devices.

minute. Each zone is equipped with four pyranometers for irradiation measurements and two sensors for temperature and humidity measurements. The sensors were positioned in such a way to ensure uniform shading that is independent of variations in the solar zenith and the associated shadow shifts. Two of the different pyranometer types (CMP10 and SP-110) were installed in the center of each zone and one pyranometer (SP Lite2) was placed closer to the edge of the zones to monitor the boundary conditions caused by incoming radiation from the side walls. The

Figure 11: Microclimatic measuring and Monitoring devices. a. shows SP-110, b. shows CMP10, c. shows SP Lite2, d. shows Elitech RC-51.

pyranometers were calibrated in advance for the experiment (Kujawa et al., 2023). The SP-110 is manufactured by Apogee Instruments and measures solar radiation (GHI) incident on a horizontal surface. The sensor contains a silicon-cell-photodiode and measures in the spectral range between 360 and 1120 nm with a relative error of less than 3% when measuring irradiance at 1000 W/m^{-2} (Apogee Instruments, 2022). The CMP10 Class A pyranometer from Kipp & Zonen has a spectral range from 285 to 2800 nm. The incoming solar radiation produces a continuous millivolt output that is converted to irradiance in W/m² in a data logger using the calibrated sensitivity. The relative error for the CMP10 pyranometer due to temperature sensitivity is less than 1% (OTT HydroMet B.V., 2023). The SP Lite2, also developed by Kipp & Zonen, measures in a spectral range from 400 to 1100 nm. The sensor measures the solar

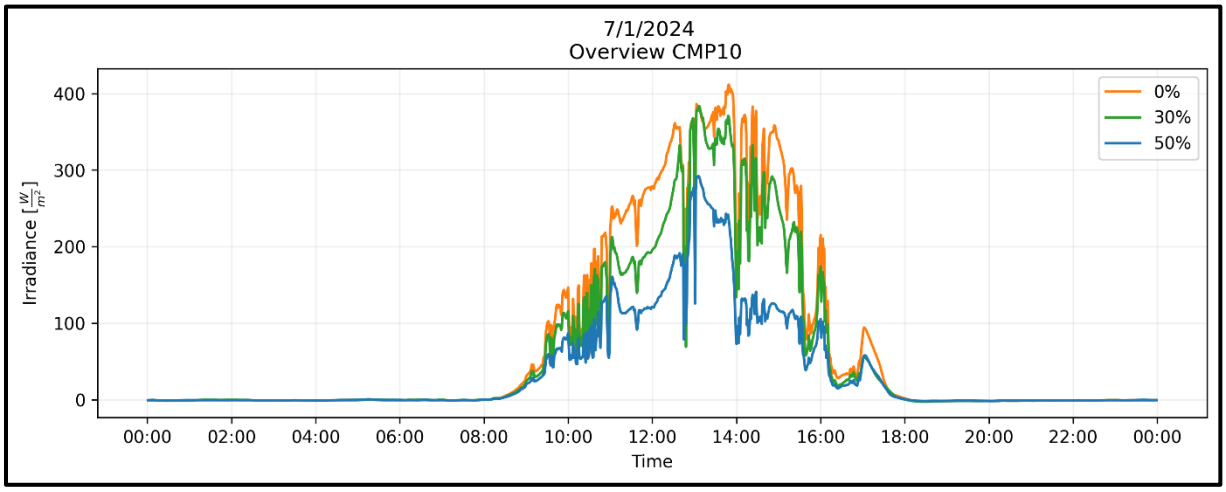

Figure 12: Overview of all CMP10 GHI measurements in minutely time resolution for 07.01.2024 as an example. It also shows the radiation differences between the three shading zones.

energy received from the entire hemisphere and uses a photodiode detector that generates a voltage proportional to the incident radiation. The relative error for the SP Lite2 pyranometer due to temperature response is less than 0.15% per °C (OTT HydroMet B.V., 2016). All three Pyranometers per zone are connected to a CR1000x data logger, which receives the mV signals. To measure temperature and relative humidity (RH), two Elitech RC-51 devices were used per zone, which were placed adjacent to the CMP10 pyranometer and the Northern Apogee pyranometer (as indicated in figure 10). The device can measure precisely (uncertainty of ± 0.5 at -20°C/+40°C; ± 1.0 at other range or ±3% RH at 25 °C, 20% ~ 90% RH; ±5% RH at other range) and in different time resolutions depending on the setting. Data acquisition is performed manually by connecting the sensor to a computer (Elitech Technology, n.d.). All these measuring and monitoring devices guarantee continuous data monitoring of the microclimate in the greenhouse with a minutely time resolution. An example of the minute-by-minute GHI recording from the CMP10 in the respective zones can be seen in Figure 12.

In order to be able to use the data for the reduced TOMGRO model (see 3.2), the data must first be prepared accordingly. First of all, TOMGRO requires hourly PPFD values as irradiance input data. PPFD describes the perceived irradiance in the photosynthetic active radiation (PAR) spectrum. The PAR spectrum includes wavelengths from 400nm to 700nm, as depicted in Figure 13. Therefore, measured GHI [W/m^2] data is converted into PPFD [$\mu mol/m^2$ s] using the linear regression model

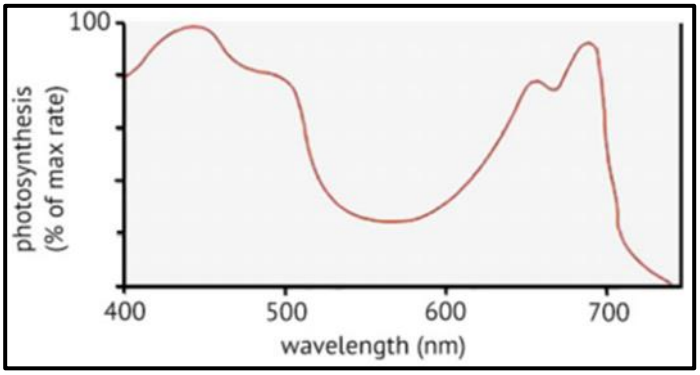

Figure 13: Action spectrum for photosynthesis (Hamblin et al., 2019).

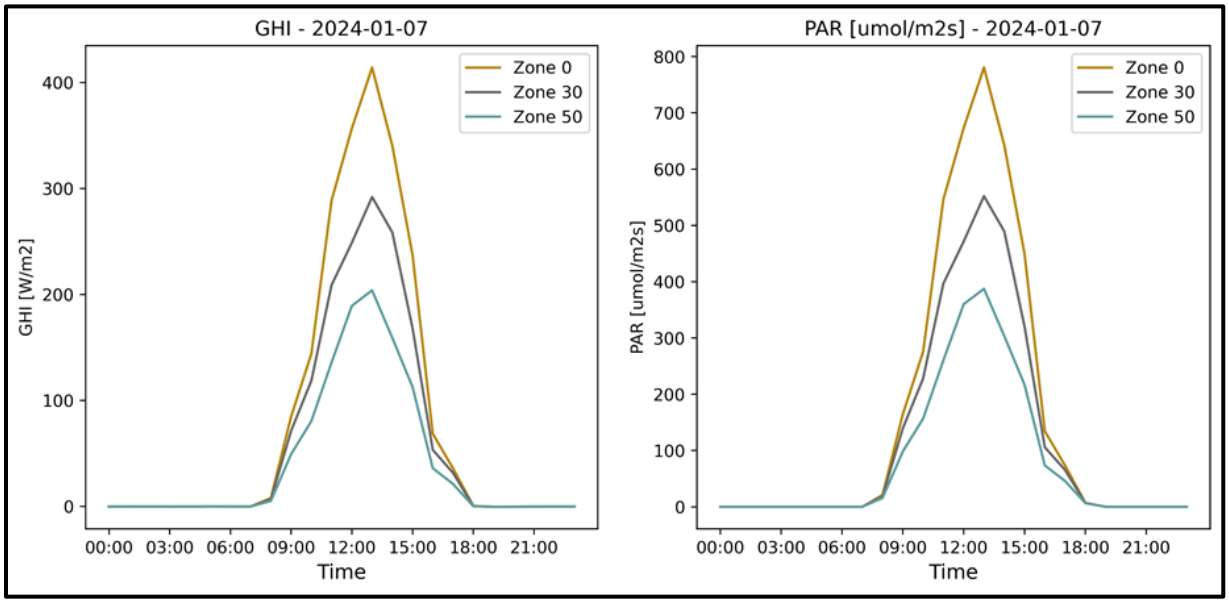

Figure 14: Comparison between hourly mean GHI and hourly mean PPFD or PAR for the selected day 07.01.2024 as an example.

according to Vindel et al. (2018). In the study, PPFD is estimated from GHI measurements using linear regression models based on strong correlations between these two variables and developed for specific regions of Spain. Figure 14 shows the GHI-to-PAR conversion for an example day. In addition, another important step is necessary. The model requires hourly temperature and PPFD values as input, so the minutely values were first averaged to hourly values. The average value from all sensors was then calculated for each zone in order to obtain one value per hour and per zone.

With the help of GHI-to-PAR conversion and averaging to hourly values, the inputs required for reduced TOMGRO were correctly prepared for all three shading zones.

3.1.4 Greenhouse conditions

In order to better interpret and understand the results with regard to the research questions of tomato yield and also the results of the modeling, it is important to know which microclimatic conditions actually existed in the greenhouse during the growing season. The radiation is shown firstly (fig. 15). The difference between the various zones can be clearly seen. There is always a higher PPFD in zone 0 than in the zones with shading, which show varying deviations. In zone 30 the difference is up to -60 µmol/m²/s and in zone 50 even up to -140 µmol/m²/s. In relative terms, there is a reduction of up to 25 % in zone 30 and up to 50 % in zone 50.

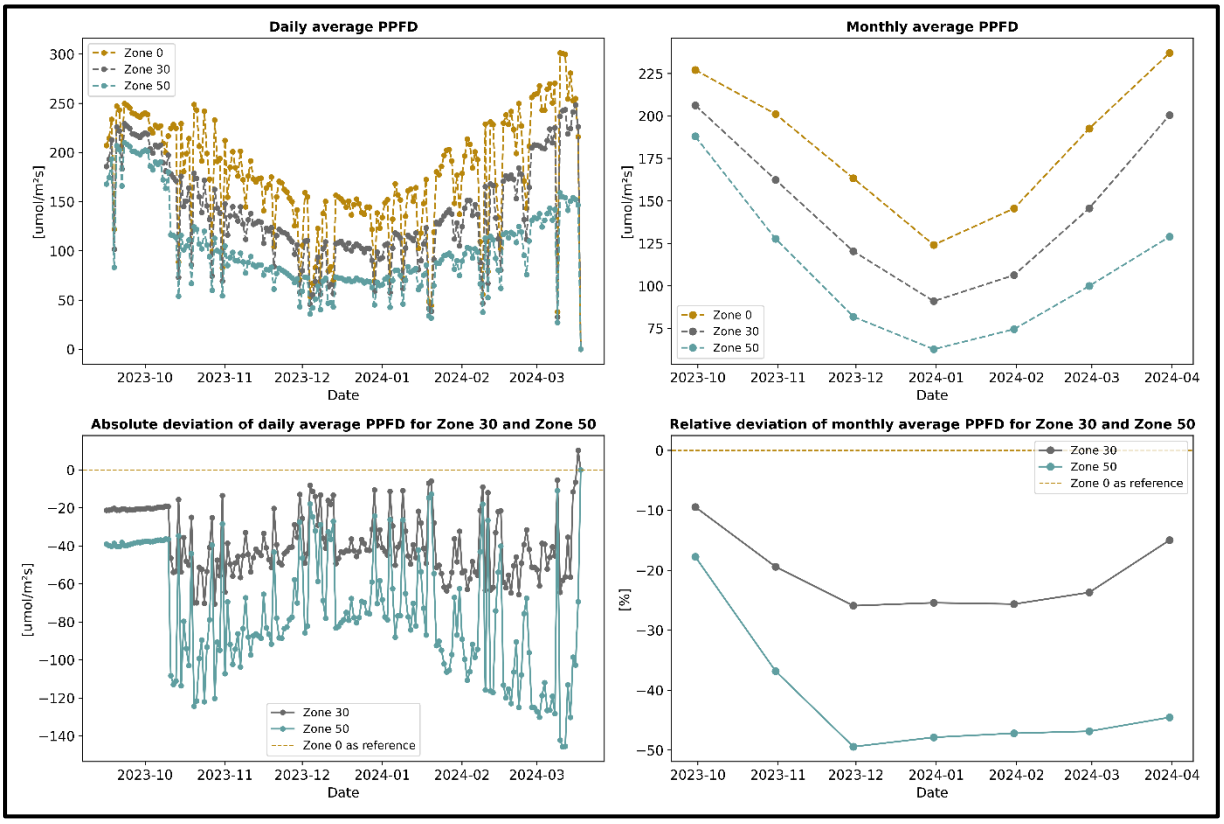

Figure 15: Existing PPFD measured values for each shading zone and their deviations from the control zone in the greenhouse during the entire growing season.

The differences in temperature are not quite as clear as in radiation (fig. 16). Zone 30 and Zone 0 are very similar this time and in some cases, there are also higher temperatures in Zone 30 than in the control zone. It is noticeable that zone 50 is always

cooler compared to the other two zones. Compared to the control zone, temperatures there are up to just under 2 °C or up to almost 0.5 % (in Kelvin) cooler in relative terms.

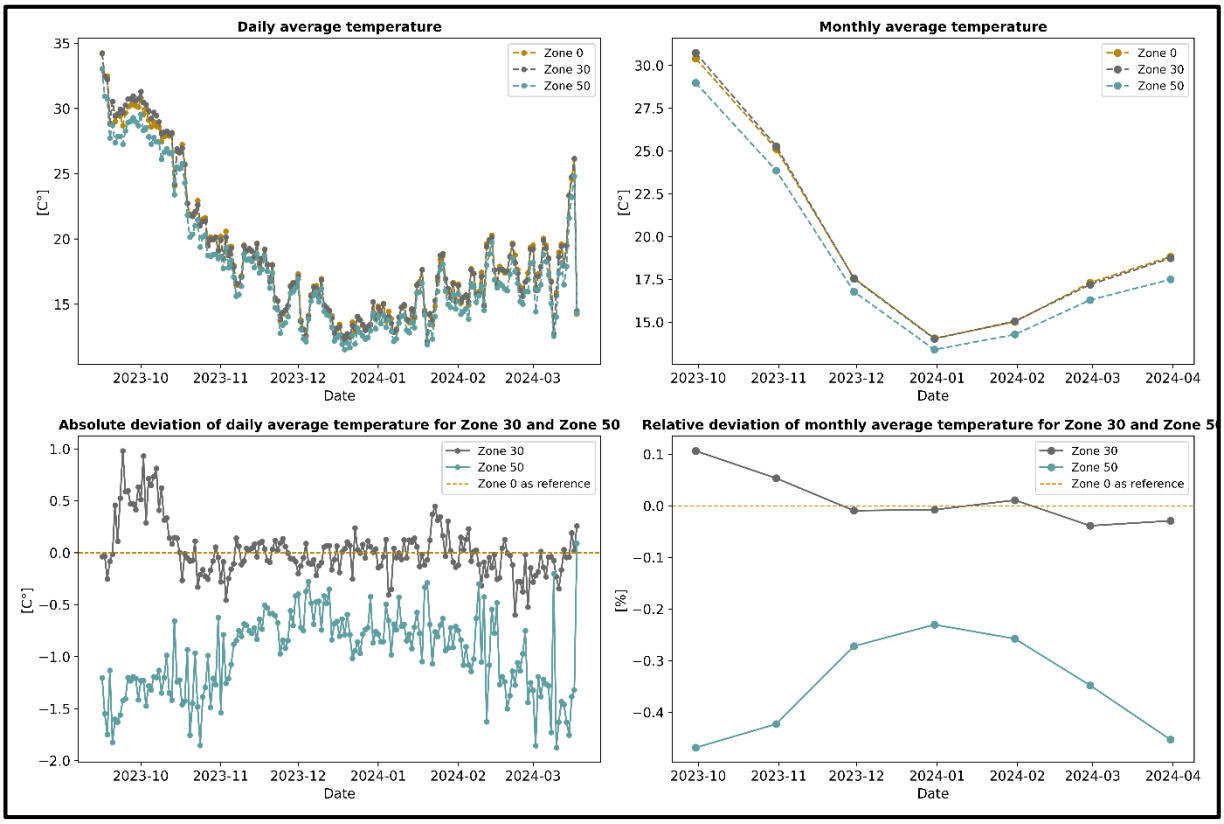

Figure 16: Existing temperature measured values for each shading zone and their deviations from the control zone in the greenhouse during the entire growing season. To calculate the relative deviation, it was previously converted to Kelvin.

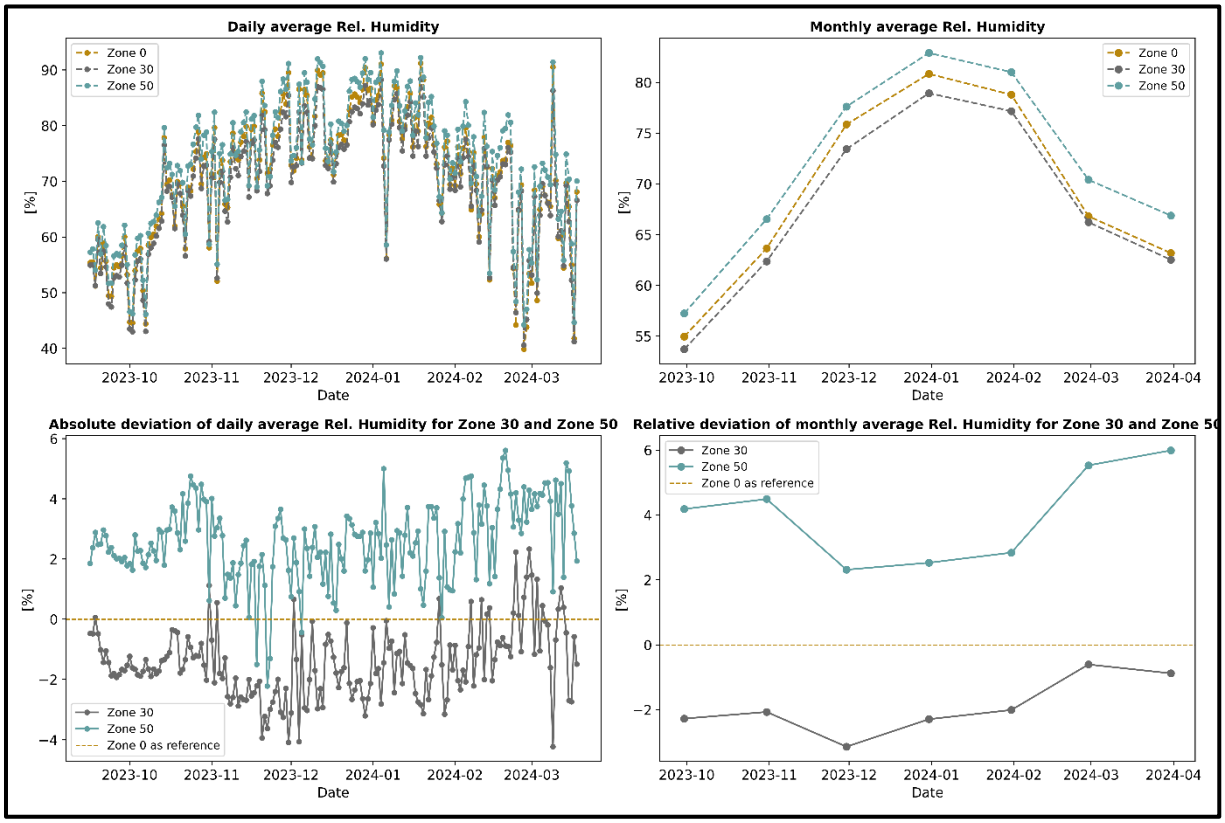

Figure 17: Existing relative humidity measured values for each shading zone and their deviations from the control zone in the greenhouse during the entire growing season.

At RH (fig. 17), zone 50 always has the highest humidity with up to over 90%, followed by the control zone. Zone 30 has the lowest. Zone 50 therefore has a higher RH and zone 30 a lower RH than the control zone. It is assumed that the unexpected temperature and humidity values in zone 30 are due to the orientation of the greenhouse (cf. 3.1.2).

3.1.5 Tomato Plants: Information, treatment and measurements

The plants have been provided and planted by the company ANECOOP S.COOP on 16.09.24. Two tomato plants were planted per square meter. The tomato plants had two main stems. It concerns the novel tomato variety 'Lygalan', which was developed in a project by ANECOOP S.COOP with a flavor-enhancing taste (Rubio, 2022). Similar to the Canary tomato, it is a versatile tomato variety with a sweet and juicy flavor that is mainly grown in the Canary Islands and in the south of the Iberian Peninsula, where the subtropical climate favors its development. They are grown outdoors or in greenhouses and are harvested all year round, with the main season running from spring to fall. Canary tomatoes are usually medium to large in size and have a round or oval shape (ANECOOP S.COOP, 2024). They have a thin, shiny skin and their color can vary from yellow-orange to deep red when ripe (fig. 18).

Figure 16: Harvest of fresh ripe Lygalan tomatoes

Treatment	Date	Product	Pest
1	28.09.2023	Affirm	Lepidoptera
		Movento	Mites
2	10.11.2023	Altacor	Caterpillar
		Costar	Caterpillar
3	16.01.2024	Prestop	Botrytis
4		Xanilo 45	Mildew
	25.01.2024	Costar	Caterpillar
		Copper	Botrytis
5		Azufega 80	Botrytis
	30.01.2024	Costar	Caterpillar
		Altacor	Caterpillar
		Fliper	Thrips
6		Switch One	Botrytis
	13.02.2024	Spintor	Caterpillar
		Cordial	Thrips
7		Spintor	Caterpillar
	20.02.2024	Altacor	Caterpillar
		Costar	Caterpillar
		Zenith	Tuta
8		Movento	Mites
	01.03.2024	Spintor	Caterpillar
		Azufega 80	Botrytis

Table 1: Plant protection products supplied and associated pests found.

Date	Fertilizer	Quantity (Kg)
20.09.2023	Calcium nitrate	12.5
20.09.2023	Mono-potassium phosphate	12.5
20.09.2023	Potassium nitrate	12.5
20.09.2023	Potassium sulfate	12.5
20.09.2023	Magnesium sulfate	12.5
31.10.2023	Calcium nitrate	29
31.10.2023	Potassium sulfate	7.34
30.11.2023	Calcium nitrate	10.78
30.11.2023	Magnesium sulfate	16.17
30.11.2023	Potassium sulfate	10.78
31.12.2023	Mono-potassium phosphate	14.21
31.12.2023	Calcium nitrate	21.32
31.12.2023	Potassium nitrate	14.21
31.12.2023	Magnesium sulfate	14.21
31.12.2023	Potassium sulfate	28.42
31.01.2024	Mono-potassium phosphate	13.45
31.01.2024	Calcium nitrate	12.35
31.01.2024	Potassium nitrate	31.22
31.01.2024	Magnesium sulfate	16.87
31.01.2024	Potassium sulfate	32.22
29.02.2024	Mono-potassium phosphate	9.41
29.02.2024	Calcium nitrate	10.22
29.02.2024	Potassium nitrate	33.45
29.02.2024	Magnesium sulfate	15.22
29.02.2024	Potassium sulfate	29.15

Table 2: Amount of fertilizer used and date of addition.

The treatments and measurements were also carried out by ANECOOP S.COOP. Tables 1 and 2 show the treatments with fertilizers and plant protection products. All zones were treated homogeneously.

Another important physiological measure that was taken was the cutting off of the tomato heads on 09.01.2024. This step is common practice among tomato growers and is done so that the plants no longer grow in height and the tomatoes that already exist on the plant grow larger and ripen faster in order to end the growing period more efficiently and on time. In this case, the growing period for the company was scheduled until mid-March. Once the tomatoes had reached the ripening stage and were ready for harvesting, the fresh and dry weights were measured. The first tomatoes developed in December 2023, so that the first weight (fresh and dry) measurements of the ripe fruit could be carried out in mid-January. Thereafter, the weight per zone was measured at approximately weekly intervals for six representative and centrally located plants until the tomato cycle was completed. The last measurement was taken on 04.03.2024. Table 3 shows the exact days of the measurements. In order to measure the dry weight, the tomatoes were dried in an oven. The Memmert Universal drying oven (Memmert GmbH + Co.KG, n.d.) was programmed at 70 ºC for 5-6 days.

Table 1: Dates of weight measurements.

18.01.2024
30.01.2024
07.02.2024
20.02.2024
26.02.2024
04.03.2024

3.2 Use and development of a model to simulate tomato yield in a 'Raspa y amagado' greenhouse in Southern Spain

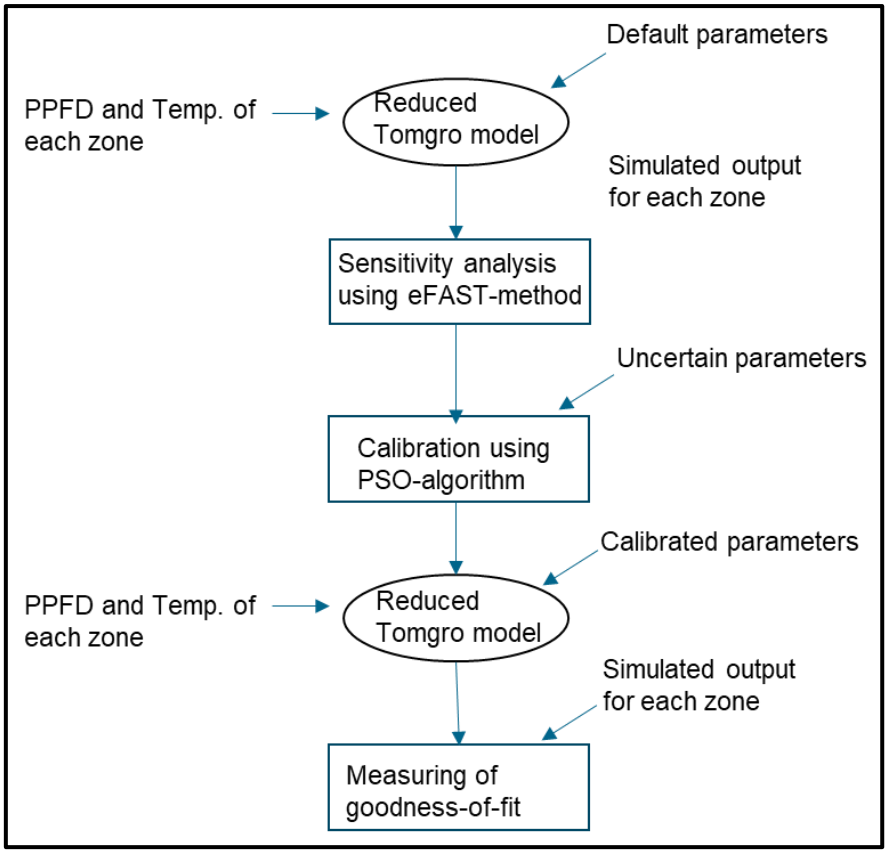

Figure 17: Framework for the calibration process of the reduced TOMGRO model.

This subsection explains the methodology of the entire model calibration in order to adapt the TOMGRO parameters to crop and site dependent microclimatic conditions (fig. 19). First, the model used is described and explained (3.2.1). Then the sensitivity method is described, which uses the default parameters of the model to identify the sensitive parameters (3.2.2). This is followed by an explanation of the actual calibration process, which was carried out exclusively with the sensitive or uncertain parameters (3.2.3). Once the sensitive model parameters have been successfully calibrated, the model can be restarted with the calibrated parameters and the model quality can be assessed using various quality criteria.

3.2.1 The reduced state-variable Tomato Growth Model (reduced TOMGRO)

The 'Reduced State-Variable Tomato Growth Model' (reduced TOMGRO) was developed by Jones et al. (1999) and focuses on the investigation of the relationship between plant growth and environmental factors in the greenhouse such as solar radiation and temperature. It corresponds to a biophysical model-based approach, which is better suited for this work than, for example, a machine learning method in view of the non-existent historical period or data and the better interpretability. Such a so-called 'Functional Structural Plant Model' (FSPM) estimates the entire plant behavior and yield under dynamic conditions (Vos et al., 2010). It uses a source-sink approach for carbohydrate partitioning among plant organs (Jones et al., 1999). TOMGRO was originally developed by Jones et al. (1991) as a comprehensive model that takes into account hundreds of state variables that influence tomato growth. Therefore, it was not suitable in its applicability or too complex to figure out how to optimally control the greenhouse environment to maximize profit over time (Jones et al., 1991). For this reason, the reduced TOMGRO was developed. It aims to reduce the complexity of the original model without neglecting key aspects of tomato growth.

Table 2: Definition of the five state variables and their differential equations of the reduced TOMGRO model (Jones et al., 1999; Vazquez-Cruz et al., 2014).

Variable	Definition	Equation	Units
N	Number of nodes on mainstem	$\frac{dN}{dt} = N_m \cdot f_N(T)$	Nodes
LAI	Leaf area index	$\frac{d(\mathrm{LAI})}{dt} = \rho \cdot \delta \cdot \lambda(T_d) \frac{\exp[\beta \cdot (N - N_b]}{1 + \exp[\beta \cdot (N - N_b]} \cdot \frac{dN}{dt}$	m^2/m^2
W	Above ground dry weight	$\frac{dW}{dt} = \frac{dW_F}{dt} + (V_{max} - p_1) \cdot \rho \cdot \frac{dN}{dt}$	g/m^2
W_F	Total fruit dry weight	$\frac{dW_F}{dt} = GR_{net} \cdot \alpha_F \cdot f_F(T_d) \cdot [1 - e^{-\vartheta(N - N_{FF})}] \times g(T_{daytime})$	g/m^2
W_M	Mature fruit dry weight	$\frac{dW_M}{dt} = D_F(T_d) \cdot (W_F - W_M)$	g/m^2

This was achieved by identifying and eliminating fewer important state variables and processes. Ultimately, the model was reduced from the original 574 state variables (TOMGRO ver. 3.0) to five main state variables (table 4). The state variables are based on a set of time-dependent differential equations (table 4) related to photosynthesis, transpiration, nutrient transport and other physiological processes to quantify changes in stems, leaves and fruits, and generally the growth of the whole plant. The differential equations require numerous parameters that are included in the calculation of the state variables (table 5; Jones et al., 1999).

α_F	Maximum partitioning of new growth to fruit	[fraction] d^{-1}
β	Coefficient in expolinear equation	$node^{-1}$
δ	Maximum leaf area expansion per node, coefficient in expolinear equation	m^2[leaf] $node^{-1}$
κ_F	Development time from first fruit to first ripe fruit	node
$\lambda(T_d)$	Temperature function to reduce rate of leaf area expansion	unitless (0 to 1 function)
ϑ	Transition coefficient between vegetative and full fruit growth	$node^{-1}$
ρ	Plant density	no.[plants] m^{-2}[ground]
$D_F(T_d)$	Function for rate of development of fruit vs. average daily temperature, T_d	d^{-1}
E	Growth efficiency, ratio of biomass to photosynthate available for growth	g[d.w.] $g^{-1}[CH_2O]$
$f_F(T_d)$	Function to modify partitioning to fruit vs. Average daily temperature, T_d	unitless (0 to 1 function)
$f_N(T)$	Function to modify node development rate as a function of hourly temperature	unitless (0 to 1 function)
$f_R(N)$	Fraction partitioning of biomass to roots as a function of development (node)	unitless (0 to 1 function)
$g_F(T_{daytime})$	Function to reduce growth due to high daytime temperature, $T_{daytime}$	unitless (0 to 1 function)
GR_{net}	Net aboveground growth rate	g[d.w.] m^{-2}[ground] d^{-1}
LAI_{max}	Maximum leaf area index	m^2[leaf] m^{-2}[ground]
N_b	Coefficient in expolinear equation, projection of linear segment of LAI vs N to horizontal axis	node
N_{FF}	Nodes per plant when first fruit appears	node
N_m	Maximum rate of node appearance (at optimal temperatures)	node d^{-1}
p_l	Loss of leaf d.w. per node after LAI_{MAX} is reached	g[leaf] $node^{-1}$
P_g	Gross photosynthesis, integrated over a day	g[CH_2O] m^{-2}[ground] d^{-1}
Q_{10}	Coefficient in maintenance respiration equation	unitless
r_m	Maintenance respiration coefficient	g[CH_2O] g^{-1}[d.w.] d^{-1}
R_m	Daily maintenance respiration	g[CH_2O] m^{-2}[ground] d^{-1}
t	time	d
T	Hourly temperature	°C
T_{CRIT}	Mean daytime temperature above which fruit abortion starts	°C
T_d	Average daily temperature	°C
$T_{daytime}$	Average temperature during daytime hours	°C
V_{max}	Maximum increase in vegetative tissue d.w. growth per node	g[d.w.] $node^{-1}$

Table 3: The reduced TOMGRO parameter (Jones et al., 1999).

The reduced TOMGRO provides a comprehensive insight into the complex growth behavior of tomato plants and it is the most representative model in the current greenhouse tomato growth simulation research (Bertin and Gary, 1993; Dayan et al., 1993). Compared to other models, it is convincing in its simplicity, efficiency and applicability for reliable and accurate predictions, which have already been validated at different locations and growth conditions. The parameters for vegetative growth were consistent, but the parameters for fruit development varied according to location and variety, so that these from Table 5 must be adapted or calibrated to local conditions (Jones et al., 1999). The model can also be adapted to simulate other greenhouse crops, such as pepper (Hernández-Hernández et al., 2011). Further information on the state variables, model evaluation and parameter development can be found in Jones et al. (1999).

In order to use the model for this work and to simulate the tomato yield in different shading zones in a 'raspa y amagado' greenhouse, the differential equations were implemented in a Python code with the data from the greenhouse, similar to Yonryu (2020). Figure 20 shows the model framework of the yield simulation used in this work.

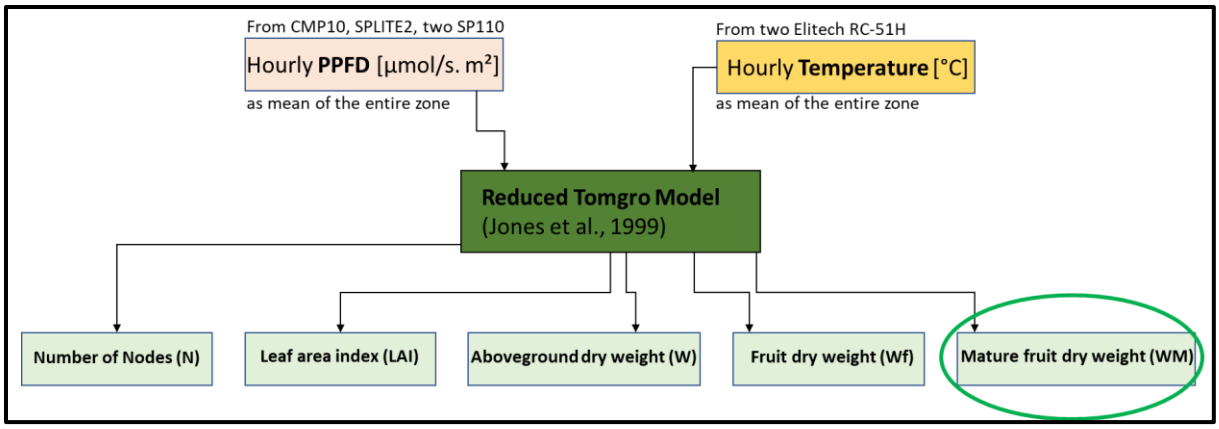

Figure 18: Model framework for tomato yield simulation.

It is important to note that the node and LAI dynamics are temperature dependent and the dry matter production and distribution, which is influenced by photosynthesis and transpiration, is dependent on the PAR-radiation present in the greenhouse. This results in the two inputs required by reduced TOMGRO: temperature and PPFD. The measured hourly mean values of the respective zone are transferred to the model and the five daily outputs or state variables for each zone are simulated using the differential equation system. The CO_2 concentration is assumed to be constant in the model. Based on a single test measurement, it was quantified as 500 µmol mol^{-1} for this study. In order to adapt the model to the conditions in the greenhouse, it must be calibrated using actual measured values from the greenhouse. As the tomato cycle had already started in September and it was only decided in mid-November to carry out precise crop modeling as with TOMGRO in the entire agrivoltaic project of the German Aerospace Center (DLR) - Institute of Solar Research in collaboration with the cooperation company ANECOOP S.COOP, not all the data required for the reduced TOMGRO could be measured from the outset. Therefore, the calibration of the model focuses exclusively on the dry weight of the ripe tomato fruits (WM), while the other four state variables are not considered in this work, as no actual measured values from the greenhouse exist for them.

3.2.2 Sensitivity analysis using the Extended Fourier Amplitude Sensitivity Test (eFAST) method

In order to be able to calibrate the model, it is first of all crucial to find the parameters from Table 5 that have a major influence on the model output. In this way, the parameters that should be optimized can be identified. The main principle of finding the parameters to be calibrate is to obtain the sensitivity of a parameter by changing the input parameter (Lin et al., 2019). It is the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input (Saltelli et al., 2004). The analysis and quantification of the influence and uncertainty of individual model parameters on the model result and thus the identification of the parameters to be calibrated therefore generally depends on the sensitivity analysis. A distinction is made in sensitivity analysis between two categories: local sensitivity analysis and global sensitivity analysis. A local analysis is particularly suitable if the model parameters do not influence each other. If the

parameters are interrelated, a global sensitivity analysis is advantageous (Lamboni et al., 2009; Saltelli, 2008). Since previous studies such as López-Cruz et al. (2012); Vazquez-Cruz et al. (2014) showed that a global analysis is more suitable for crop models and that most parameters in the Reduced TOMGRO Model interact with each other, a global sensitivity analysis was carried out.

There are various methods within global sensitivity analysis that can be looked up in the related literature (Cukier et al., 1978; Morris, 1991; Saltelli et al., 2012; Sobol, 2001). One methodology is the 'Extended Fourier Amplitude Sensitivity Test' (eFAST) method. It analyzes the variations of the model response to periodic changes in the input parameters using the Fourier amplitudes to calculate the sensitivity indices of the individual parameters. It extends the FAST method by additionally taking into account the interactions of the individual parameters and calculating higher-order sensitivity indices, thus enabling greater stability and higher precision compared to the other methods (Mulla, 2003; Saltelli, 1999). Studies specializing in the sensitivity analysis of the reduced state-variable TOMGRO model for Central Mexico and China also carried out a global sensitivity analysis using the EFAST methodology (Lin et al., 2019; Vazquez-Cruz et al., 2014). For this reason, the eFAST method was also used in this study. The EFAST method can therefore be used to calculate not only the first order sensitivity index (S_i), which only considers the direct influence of a parameter without interaction, but also the total effect index (S_{Ti}), which includes both the direct influence of the individual parameter and the influence of the parameter through interactions with other parameters (López-Cruz et al., 2012). Equation 1 and 2 show the calculation of the indices, where Si is the main effect sensitivity index and S_{Ti} is the total effect index of the i-th input variable, V_i is the variance of the i-th input variable of the model response and $V_{\sim i}$ is the variance of the model response attributable to all other variables except the i-th. V is the total sum of the variance of the model response. These metrics are dimensionless and represent the proportions of the variance of the model result (Saltelli et al., 2012). The open source Python implementation 'Sensitivity Analysis Library' (SALib) by Herman and Usher (2017) and Iwanaga et al. (2022) was used for this purpose.

$$S_i = \frac{V_i}{V} \tag{1}$$

$$S_{Ti} = 1 - \frac{V_{\sim i}}{V} \tag{2}$$

To apply the eFAST method, the value ranges of the parameters in which they can take on values must be defined. These were selected or considered from numerous literature references, as some of these used different default values and value ranges and the reasons for this were often not clear (Bacci et al., 2012; Gong et al., 2021; Hernández-Hernández et al., 2011; Lin et al., 2019; Ramírez et al., 2004; Vazquez-Cruz et al., 2014). A range of values was summarized for each parameter from all available literature data to ensure that all possible values were covered. If available,

the value range of the study by (Ramírez et al. (2004) from Southern Spain was prioritized. However, if other studies gave values outside this range, the range of values was extended accordingly. If only the original value was available and thus no value range, a probability density function (PDF) was selected. As in Vazquez-Cruz et al. (2014), the PDF was set with a 10% deviation around its original value. Table 6 shows the value ranges used for the respective parameters with the respective sources. Within these value ranges, the parameter values are changed using the eFAST method from the SALib library with 3000 samples and the reduced TOMGRO model is run in each case. With 5000 resamples and a confidence interval of 0.95, the sensitivity indices S_i and S_{Ti} of the parameters were then calculated with regard to the model output. Then, according to Varella et al. (2010), only those parameters that yielded more than 10% of the total effects were classified as sensitive.

Table 4: Value ranges of the reduced TOMGRO parameters. The value ranges were summarized from various studies described below.

Parameter	Description	Unit	Range of variation	Original value ([a] and [b])	Optimized values from other authors
N_m	Maximum rate of node appearance (at optimal temperatures)	node d^{-1}	**[0.1 - 0.9][c]**	0.5[a]	0.495[c] ,0.495[d], 0.5[e], 0.4039[f]
delta	Maximum leaf area expansion per node, coefficient in expolinear equation	m^2[leaf] $node^{-1}$	**[0.01 - 0.1][c]**	0.038[b]	0.041[c] , 0.041[d], 0.1[e], 0.0875[f]
beta	Coefficient in expolinear equation	$node^{-1}$	**[0.06 - 0.5][c]**	0.169[b]	0.22[c], 0.22[d], 0.169[e]
N_b	Coefficient in expolinear equation, projection of linear segment of LAI vs N to horizontal axis	node	**[8.0 - 25.0][c]**	16.0[b]	13.0[c], 18.5[d], 16.0[e]
LAI_{max}	Maximum leaf area index	m^2[leaf] m^{-2}[ground]	**[±10% 4.0][b]**	4.0[b]	-
p1	Loss of leaf d.w. per node after LAIMAX is reached	g[leaf] $node^{-1}$	**[±10% 2.0][b]**	2.0[b]	-
V_{max}	Maximum increase in vegetative tissue d.w. growth per node	g[d.w.] $node^{-1}$	**[2.0 - 12.0][c]**	8.0[b]	6.0[c], 10.0[e]
N_{FF}	Nodes per plant when first fruit appears	node	**[4.0[f] - 22.0[b]]**	22.0[b]	14.88[d], 4.0116[f]
k_F	Development time from first fruit to first ripe fruit	node	**[0.5[d] - 5.0[b]]**	5.0[b]	0.58[d]
tau	Carbon dioxide use efficiency (Part of Pg calculation in non-reduced model (Jones 1991))	µmol m^2 s^{-1}	**[0.01 - 0.5][c]**	0.0693[a]	0.12[c], 0.0664[d], 0.084[e]
D	Coefficient to convert Pg from CO2 to CH2O (Part of Pg calculation in non-reduced model (Jones 1991))	$gm^{-2}h^{-1}$	**[0.1[d] - 2.6[b]]**	2.593[a]	0.108[d], 0.317[f], 2.593[g]
K	Light extinction coefficient (Part of Pg calculation in non-reduced model (Jones 1991))	unitless	**[0.3 - 0.9][c]**	0.58[a]	0.61[c], 0.58[e], 0.3888[f]
m	Leaf light transmission coefficient (Part of Pg calculation in non-reduced model (Jones 1991))	unitless	**[±10% 0.1][a]**	0.1[a]	0.1[d]
Qe	Leaf quantum efficiency (Part of Pg calculation in non-reduced model (Jones 1991))	mol(CO2 fixed) /mol(photons)	**[0.01 - 0.5][c]**	0.0645[a]	0.09[c], 0.09[d]
Q10	Coefficient in maintenance respiration equation	unitless	**[1.4[a] - 1.6[d]]**	1.4[a]	1.6[d], 1.4[g]
r_m	Maintenance respiration coefficient	g[CH2O] g^{-1}[d.w.] d^{-1}	**[0.005[d] - 0.016[b]]**	0.016[b]	0.006[d]
E	Growth efficiency, ratio of biomass to photosynthate available for growth	g[d.w.] g^{-1}[CH_2O]	**[0.5 - 0.9][c]**	0.7[a]	0.7[c], 0.7[d], 2.0663[f], 0.75[g]
T_CRIT	Mean daytime temperature above which fruit abortion starts	°C	**[17.0 - 29.0][c], but extended to [17.0 - 30.0[e]]**	24.4[b]	24.0[c], 30.0[e]
alpha_F	Maximum partitioning of new growth to fruit	[fraction] d^{-1}	**[0.1 - 0.8][c], but extended to [0.1 - 0.96[e]]**	0.8[b]	0.95[c], 0.95[d], 0.96[e]
v (ϑ)	Transition coefficient between vegetative and full fruit growth	$node^{-1}$	**[0.05 - 0.9][c]**	0.135[b], 0.2[b]	0.24[c] , 0.24[d], 0.16[e]
fF_	Function to modify partitioning to fruit vs. Average daily temperature, Td	unitless (0 to 1 function)	**[0-1][a]**	not specified	0.825[f]

[a]Jones et al. (1991); [b]Jones et al. (1999); [c]Ramirez et al. (2004); [d]Hernandez-Hernandez et al. (2011); [e]Bacci et al. (2011); [f]Vazquez-Cruz et al. (2013); [g]Lin et al. (2019); [h]Gong et al. (2021)

3.2.3 Calibration of the model

The sensitive parameters or the parameters with the greatest influence have been identified and are now being calibrated. This chapter explains the methodology for this. Firstly, the algorithm used is presented and its functionality explained (3.2.3.1). Before the application of the algorithm to the reduced TOMGRO model (3.2.3.3) is presented, an additional step is explained in order to fill the data gaps in the actual measured data from the greenhouse (3.2.3.2). Finally, the individual quality criteria used to evaluate the model performance are presented (3.2.3.4).

3.2.3.1 The Particle Swarm Optimization (PSO)

The calibration was carried out using the 'Particle swarm optimization algorithm' (PSO). In addition to the PSO algorithm, there are other evolutionary algorithms, such as the genetic algorithm (García-Martínez et al., 2019) and the differential evolutionary algorithm (Suganthan, 2012). The genetic algorithm was used, for example, by Vazquez-Cruz et al. (2014) for the calibration of TOMGRO. However, since it has not yet been investigated how other optimization algorithms can be used to calibrate the reduced Tomgro model, Gong et al. (2021) evaluated different evolutionary algorithms regarding the performance of the reduced TOMGRO for a real-scale greenhouse tomato yield simulation based on three datasets obtained from a real tomato grower. The model calibrated by the PSO algorithm achieved the best performance for modeling the WM, with the smallest Root Mean Squared Errors (RMSE), relative RMSEs (r-RMSE) and Mean Absolute Errors (MAE) for all three datasets. For this reason, the PSO algorithm was also used for this work.

The PSO algorithm is a stochastic population-based optimization method proposed by Kennedy and Eberhart (1995). The optimization here refers to the investigation of the minimization of an objective function by searching for the best values for the parameters within the permissible range of values (Elbes et al., 2019). The ideas for this algorithm come from artificial intelligence, social psychology and swarm theory (Kennedy and Eberhart, 1995). Swarm intelligence is a type of artificial intelligence and

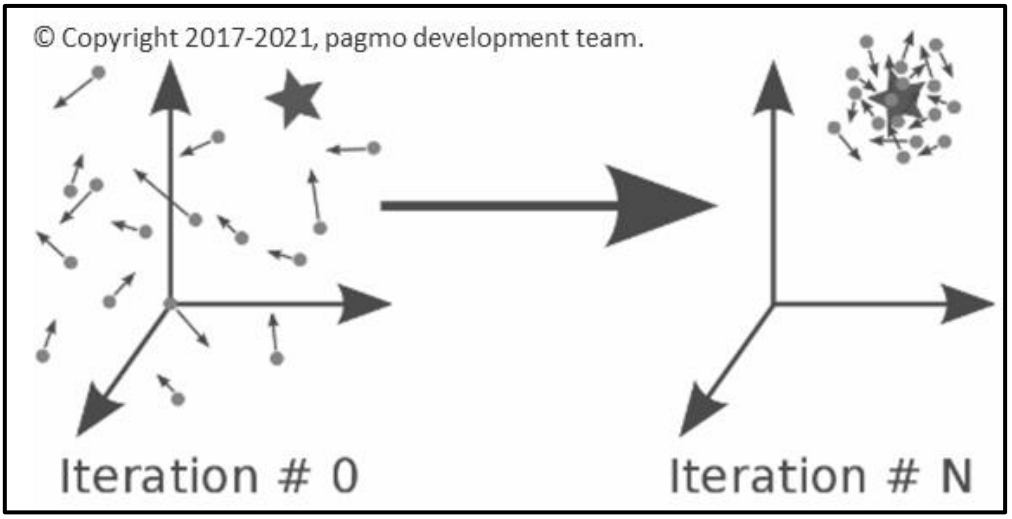

Figure 19: Illustration of the PSO algorithm. Each point represents one possible set of parameter values within the parameter space or within their parameter ranges. The star represents the optimal solution (pagmo development team, 2021).

is based on intelligent the collective behavior of decentralized or self-organized systems such as flocks of birds or schools of fish (Elbes et al., 2019).

In specific, initially N solution candidates or also called particles [$X_1{}^0$, …, $X_N{}^0$] and their associated velocities [$V_1{}^0$, …, $V_N{}^0$] are generated. Each particle represents one possible set of parameter values and a potential solution position within the parameter space (cf. fig. 21). All candidates have a fitness value, which can be calculated using a fitness function associated with a particular optimization problem. The values of the particles or their positions, which are based on the successful or best solutions of the individual particle (local optimum), but also of the whole swarm (global optimum), found during the optimization process after each iteration until the optimal solution is found using equations 3 and 4:

$$v_i(t+1) = wv_i(t) + c_1r_1\big(p_i - x_i(t)\big) + c_2r_2(g - x_i(t)) \quad (3)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (4)$$

where $v_i(t)$ is the velocity of particle i at time t, $x_i(t)$ is the position of particle i at time t, ω is the inertia coefficient, c_1 (cognitive component) and c_2 (social component) are the acceleration coefficients, r_1 and r_2 are random values between 0 and 1, p_i represents the best position found so far by particle i (personal best), g represents the best position (global best) found by the entire swarm (Kennedy and Eberhart, 1995).

3.2.3.2 Gap filling for the actual data with regression substitution

As only approximately weekly measurements of the dry weight of the ripe fruit are available in the greenhouse (cf. table 3), various methods were tried out to apply the PSO algorithm to reduced TOMGRO using the actual values. One possibility, for example, is to only compare the simulation with the actual values on the days on which a measurement is available. In this way, comparisons and evaluations are really only made on the exact measurement days. However, comparisons between the individual measurement days are then missing and lead to poorer model performance. In order to close these gaps and create completeness and robustness with a higher data density, a regression substitution was applied to the actual data sets in advance so that a value for the actual dry weight of the ripe fruit is also available for every single day. The existing measured values served as the basis for the creation of the regression model. Since this is an accumulative, steadily increasing data set, a linear regression Analysis, as described with equation 5, was performed to effectively and easily fill or estimate the gaps in the data series (Gelman and Hill, 2021; Little and

$$y = ß_0 + ß_1x \quad (5)$$

Rubin, 2020; Schafer, 1997). This contains the calculation of the slope $ß_1$ and the intercept $ß_0$ of the best fit line, which describes the relationship between the data points. For each missing value, the corresponding location in the data series x was substituted into the equation to estimate the value y (Montgomery et al., 2021). The

values of the regression lines were used then for further calibration as the actual values from the greenhouse.

3.2.3.3 Application of PSO to the reduced TOMGRO model

To implement the PSO algorithm in the reduced TOMGRO model, the Python library 'Pyswarms' by Miranda et al. (2017) was used. Table 7 shows the initialization and configuration of hyperparameters and properties with which the algorithm worked best.

Table 5: Selected hyperparameters and settings. Other values for hyperparameter as well as several iterations and particles were tried, but these had no positive influence on the performance of reduced TOMGRO.

Cognitive Coefficient **C1**	Social Coefficient **C2**	Inertial weight **w**	Number of **particles**	Number of **iterations**
0.5	0.5	0.8	100	100

Figure 22 shows the calibration process method of the reduced TOMGRO model using the PSO algorithm, which takes all three zones into account. To achieve the overall best results, each parameter set, consisting of the sensitive reduced TOMGRO parameters, was passed to reduced TOMGRO from the respective iteration within the permissible range of each parameter from Table 6 and the model was run individually for each zone. The RMSE between the model predictions of WM and observed or actual tomato dry weight from the greenhouse represents the fitness function. The

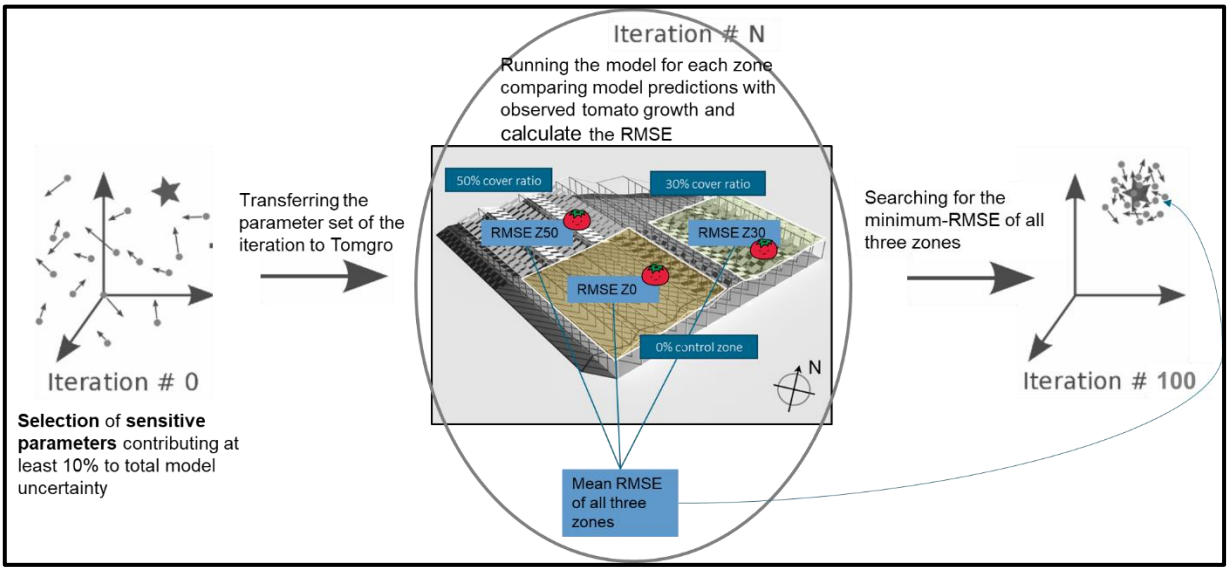

Figure 20: Procedure of the PSO algorithm to identify the optimal parameter values for minimizing the RMSE between simulated and actual values for all three zones. The star in the parameter space that is searched for represents the minimum mean RMSE from all three zones.

RMSE was calculated for each zone individually using equation 6 and then the mean RMSE was calculated from all three zones. Where N is the number of observations, y_i^{real} is the actual values, and y_i^{model} is the predicted values (Chai and Draxler, 2014; Gong et al., 2021). This means that all squared error values are considered and the

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^{real} - y_i^{model})^2} \quad (6)$$

average squared difference between the observed and predicted values is shown (Tibshirani et al., 2021). This involves a joint and continuous search in the parameter space for the minimum mean RMSE from all three zones in order to find the optimal parameter values that bring the simulated WM as close as possible to the dry weight from the greenhouse. In order to be able to apply the algorithm to this case study in the best possible way and to achieve the best and most robust results, various application methods were tried out in advance. In addition to including all three zones, firstly, the calibration was performed for each zone individually, without considering the other zones. Secondly, only the control zone was calibrated and then the other two zones were simulated using the calibrated parameter values of the control zone.

3.2.3.4 Measuring of goodness-of-fit

To evaluate the performance of the calibrated model, in addition to the RMSE (equation 6), which also served as a fitness function, other indicators of variance were used to balance the advantages and disadvantages of individual measures of simulation accuracy. Similar to Gong et al. (2021), Lin et al. (2019) and Vazquez-Cruz et al. (2014), a statistical study with a set of metrics calculation between calibrated simulated WM and actual WM from the greenhouse (including mean absolute error (MAE), relative-RMSE (r-RMSE), percent standard error of the prediction (%SEP), average relative variance (ARV), coefficient of determination (R^2)) is therefore suitable for making more robust statements about how well reduced TOMGRO can explain the variations in the data. Equations 7 - 11 each describe the quality criteria used (Chai and Draxler, 2014; Horton and Doxas, 1998; Næs et al., 2017; Ozer, 1985; Willmott and Matsuura, 2005). Where, N is the number of data points, y_i^{real} the actual values, y_i^{model} the predicted values und $y^{real\ average}$ the average of the actual values.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i^{real} - y_i^{model}| \tag{7}$$

$$r - RMSE = \frac{RMSE}{y^{real\ average}} \tag{8}$$

$$\%SEP = \frac{RMSE}{y^{real\ average}} \times 100 \tag{9}$$

$$ARV = \frac{1}{N}\sum_{i=1}^{N}(\frac{y_i^{real} - y_i^{model}}{y_i^{real}})^2 \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i^{real} - y_i^{model})^2}{\sum_{i=1}^{N}(y_i^{real} - y^{real\ average})^2} \tag{11}$$

The MAE calculates the average of the absolute errors between the actual and predicted values. It is less susceptible to outliers, which is why it is suitable for evaluating the average accuracy of the model. The r-RMSE calculates the RMSE

relative to the average of the actual values. The %SEP calculates the standard error of the prediction relative to the average of the actual values and also expresses it as a percentage compared to r-RMSE. It provides a clear interpretation of the model error relative to the average prediction, which increases comparability. The ARV compares the variance of the errors with the variance of the actual values. This makes it possible to assess how well the model explains the variance of the data. R^2 shows how well the model can explain the actual values. It calculates the share of variance in the actual values that is explained by the model predictions. Further information, advantages and disadvantages as well as application examples can be found in the relevant literature (Brogden, 1946; Chai and Draxler, 2014; Horton and Doxas, 1998; Næs et al., 2017; Ozer, 1985; Willmott and Matsuura, 2005). For a perfect match, R^2 should be close to 1.0 and the values of %SEP, MAE, RMSE ARV and r-RMSE close to 0.

4 Results

The results are now presented in this chapter. First, the crop yield measurements are presented (4.1). Then the results of the model development are presented. Starting with the results of the sensitivity analysis (4.2), followed by the calibration and simulation results (4.3).

4.1 Actual measurements

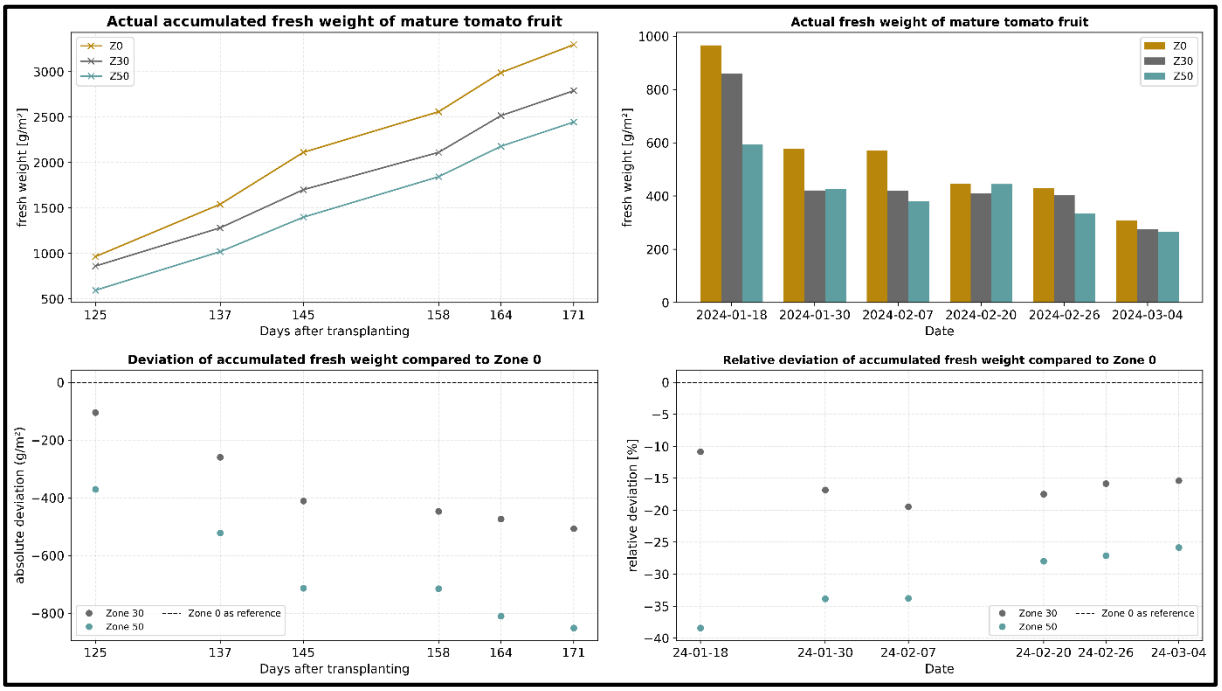

Figure 21:Analysis of the measured mature tomato fresh weight in different shading zones of the greenhouse. Top left shows the accumulation of fresh weight over time. Top right shows the fresh weight of the individual measurements. Bottom left shows the absolute deviation of the accumulated fresh weight of the two shading zones compared to the control zone. The bottom right shows the relative deviation.

Figure 23 shows the fresh weight of the individual zones. At the top left of the accumulated weight, the difference between the individual zones can be seen. Starting from the first measurement (18.01.2024), the control zone has the highest yield with over 950 g/m². The yields of the zones vary on the individual measurement days. The decisive factor, however, is that at the end of the entire growing season, the control

zone again has the most fresh weight with 3298 g/m², followed by zone 30 with just 2791 g/m² and zone 50 with just 2446 g/m². For zone 30 an overall reduction of 507 g/m² (bottom left) or 15%/m² (bottom right) is observed at the end of the crop cycle compared to the control zone. Zone 50 shows a reduction in fresh yield of 851 g/m² or 26%/m² compared to the control zone. In addition, a yield delay of around 10 days for zone 30 and 14 days for zone 50 zone can be seen compared to the control.

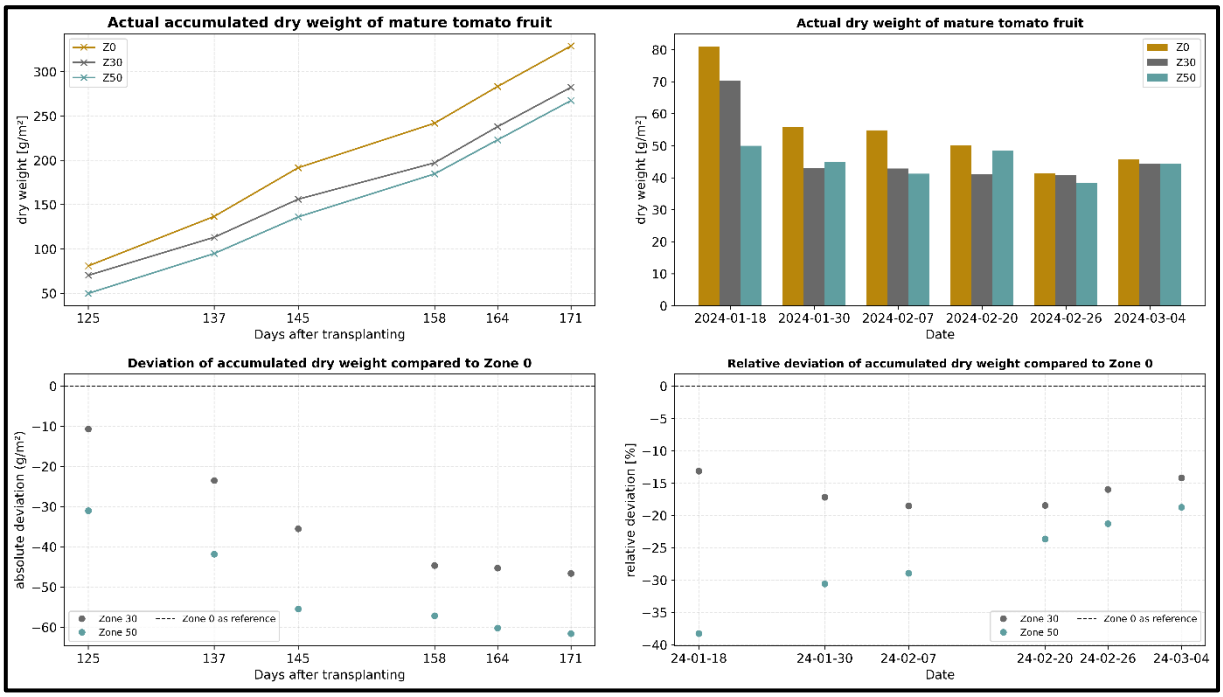

Figure 22: Analysis of the measured mature tomato dry weight in different shading zones of the greenhouse. Top left shows the accumulation of dry weight over time. Top right shows the dry weight of the individual measurements. Bottom left shows the absolute deviation of the accumulated dry weight of the two shading zones compared to the control zone. The bottom right shows the relative deviation.

Figure 24 presents the dry weight of the individual zones. At the top left of the accumulated weight, you can again clearly see the difference between the individual zones. Towards the end of the crop cycle, the deviation between zone 50 and zone 30 is not as dominant as for the fresh weight. On the first day of measurement, the control zone had the highest yield with just over 80 g/m² dry weight. Here too, the placement of the zones varies on individual measurement days, although the difference in dry weight between zone 50 and the other two zones is sometimes higher or the same over time. Here, too, the decisive factor is that at the end of the entire growing season, the control zone again has the highest dry weight with 330 g/m², followed by zone 30 with 282 g/m² and zone 50 with 267 g/m². Zone 30 therefore has an overall reduction in dry yield at the end of the growing season of approx. 46 g/m² (bottom left) or 14%/m² (bottom right) compared to the control zone. Zone 50 shows a reduction in dry yield of approx. 61 g/m² or 19%/m² compared to the control zone.

4.2 Sensitivity analysis

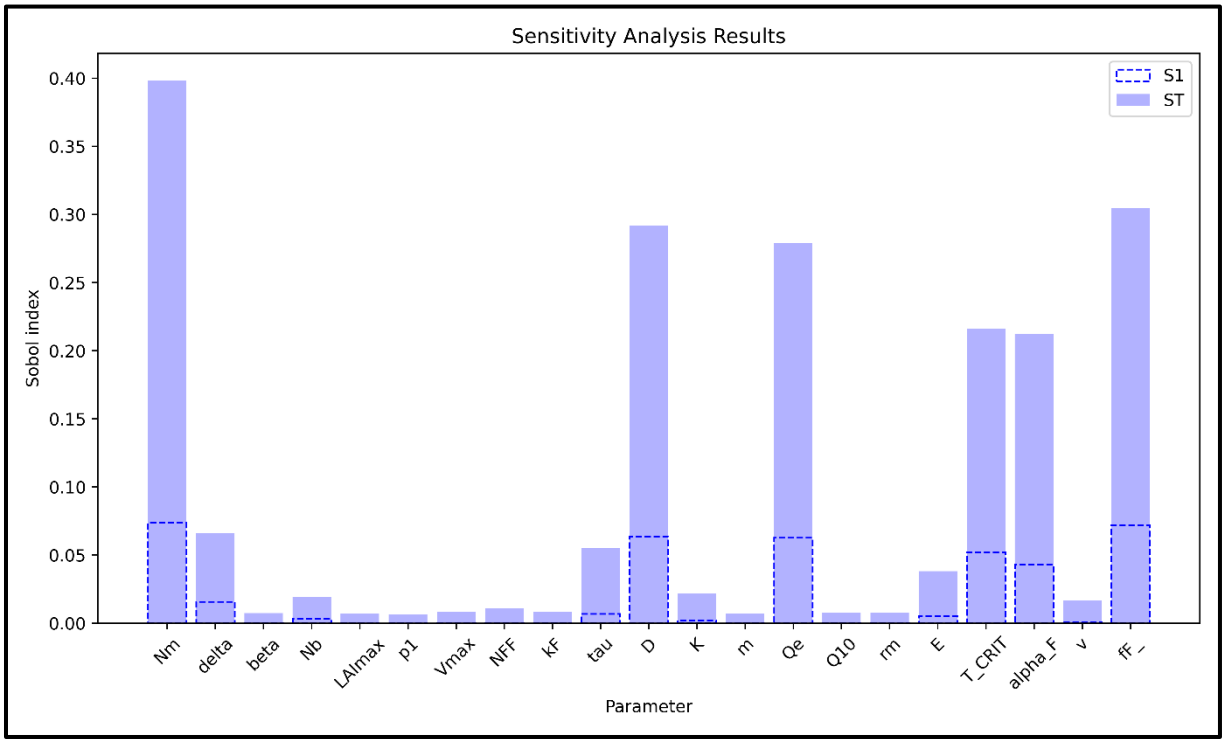

Figure 23: Sensitivity of all parameters with the Sobol indices S_i (direct) and S_{Ti} (interactive). Calculated using eFAST sensitivity method.

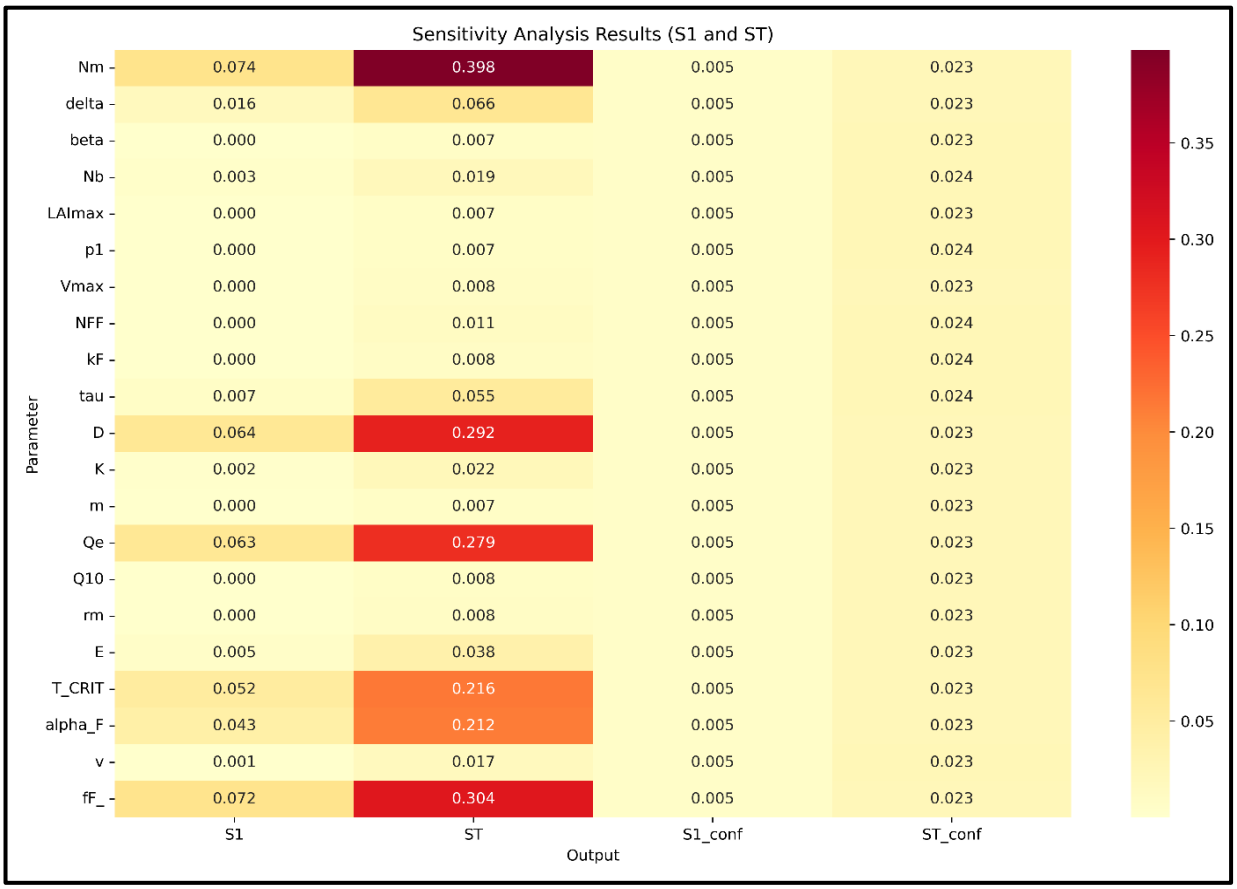

Figure 24: Heatmap of the parameters with the sensitivity indices S_i and S_{Ti}.

Figure 25 and 26 show the results of the sensitivity analysis. The parameters with more than 10% influence on the total model uncertainty and thus classified as sensitive are Nm (0.398), D (0.292), Qe (0.279), T_CRIT (0.216), alpha_F (0.212), fF_ (0.304). The exact description of the individual parameters can be looked up again in Table 6. Of a total of 21 parameters, 6 are sensitive. These are now included or considered in the calibration.

4.3 Calibration of reduced TOMGRO and simulation approach

4.3.1 Gap filling for the actual data with regression substitution

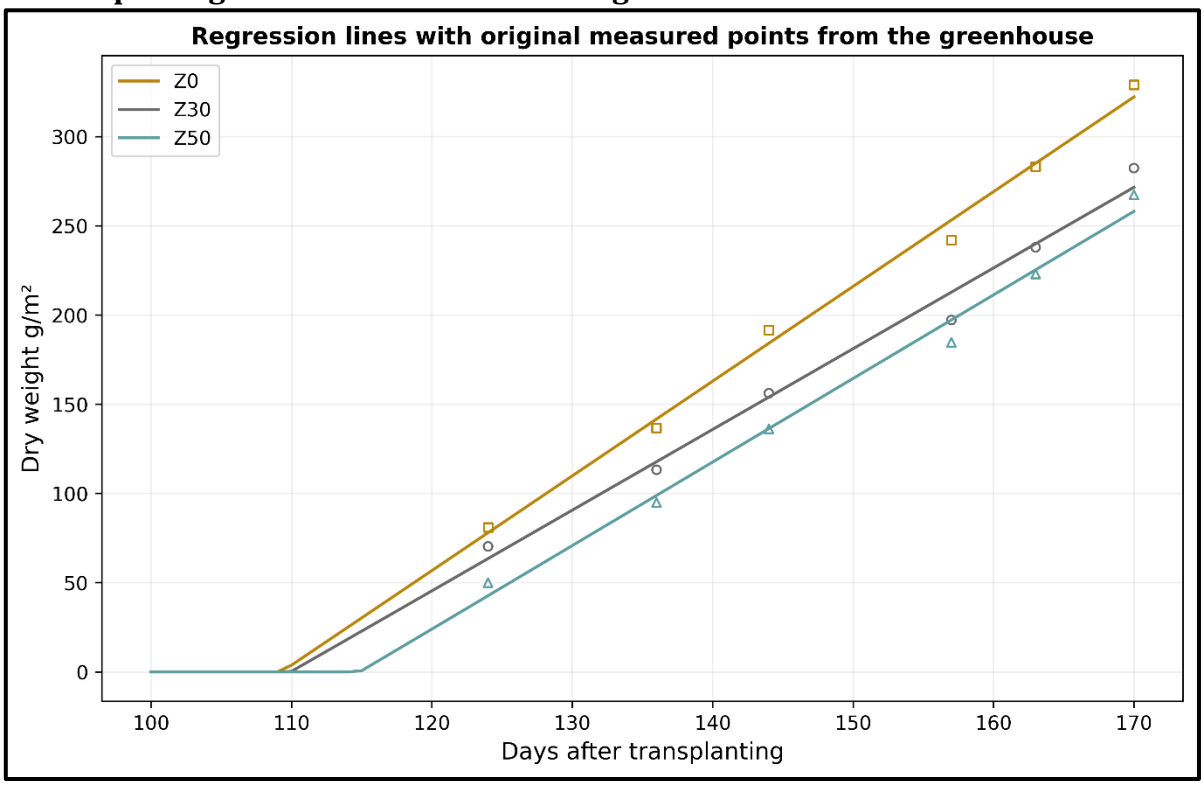

Figure 25: Regression lines that are used for further calculation as the actual values from the greenhouse. They were calculated to fill the gaps between the existing measurement days (dots). All negative values of the regression lines resulting from drawing the line were set to 0, as there is no negative growth of the ripe fruit.

Figure 27 shows the regression lines with their original measured points from the greenhouse. For the calibration the values of the regression line were used.

4.3.2 Comparison of algorithm application methods

This chapter shows the calibration results of the various methods for applying the PSO algorithm. Figure 28 shows the comparison of the simulation curves of the three different applications of the PSO algorithm for the calibration of the reduced TOMGRO model compared to the actual measured values. If each zone is calibrated individually (a.), RMSE values of 8.30 g/m² (Z0), 6.94 g/m² (Z30) and 8.54 g/m² (Z50) are achieved (cf. table 8). However, this method leads to an individual parameter adjustment for each zone, so that there are no uniform parameter values for the entire greenhouse (cf. table 9). When the calibrated parameter values of the control zone (Z0 a. and b.) are transferred to the two shading zones (b.), an RMSE of 24.57 g/m² is achieved for zone 30 and an RMSE of 60.55 g/m² for zone 50. In the last method, in which the minimum mean RMSE from all three zones was searched for (c.), an RMSE of 9.84

g/m² was achieved for the control zone, a value of 9.15 g/m² for zone 30 and a value of 9.67 g/m² for zone 50. Compared to the other two methods, the last method was therefore able to achieve uniform parameter values for the entire greenhouse and at the same time good RMSE values for all three zones and is therefore best suited for this research.

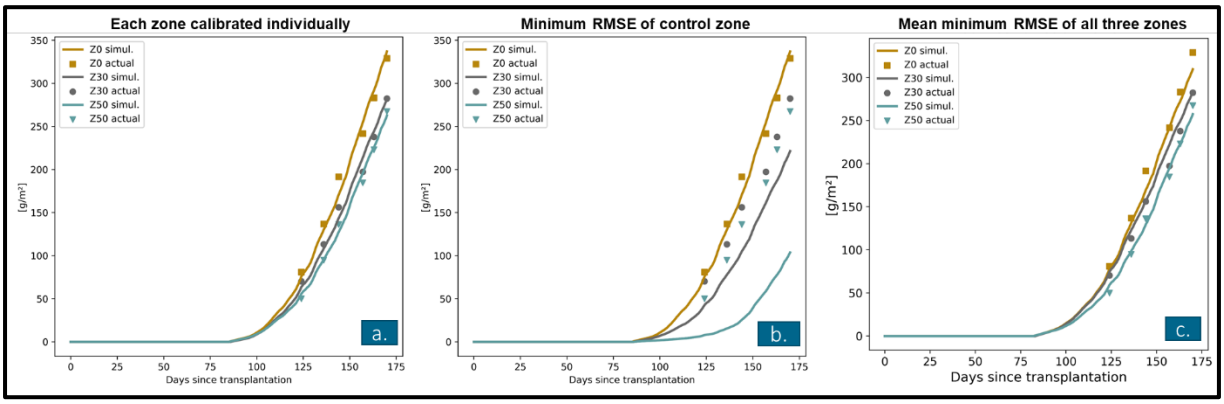

Figure 26: Comparison of the simulation curves of the three different applications of the PSO algorithm for the calibration of the reduced TOMGRO model compared to the actual measured values (dots). a. shows the curves in which the zones were calibrated individually and the minimum RMSE was searched for individually for each zone. In b. the calibrated parameters of the control zone were transferred to Z30 and Z50 and reduced TOMGRO was started with these parameter values. c. shows the curves with the uniform parameter values for all zones after the minimum mean RMSE was searched for from all three zones.

Table 6: RMSE values of the application methods.

Z0	8.30	8.30	9.84
Z30	6.94	24.57	9.15
Z50	8.54	60.55	9.67

Table 7: Calibrated parameter values after application of the various methods.

Parameter	Z0 a. and b.	Z30 a.	Z50 a.	c.
Nm [node d^{-1}]	0.374413229	0.374064896	0.375292934	0.385998652
D [$gm^{-2}h$-^{1}]	2.404519379	2.100298985	2.332767334	1.870763834
Qe [mol(CO2 fixed) /mol(photons)]	0.403494952	0.386887209	0.436927646	0.474278636
T_CRIT [°C]	22.37778773	21.56565451	19.76221041	19.62227286
alpha_F [[fraction] d^{-1}]	0.884608988	0.92209049	0.858932202	0.892947268
fF_ [unitless 0 to 1]	0.664225926	0.861967465	0.895849133	0.943267855

4.3.3 Simulation results and measurement of the goodness-of-fit of the best reduced TOMGRO simulation approach

In the following, the cost history of the fitness function is first shown before the simulation results and the model quality of the reduced TOMGRO before and after calibration with the most suitable method from 4.3.2 (c.) are presented in detail.

4.3.3.1 Cost history of PSO algorithm application

Figure 29 shows the evolution or the cost history of the mean RMSE of all three zones fitness function values with respect to the iteration number for the PSO algorithm applied on the basis of all three data sets or zones. It can be seen that as the number

of iterations increases, the minimum fitness value of the candidate solutions (represented by the black line in the figure) or the minimum mean RMSE of all three zones (cost) decreases and converges to the minimum value (9.45). This indicates that solutions for model parameter values that minimize the fitness function and the best cost can be found successfully after a certain number of iterations.

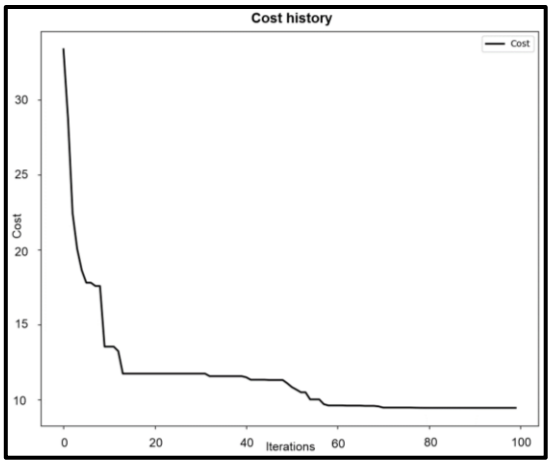

Figure 27: Cost history for the fitness function. The black line represents the minimum mean RMSE of all three zones with respect to the iteration number.

4.3.3.2 Simulation before calibration

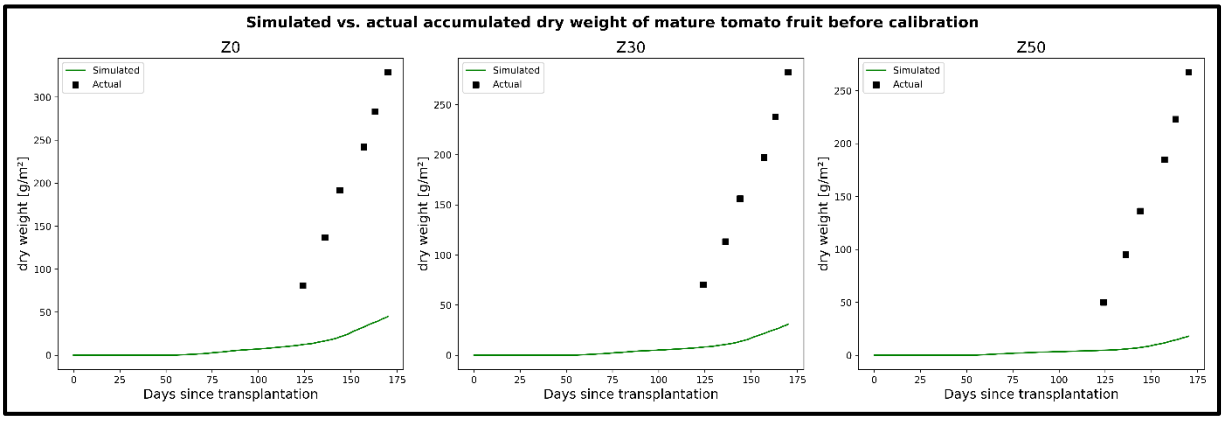

Figure 30: Comparison for all three zones between the actual measured WM from the greenhouse and the simulated WM before calibration of the model or with the default values (Table 10) for the sensitive parameters.

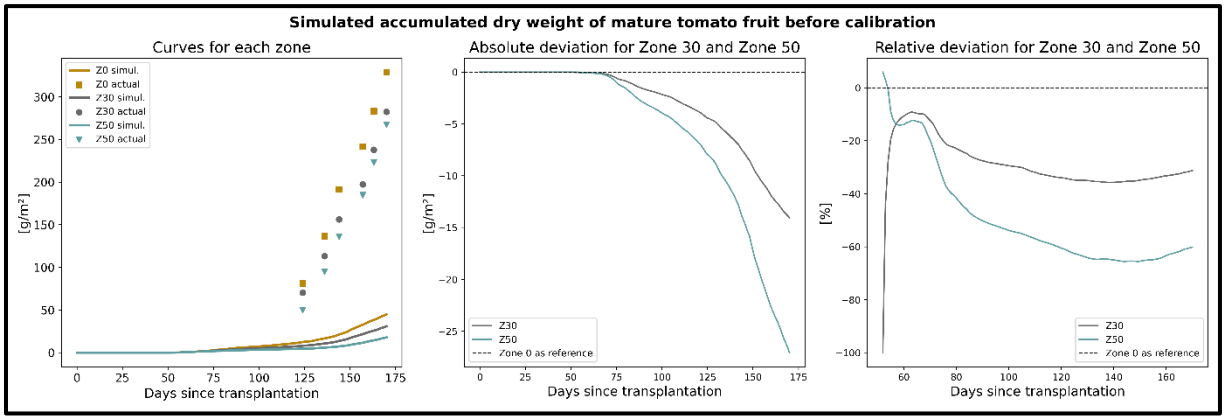

Figure 31: Comparison of the simulated WM between the individual zones before calibration. Left shows the comparison between simulated and actual again, but this time all three zones are compared in one figure. Center shows the absolute deviation of the simulated WM of the shading zones from the control zone. Right shows the relative deviation.

When looking at the simulation results with the uncalibrated parameter values of the sensitive parameters (table 11), a clear deviation between the simulated WM values and the actual measured values can be seen in all three zones (fig. 30). A comparison of the zones with each other (fig. 31) shows that the highest yield is simulated for the

control zone and that this decreases with increasing shading. At the end of the growing season, a dry yield of 45 g/m² is modeled for the control zone, with zone 30 having just under 15 g/m² and zone 50 approx. 27 g/m². The values of the quality criteria (table 10) quantify the high deviation between the simulated and actual WM values. The simulation without calibration has an RMSE above 80 in all zones and the remaining quality measures also indicate a high divergence between the uncalibrated simulation and the actual values. Overall, the uncalibrated results show that calibration is necessary to adapt the model to climatic and site-specific conditions when working with an FSPM.

Table 8: Measuring of the simulation's goodness-of-fit without calibration

	RMSE	MAE	r-RMSE	%SEP	ARV	R^2
Z0	97.63	51.67	1.68	1052.87	0.66	-0.03
Z30	84.40	44.35	1.74	1353.60	0.68	-0.10
Z50	80.44	40.44	1.90	2215.84	0.68	-0.16

Parameter	Value
Nm	0.5
D	2.593
Qe	0.0645
T_CRIT	24.4
alpha_F	0.95
fF_	0.5

Table 9: Default parameter values for reduced TOMGRO (Jones et al., 1999).

Figure 32, which shows the values for the individual measurement days and thus the non-accumulated values, also illustrates the discrepancy between the simulation and reality.

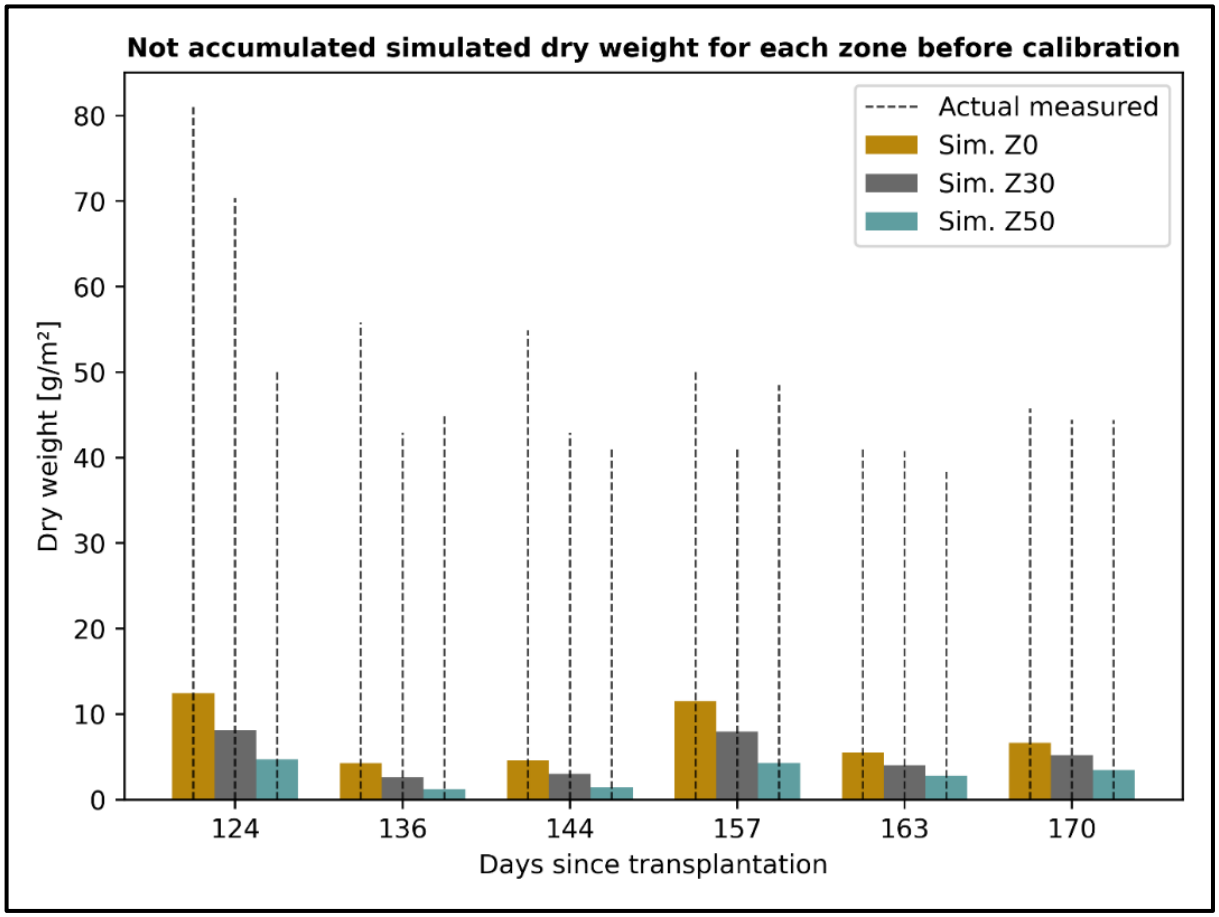

Figure 28: Comparison of the zones between non-accumulated simulated and actual WM before calibration on the individual measurement days from Table 3.

4.3.3.3 Simulation after calibration

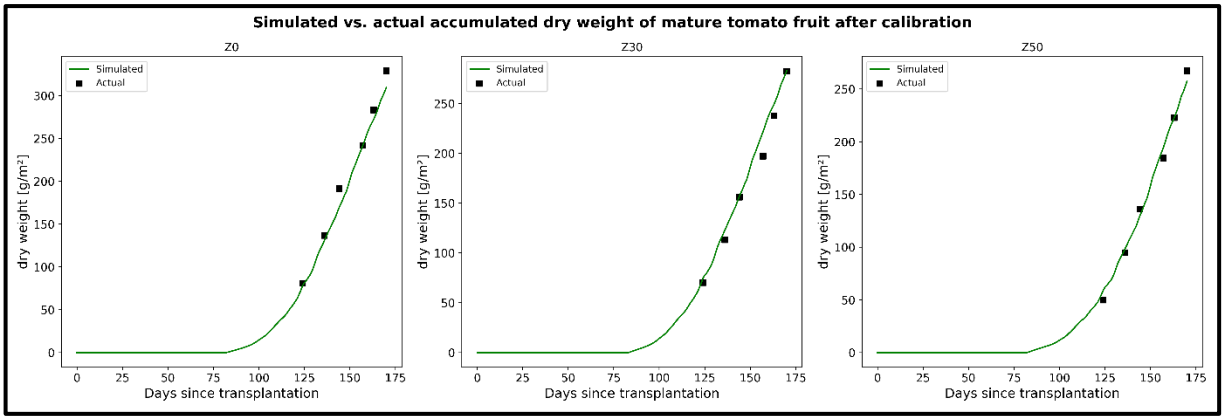

Figure 33: Comparison for all three zones between the actual measured WM from the greenhouse and the simulated WM after PSO calibration of the model.

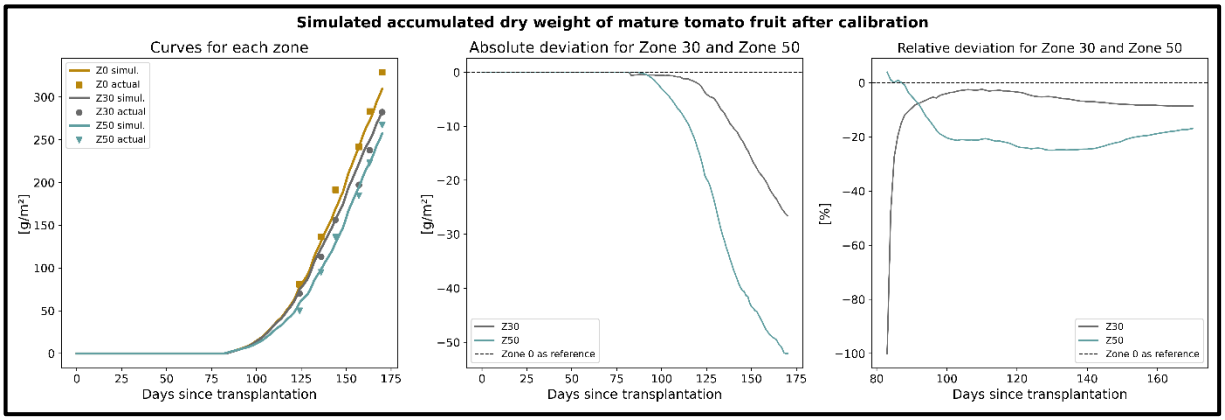

Figure 34: Comparison of the simulated WM between the individual zones after calibration. Left shows the comparison between simulated and actual again, but this time all three zones are compared in one figure. Center shows the absolute deviation of the simulated WM of the shading zones from the control zone. Right shows the relative deviation.

After calibration with the PSO algorithm, a change in the parameter values compared to their default values (table 11) is now noticeable. Nm, D, T_crit and alpha_F were reduced and Qe and fF_ were increased. This value adjustment led to a significant improvement in the model simulation. The simulation curves approach the actual values for all three zones and reproduce the WM almost perfectly (fig. 33). For the control zone and for zone 50, WM is slightly underestimated at the end of the growing season and an almost identical value is achieved for zone 30. Figure 35 illustrates this once more. It summarizes the complete simulated WM yield or the total yield of tomato production per zone and compares it with the total yield actually measured in each case. A comparison of the simulated yield between the zones shows that the control zone with an accumulation of approx. 309 g/m² simulates the highest yield, followed by zone 30 with approx. 282 g/m² and zone 50 with 267 g/m² (fig.33 and fig. 35). Zone 30 thus experiences a negative deviation of 27 g/m² or 8.7 %/m² and 42 g/m² or 13.6 %/m² less WM is simulated for zone 50 (fig. 34). The quantitative measurement of the model quality also shows an improvement in the individual quality criteria (table 13). The RMSE for the control zone has improved from

Table 10: Values of sensitive parameters after PSO calibration

Parameter	Value
Nm	0.385999
D	1.870764
Qe	0.474279
T_CRIT	19.62227
alpha_F	0.892947
fF_	0.943268

97.6 to approx. 9.8. The RMSE also improved in the two shading zones from 84.3 to 9.1 for zone 30 and from 80.4 to 9.7 for zone 50. All other quality criteria also show a strong improvement. The MAE fell from a maximum of 51.7 (Z0) to a minimum of 5.0 (Z50). The r-RMSE and the ARV are now very close to 0, the %SEP multiplied by 100 was minimized from values up to over 2000 (Z50) to at least 21 and the R^2 is almost 1 in all three zones. Overall, the model performance of the reduced TOMGRO model for all three zones thus shows a strong improvement after applying the PSO algorithm with the most suitable method from 4.3.2 and allows the modeling of the WM with lower model uncertainty. Looking at all metrics together, Z0 has the best performance, closely followed by Z30 and Z50. Accordingly, the uncertainty is highest for Z50, which is primarily reflected in the higher ARV, r-RMSE and %SEP values. Z0 and Z30 have a similar and slightly lower uncertainty.

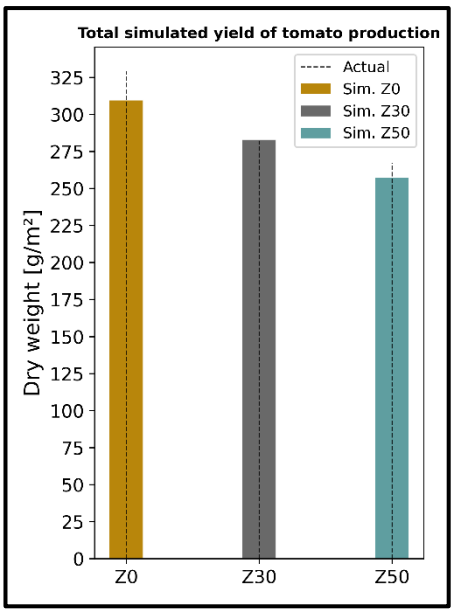

Figure 29: Simulated total yield of tomato production compared to the actual measured total yield.

Table 11: Measuring of the calibrated simulation's goodness-of-fit after finding the minimum fitness value or minimum mean RMSE of all three zones with PSO algorithm.

	RMSE	MAE	r-RMSE	%SEP	ARV	R^2
Z0	9.84	6.10	0.17	17.01	0.08	0.99
Z30	9.15	5.21	0.19	17.04	0.08	0.99
Z50	9.67	4.98	0.23	21.10	0.08	0.98

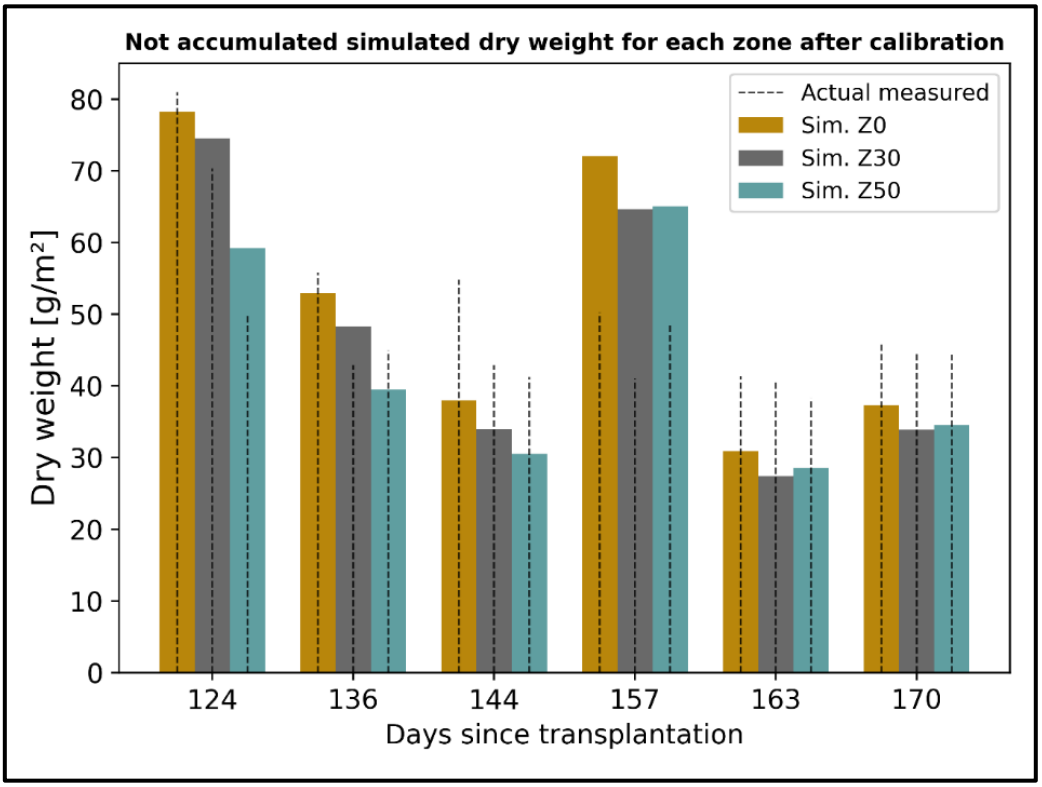

Figure 30: Comparison of the zones between non-accumulated simulated and actual WM after calibration on the individual measurement days from Table 3.

Figure 36 shows the values for the individual measurement days and thus the non-accumulated values. The improvement from simulated to actual values can be seen here again in comparison to Figure 32. Although the accumulated values between simulated and actual WM or total yield converge by the end of the growing season, the yields of the individual measurement days are not completely accurately mapped. In some cases, higher WM is simulated, especially for 20.02.24 (day 157 since transplantation), but often too little.

5 Discussion

In this chapter, the methods and results are discussed and interpreted. The measurements from the greenhouse are discussed first (5.1). Before the calibration methodology and its various applications and results (5.3) are discussed, the sensitivity analysis (5.2) is examined.

5.1 Actual measurements

First of all, it is important to mention that when measuring crop yields, using dry weight instead of fresh weight is often preferred, as uncertainty factors can be minimized. Measuring dry weight, for example, eliminates variability in the measurement data caused by the highly variable water content over time and environmental conditions. Although dry weight is directly less relevant for the economic yield aspect, it allows for more detailed analyses to assess the effectiveness of treatments and management practices, as the actual growth of the plant or tomato material can be analyzed without water content (Hort Americas, 2021; Huang et al., 2019). The uncertainty of the fresh weight should therefore be taken into account when considering the results. However, these results can serve as an estimate for the economic sales evaluation. Nevertheless, the results of the dry weight measurements should be given priority for all future plant physiological analyses and comparisons.

In addition to the general measurement uncertainties, it should also be noted that 20 plants were included per zone for the fresh weight for a representative yield measurement, while for logistical reasons only 6 out of 20 plants from a representative area were included for the dry weight. However, as only one oven would not have been sufficient for the yield of 60 plants (20 per zone), this compromise had to be made. Nevertheless, the results show a good correlation between fresh and dry weight.

Furthermore, it is assumed that the differences in radiation and temperature in the different zones influence the water content of the fruit and therefore the relative ratio between the zones are different for dry weight than for fresh weight (Heuvelink, 2018). The noticeable temperature difference in zone 50 (fig. 17) could also be the reason why the negative deviation compared to the control zone for fresh weight fell from 26% to only 18% for dry weight.

In addition, a temporal offset of the shading zones can be seen in the yield results. This was actually already apparent during the growth phase, as the plant grew more in height and leaf size, especially in zone 50. This means that less energy was invested

in the transition to the fruit development stage and more in plant growth. The greater the shade, the taller the tomato plants were. It could therefore be observed that the distances between the tomato branches were greater than in the control zone and therefore a lower fruit production was expected relatively early for zone 50. Tomato production is thus not only negatively influenced directly, but also indirectly, as the larger and more densely overgrown leaves can provide a larger attack surface for pests (Heuvelink, 2018).

Overall, the results of the actual measurements with the differences in the individual zones were to be expected. When extrapolating from g/m² to the total area of 440m² (20m x 22m) of the zones, the difference is of course more significant (cf. fig. 37). Nevertheless, it should be considered that the negative economic impact on tomato production due to shading by PV modules could be offset by a simultaneous positive economic impact due to electricity production.

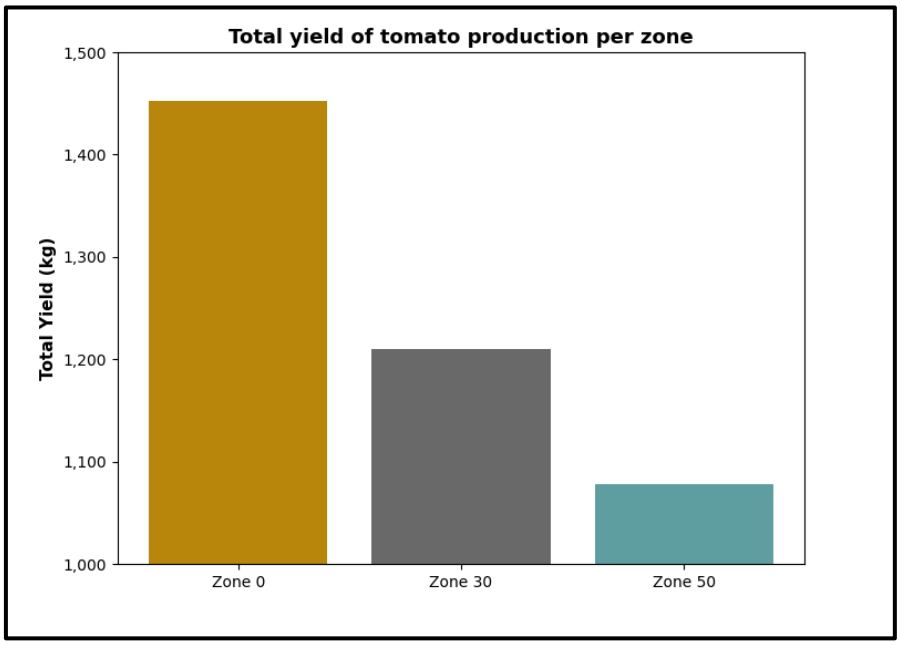

Figure 31: Total production of the respective zone.

5.2 Sensitivity analysis

The results of the sensitivity analysis correspond with the previous results from earlier studies on the sensitivity of the reduced TOMGRO parameters (Lin et al., 2019; Vazquez-Cruz et al., 2014). However, so far there have not been many studies that have applied a sensitivity analysis to the reduced TOMGRO model. A sensitivity analysis can be used to identify the parameters with the greatest influence on model uncertainty and to improve the model quality by adjusting the parameter values. The risk of overfitting the model is thus also reduced (Saltelli et al., 2004). It can therefore have advantages to carry out a sensitivity analysis that was omitted in other studies and thus possibly miss out on better results. For example, in the study by Gong et al. (2021), only all available parameters were used for calibration and no distinction was made between sensitive parameters. To be on the safe side and to clarify the extent to which a sensitivity analysis is useful or necessary at all, a calibration with all

parameters was also tried out. This means that not only the sensitive parameters were implemented in the calibration, but all 21 parameters. However, the results did not lead to further improvements in the model performance.

A different selection of methods, as mentioned in 3.2.2, could also have been considered for the sensitivity analysis. However, as the advantages of the eFAST method described in 3.2.2 outweigh those of other methods and the previous studies were also able to achieve relatively good results, no other method was tried. In addition, the existing studies could also serve as orientation in order to improve and expand their approach.

This aspect leads to another point of discussion: the definition of the value ranges of the individual parameters. As predefined value ranges are not available for all parameters and the literature sometimes even uses different value ranges for sensitivity analysis and calibration, a detailed literature search was carried out in order to cover all possible acceptable and potentially optimal values. Thus, values were also covered that were identified in other studies as the optimal value after calibration, but which may be unrealistic from a plant physiological point of view. As the results of the sensitivity analysis were consistent with the results of previous studies and the calibration also worked satisfactorily, the value ranges of the parameters have not been further changed.

Overall, the entire methodology could be expanded with the help of a sensitivity analysis compared to other studies. In this study, a sensitivity analysis using the eFAST method proved to be useful.

5.3 Calibration of reduced TOMGRO and simulation approach

5.3.1 PSO algorithm as a suitable method for calibrating the reduced TOMGRO model?

Calibrating the sensitive parameters of the reduced TOMGRO model is an optimization problem. This means that the parameters are optimized in such a way that the simulation values of WM of the reduced TOMGRO model come as close as possible to the actual values. There are several reasons for selecting an optimization algorithm and for selecting the PSO algorithm: The existing data basis without historical data is more suitable for an optimization algorithm than some other calibration methods, such as a machine learning method, which would require a broader set of historical training data sets. The PSO algorithm is also very easy to implement with its comparatively low computation time and required fewer hyperparameter adjustments than genetic algorithms. In addition, the study by Gong et al. (2021) showed that the PSO algorithm achieves better results than other optimization algorithms such as the genetic algorithm and the differential evolution algorithm. These aspects spoke in favor of selecting and implementing the PSO algorithm. However, even though the algorithm has a low susceptibility to premature convergence compared to other algorithms, it was observed that the algorithm can still get stuck in local minima and therefore did not find the expected global minimum in all experimental runs (Eberhart and Shi, 1998; Eslami

et al., 2012; Gad, 2022; Jain et al., 2022). However, this was not a major problem when multiple trials were started, as in most cases the algorithm was able to define the global minimum.

5.3.2 Gap filling to improve calibration?

Another point of discussion is the use of regression substitution for gap filling in the actual WM data. With regression substitution, the actual measured values are replaced by a regression model and then no longer correspond 100% to the actual measured values and therefore no longer correspond to reality. However, as the available data basis or the period of the growing season for a calibration is not particularly large anyway, the advantages of a larger and better data basis outweigh the disadvantages here. When trying out the calibration with only the individual measurement days as the basis for the actual measured values, reduced TOMGRO achieved acceptable but weaker results than with actual values on a daily basis. This allows the optimization algorithm to include values between the individual actual measurement days and thus better adjust the parameters that are of greater importance for the simulation between the individual days from a plant physiological point of view, so that reduced TOMGRO can perform a more complete and robust simulation with better timing.

5.3.3 Mean RMSE from all three zones as the best method for implementing the PSO algorithm?

As mentioned in 3.2.3.3, various methods for the application or implementation of the PSO algorithm in the reduced TOMGRO calibration were tried out. It is now necessary to discuss which method appears to be the most appropriate for this work. The method in which the model is calibrated individually for each zone produces the best results (cf. 4.3.2). However, since the aim of this work is to develop a uniform model for a greenhouse system that can also handle different shading or radiation (PPFD) and temperature depending on the input, this method is not expedient. The individual calibration results in different parameter values for each zone (cf. table 9), so that this does not result in a uniform model with a uniform parameter configuration for a greenhouse system. This also shows the importance of a validation period, which does not exist in this work due to the novel project and the first available growing season. In other words, a validation with an independent data set in order to check the accuracy and reliability of the model is normally needed. The problem becomes apparent when method b. is applied and the reduced TOMGRO model is started with the calibrated parameters of the control zone for the simulation of WM of the two shading zones, i.e. for an independent data set. For the control zone, the simulation represents the experimental data completely with a very low RMSE (8.30 g/m²), but for the two shading zones the simulation curve deviates significantly from the observed values. With this or these two methods, the problem of overfitting occurs. The model is then too strongly adapted to a calibration data set and can therefore be poorly generalized to a new data set (Murphy, 2012). Although there is no validation period for this work, this problem could be counteracted with the last method (c.) by including all three zones, and thus three different and independent data sets, in the calibration of a

uniform model for a greenhouse system. The same parameter values of the reduced TOMGRO model could now represent simulation curves that match the actual measured values from the greenhouse for all three zones, thus ensuring a more robust and universal model prediction.

5.3.4 Parameter values

Through calibration using the PSO algorithm with the mean RMSE from all three zones as a fitness function, the sensitive parameters were adjusted or optimized so that the simulation of the reduced TOMGRO model has the minimum error between the actual measured WM and the simulated WM (cf. table 11 and 12).

Nm, the maximum rate of node appearance, has changed from its default value of 0.5 to 0.386. If the parameter is considered individually, this means that the node appearance was slowed down and plant development is thus delayed. The model could thus reduce the overall yield and delay the time to fruit maturity or the timing of fruit production. However, a comparison of the simulation results before and after calibration shows that the total yield was lower in all zones before calibration and that ripening also started later than after calibration. This again shows the effect of the interaction between the individual parameters and underlines the importance of considering all parameters.

This is because the value for D, the coefficient to convert Pg from CO_2 to CH_2O also fell from the default value of 2.539 to 1.871 after calibration.

$$6CO_2 + 6H_2O + light \rightarrow C_6H_{12}O_6 + 6O_2 \quad (12)$$

Equation 12 shows the chemical reaction equation of photosynthesis. This shows that for every molecule of carbon dioxide and water, one molecule of CH_2O is produced, which describes the resulting biomass together with one molecule of oxygen (Encyclopedia Britannica, 2024). The reduction in value of D means that the model adapted or calibrated specifically to the 'raspa y amagado' greenhouse now has a lower efficiency in the conversion of photosynthesis products, such as glucose, into biomass (to CH_2O), which actually also leads to lower growth rates and a reduced tomato yield. However, the calibration resulted in a significant increase in the parameter Qe, i.e. leaf quantum eficiency. This rose from 0.065 to 0.474. This value now indicates that the tomato plants very efficiently and effectively convert photons absorbed by the leaves of the tomato plant into the production of photosynthesis products. The rate of photosynthesis has therefore been increased, which increases the growth and yield of the tomato plant. To summarize, the step of converting light energy into chemical energy (Qe), which must occur before the conversion into biomass (D), now has a significantly higher efficiency, while the conversion into biomass has a lower efficiency than before the calibration.

With T_crit, it is first of all helpful to take a closer look at the formula for $g(T_d)$ in order to recognize what the value of T_crit actually means and which optimal daytime temperature range results from it. $g(T_d)$ modifies partitioning to fruit and reduction of

growth under hot daytime conditions and thus shows at which daytime temperature (T_d) the fruit abortion starts (Jones et al., 1999). If the formula for $g(T_d)$ is now used (equation 13) and there is no partitioning of growth to the fruit, $g(T_d)$ is equal to 0. The 0.154 slope was based on the experiments of (Vallejos et al., 1997). If the formula is solved for T_d, this gives the temperature from which the equation becomes 0 and thus from which fruit abortion starts, i.e. T_crit added with 6.5 °C. According to Jones et al. (1999), the equation has the condition that T_d must be greater than T_crit. This means that if T_d is smaller than T_crit, there is no partitioning over the fruit. This results in a partitioning of growth on the fruit if T_d is greater than T_crit but less than T_crit added with 6.5.

$$g(T_d) < 1 - 0.154 \cdot (T_d - T\_crit) \quad (13)$$

$$0 < 1 - 0.154 \cdot (T_d - T\_crit)$$

$$-1 < 0.154 \cdot (T_d - T\_crit)$$

$$1 > 0.154 \cdot (T_d - T\_crit)$$

$$\frac{1}{0.154} > T_d - T\_crit$$

$$6.5°C > T_d - T\_crit$$

$$T_d < T\_crit + 6.5°C$$

Thus, with a value of 24.4 for T_crit, the optimal temperature range of T_d for splitting growth to fruit was 24.4 °C to 30.9 °C before calibration and 19.6 °C to 26.1 °C after calibration with a value of 19.6 for T_crit. This means that the tomato plants are more sensitive to high temperatures and the model tends towards earlier fruit abortion and reduced yield in warmer conditions, but the plants are less sensitive at medium temperatures. The fact that T_crit has such a low value is mainly due to the adaptation to zone 50, where the cooler temperatures are present. In this work, T_d was calculated from the average of the hourly temperatures between 8 a.m. and 5 p.m., as this allowed a T_crit value to be found with which all zones could be better simulated. This could be mainly due to the fact that temperature effects due to the greenhouse orientation could be compensated with a longer time period. If fewer daytime hours were included (e.g. 12 to 4 p.m.), the algorithm was unable to find an ideal value for T_crit at which zone 50 could be simulated well in addition to zones 0 and 30.

In addition, the algorithm identified a high value of 0.943 for fF_, i.e. the modification of the fruit partitioning based on temperature. This means that the reduced TOMGRO model adapts the partitioning more strongly to the temperature change or the temperature change is given a higher weight for the partitioning of growth to the fruit, which is indirectly transferred to T_crit and thus increases the influence on the model performance. At the same time, the value for alpha_F was slightly reduced compared to the default value. As a result, more biomass is potentially allocated to vegetative growth instead of the fruits than before the calibration. However, the value is still high, so that the proportion of biomass allocated to the fruit still results in a high yield.

Overall, the adjustment of the parameters Nm, D and alpha_F has more of an effect on reducing the yield compared to the simulation without calibration, at least when viewed individually. However, if the interaction of all parameters is considered, a clear increase in yield can be seen after calibration. This is due to the fact that the changes in the parameters Nm, D and alpha_F were rather slight in relation to the range of values, while the adjustments were stronger for the parameters Qe, T_crit and fF_. Nevertheless, the opposite adjustments to Nm, D and alpha_F should not be misunderstood. These changes are probably necessary in order to adjust the timing of fruit ripening and fruit production to the values actually measured in the greenhouse.

5.3.5 Model performance

The results of the calibration have shown that the application of the PSO algorithm to the reduced TOMGRO model has led to a significant improvement in model simulation and represents an efficient and effective method for calibrating complex models such as the reduced Tomgro model.

All error measures were significantly reduced or improved by the calibration. The differences or contradictions between the individual quality criteria and the individual zones are also interesting. For the RMSE and MAE, which are both a measure of the average error, but once with squaring and once without (cf. 3.2.3.4), the control zone has the highest value for RMSE. This indicates that the absolute errors or differences between observed and simulated values are greatest in this zone, which result from the conditions without shading. However, the control zone shows the best values for the relative error measures (r-RMSE, %SEP, ARV). This means that the model works best in the zone without shading or with the highest PPFD values when the errors are relativized. This means that if the errors are considered in relation to the size of the observed values, the shaded zones perform worse. This is because the r-RMSE and %SEP normalize the errors based on the observed values. Since the observed values of WM or the yield are highest in the control zone due to the apparently better growing conditions, the errors are smaller in relative terms, although the absolute difference is highest compared to the other zones. The variance (ARV) is also measured in relative terms. This means that the errors in the control zone are relatively more evenly distributed than in zones 30 and 50. The same applies to R^2, which looks at the proportion of the variance explained by the model in the total variance. This proves that the reduced TOMGRO model can explain the data in the control zone well with a slight advantage over the other zones despite the higher absolute errors.

Overall, the qualitative and quantitative results of the model performance for all three zones can be assessed as very good. The dry yield could be simulated well with the PSO calibrated reduced TOMGRO model for all shading levels. The slightly better model performance in the control zone despite higher absolute errors compared to the shaded zones can be explained by the higher yield values, as the error values between the actual measured and modelled WM decrease in relative terms due to the normalization and relativization of the errors. This shows how important it is to consider

several error measures in order to be able to comprehensively evaluate the model performance. At the same time, the values of the quality criteria show that there is still room for improvement. Further adjustments, additional calibration steps, or a further or longer growing season for a better data basis could be helpful for future work in order to further improve model accuracy.

5.4 Overall limitations

Although the results have shown that the biophysical model-based approach with the reduced TOMGRO model and its calibration with the PSO algorithm is very well suited as a tool for the simulation of tomato yield with different shading and an useful analysis of the tomato yield from different shaded zones could be carried out, this chapter should show and discuss the overall limitations and uncertainties of this work.

First of all, there are limitations in this work with regard to the data basis, which could not be avoided due to the novel and challenging design of the experiment. Although an initial project partner, such as the UAL-ANECOOP Foundation, was found to take on the role of a 'tomato grower' and allowed shading on their tomatoes, thereby risking a potential reduction in yield, it was not initially assumed that greenhouse crop modeling would be conducted during this growing season. Only the actual yields and changes in the tomato plants and the microclimate between the individual zones were to be analyzed. Before it was decided that modeling could also be carried out for this work, the growing season had already started and the tomato plants had been planted for six weeks. For this reason, although microclimatic measurements such as radiation, temperature and humidity had been taken since the beginning, numerous other (physiological) measurements that would be helpful for a complete model simulation with the reduced TOMGRO were not taken. In addition to the WM, the reduced TOMGRO model can also simulate four other state variables (cf. fig. 20), for which corresponding measurements such as LAI measurements, node counts, etc. should have been carried out from the outset. As a result, it was no longer possible to perform a complete calibration of all five state variables of the reduced TOMGRO for this growing season. Not only would the calibration have been much more complex, it would also have required significantly more capacity, which was not available within the context of this work and the cooperation project. At the same time, however, the simulation should then also achieve more representative, holistic and robust results, as it would then be based on a more comprehensive database and be able to make better predictions about various aspects of plant growth by considering several plant physiological measurements such as the LAI values and number of nodes. This would not only lead to an extended applicability for different scenarios with different aspects of plant growth, but also to an extended validation capability, as all calibrated state variables could be validated and the exact areas where the model quality needs to be improved could be identified. However, from an economic point of view and with regard to a future integrated agrivoltaic yield model that takes into account both crop and electricity production to determine the most lucrative combination of both yields, the focus on dry yield in crop modeling is sufficient in the context of this work.

In addition to the data basis, there are other uncertainties or several factors that cannot be considered that influence the growth process of the tomato plant and can therefore affect the accuracy of the simulation with reduced TOMGRO. These include information on soil, irrigation and even diseases or pests that are not directly included in reduced TOMGRO, as well as the assumption of a constant $C0_2$ content. Although this simplifies the modeling, the soil properties, the way in which the tomato plants are irrigated and the fluctuations in the CO_2 concentration in the greenhouse play a significant role in how the plant develops. The consequence of such simplifications is that the parameters in the calibration may incorrectly take on different values in order to achieve the best possible results, as additional parameters are missing and therefore certain stress factors or growth conditions may not be correctly mapped in a simulation with an independent data set.

The measurements of the sensors represent a further uncertainty. Technical limitations and defects, environmental influences and soiling always have a risk of inaccuracies that affect the database and thus impair the precision of the model. The sensors were therefore carefully calibrated in advance (Kujawa et al., 2023) and the data was also constantly checked and the sensors cleaned during the measurements in order to minimize uncertainties. Nevertheless, a residual risk remains, which should always be taken into account when evaluating the data.

In addition, the PAR was estimated by GHI using linear regression, as no PAR spectrometer was available for this work. Although this method has been shown to be suitable in past studies and the deviations are minimal, it should be kept in mind that it may introduce additional, albeit small, uncertainties, as the relationship between GHI and PAR is not always linear and can be influenced by environmental conditions (Vindel et al., 2018). However, the fact that the GHI was measured in this work is also relevant with regard to future PV modeling.

Another limitation of this work is the consideration of the shading zones in a greenhouse or that all three zones are located within the same greenhouse. Even if clear differences in the microclimatic conditions are recognizable in this work, there are temperature and other microclimatic mixtures between the zones within the greenhouse, which easily distort the individual conditions and thus make it difficult to evaluate their actual influence on the plant growth process. In addition, the greenhouse had an irrigation system that watered all plants equally, so that irrigation management would not be affected if one zone needed less or more water. Ideally, a separate greenhouse would be needed for each shading zone to ensure that the specific conditions of each zone could be studied in isolation and without interference from neighboring zones. It is possible that the differences in microclimatic conditions between the various shading levels would be even more pronounced and the effects on different yields of the zones would be greater. However, this can hardly be realized in practice due to limited capacities of any kind. Nevertheless, a good initial yield analysis and diagnostic model calibration of the actual state of the greenhouse system

used in this work could be carried out, which could also be used in reality with this design.

6 Conclusion and outlook

To address the issue of the incomplete understanding of the shading influence on tomato plants, this work demonstrates how the decreased irradiance intensity from installed PV dummies on the roof of a 'raspa y amagado' greenhouse affects tomato yield. Yield measurements from the greenhouse per square meter showed approximately 14% less dry yield and about 15% less fresh yield in the zone with 30% shading compared to the control zone, while in the 50% shading zone, there was a 19% reduction in dry yield and a 26% reduction in fresh weight compared to the control zone. Additionally, a delay in ripening and yield of about 10 days was observed compared to the control zone.

In relation to the modeling of an integrated photovoltaic yield model, which aims to identify the most lucrative method of combining cultivation and electricity production, this work focused exclusively on the plant growth component. It was found that the biophysical modeling approach, with the PSO calibrated reduced TOMGRO model and a sensitivity analysis using the eFAST method, is excellently suited for simulating tomato yield in a typical 'raspa y amagado' greenhouse in southern Spain, considering various shading conditions and effects. The results show good agreement and accurate performance between yield simulation with reduced TOMGRO and actual measured dry weight in all zones, with RMSE ranging from 9.15 to 9.84. Additionally, all other goodness-of-fit criteria show minimal deviations in model performance. Therefore, the use and adaptation of precise crop models to simulate and predict the yield of greenhouse tomatoes appears to be very advantageous and worthwhile based on the results. However, the different calibration methods with the PSO algorithm also indicate that a validation period with an independent growing season is advisable to ensure the reliability and transferability of the diagnostic model results. This should be considered for the next winter cycle for tomatoes.

In future work, the crop model developed in this study should be implemented into the overall integrated agrivoltaic greenhouse model. With various radiation values due to shading as input data for the model, tomato yield and simultaneous electricity production for each possible coverage combination could be calculated, determining the most lucrative method between electricity and crop production. In this context, an economic profitability analysis could also be conducted, including an analysis of costs such as operational, investment, and maintenance costs of both production systems, and revenues from the sale of tomatoes and generated electricity.

Furthermore, future work could also explore the estimation of other state variables besides modeling the WM yield, such as leaf area index, above-ground biomass, or number of nodes, while more parameters would need to be measured in the

greenhouse. This would create a more holistic and comprehensive modeling approach and could lead to the identification of additional plant physiological aspects.

Additionally, the methodology used in this work could be applied to various other greenhouse crops. Yield predictions could be made for, for example, potatoes or strawberries, etc.

Moreover, as in the study by Gong et al. (2023), the reduced TOMGRO as a biophysical model approach could be tested in model fusions with various combinations of more advanced machine learning models with a larger or growing data base to achieve even more accurate yield predictions.

References

AEMET (2024). Radiación y ozono. Radiación global, directa y difusa: Almería Aeropuerto. Global, directa y difusa. https://www.aemet.es/es/eltiempo/observacion/radiacion/radiacion?l=almeria. Accessed 20 April 2024.

ANECOOP S.COOP (2024). Tomate canario. E-Mail.

Apogee Instruments (2022). OWNER'S MANUAL Pyranometer Models SP-110 and SP-230. https://www.apogeeinstruments.com/content/SP-110-manual.pdf. Accessed 9 May 2024.

Aroca-Delgado, R., Pérez-Alonso, J., Callejón-Ferre, Á.-J., Díaz-Pérez, M. (2019). Morphology, yield and quality of greenhouse tomato cultivation with flexible photovoltaic rooftop panels (Almería-Spain). *Scientia Horticulturae* 257, 108768.

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

Bertin, N., Gary, C. (1993). TOMATO FRUIT-SET : A CASE STUDY FOR VALIDATION OF THE MODEL TOMGRO. *Acta Horticulturae* (328), 185–194.

Brogden, H.E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology* 37 (2), 65–76.

Chai, T., Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development* 7 (3), 1247–1250.

Cohen, S., Gijzen, H. (1998). THE IMPLEMENTATION OF SOFTWARE ENGINEERING CONCEPTS IN THE GREENHOUSE CROP MODEL HORTISIM1. *Acta Horticulturae* (456), 431–440.

Cossu, M., Cossu, A., Deligios, P.A., Ledda, L., Li, Z., Fatnassi, H., Poncet, C., Yano, A. (2018). Assessment and comparison of the solar radiation distribution inside the main commercial photovoltaic greenhouse types in Europe. *Renewable and Sustainable Energy Reviews* 94, 822–834.

Cossu, M., Yano, A., Solinas, S., Deligios, P.A., Tiloca, M.T., Cossu, A., Ledda, L. (2020). Agricultural sustainability estimation of the European photovoltaic greenhouses. *European Journal of Agronomy* 118, 126074.

Cukier, R., Levine, H., Shuler, K. (1978). Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics* 26 (1), 1–42.

Dayan, E., van Keulen, H., Jones, J.W., Zipori, I., Shmuel, D., Challa, H. (1993). Development, calibration and validation of a greenhouse tomato growth model: I. Description of the model. *Agricultural Systems* 43 (2), 145–163.

Dinesh, H., Pearce, J.M. (2016). The potential of agrivoltaic systems. *Renewable and Sustainable Energy Reviews* 54, 299–308.

Dueck, T., Janse, J., Li, T., Kempkes, F., Eveleens, B. (2012). INFLUENCE OF DIFFUSE GLASS ON THE GROWTH AND PRODUCTION OF TOMATO. *Acta Horticulturae* (956), 75–82.

Eberhart, R.C., Shi, Y. (1998). Comparison between genetic algorithms and particle swarm optimization. In: Porto, V.W. (Ed.) *Evolutionary programming VII*. *Proceedings*, vol. 1447. Springer, Berlin, London, pp. 611–616.

Elbes, M., Alzubi, S., Kanan, T., Al-Fuqaha, A., Hawashin, B. (2019). A survey on particle swarm optimization with emphasis on engineering and network applications. *Evolutionary Intelligence* 12 (2), 113–129.

Elitech Technology (n.d.). RC-51 Multi-use Temperature Data Logger. https://www.elitechlog.com/staticfile/upload/file/20240415/rc-51-instructions.pdf. Accessed 7 May 2024.

Encyclopedia Britannica (2024). Photosynthesis | Definition, Formula, Process, Diagram, Reactants, Products, & Facts. https://www.britannica.com/science/photosynthesis. Accessed 1 July 2024.

Eslami, M., Shareef, H., Khajehzadeh, M., Mohamed, A. (2012). A Survey of the State of the Art in Particle Swarm Optimization. *Research journal of Applied Sciences, Engineering and Technology* 4, 1181–1197.

Ezzaeri, K., Fatnassi, H., Bouharroud, R., Gourdo, L., Bazgaou, A., Wifaya, A., Demrati, H., Bekkaoui, A., Aharoune, A., Poncet, C., Bouirden, L. (2018). The effect of photovoltaic panels on the microclimate and on the tomato production under photovoltaic canarian greenhouses. *Solar Energy* 173, 1126–1134.

Fink, M., Daniels, A., Qian, C., Velásquez, V.M., Salotra, S., Wollherr, D., 2023. Comparison of Dynamic Tomato Growth Models for Optimal Control in Greenhouses.

Fraunhofer ISE (n.d.). Agrivoltaics. Fraunhofer Institut for Solar Energy Systems ISE. https://www.ise.fraunhofer.de/en/press-media/press-releases.html. Accessed 30 April 2024.

Gad, A.G. (2022). Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review. *Archives of Computational Methods in Engineering* 29 (5), 2531–2561.

García-Martínez, C., Rodriguez, F.J., Lozano, M. (2019). Genetic Algorithms. In: Martí, R., Pardalos, P.M., Resende, M.G.C. (Eds.) *Handbook of heuristics*, [Enhanced Credo edition] ed. Springer; Credo Reference, Cham, Switzerland, Boston, Massachusetts, pp. 431–464.

Gelman, A., Hill, J. (2021). *Data analysis using regression and multilevel/hierarchical models,* 23rd printing ed. Cambridge University Press, Cambridge.

Gong, L., Yu, M., Cutsuridis, V., Kollias, S., Pearson, S. (2023). A Novel Model Fusion Approach for Greenhouse Crop Yield Prediction. *Horticulturae* 9 (1), 5.

Gong, L., Yu, M., Jiang, S., Cutsuridis, V., Kollias, S., Pearson, S. (2021). Studies of evolutionary algorithms for the reduced Tomgro model calibration for modelling tomato yields. *Smart Agricultural Technology* 1, 100011.

Herman, J., Usher, W. (2017). SALib: An open-source Python library for Sensitivity Analysis. *The Journal of Open Source Software* 2 (9), 97.

Hernández-Hernández, F., Lopez-Cruz, I., Guevara-Gonzalez, R., Rico-García, E., Herrera-Ruiz, G., Chavira, M., Pacheco, I. (2011). Growth and development

simulation of pepper (Capsicum Annumm L.) under greenhouse conditions. *Revista Mexicana de Ciencias Agricuolas* 2, 385–397.
Heuvelink, E. (1999). Evaluation of a Dynamic Simulation Model for Tomato Crop Growth and Development. *Annals of Botany* 83 (4), 413–422.
Heuvelink, E. (2018). *Tomatoes,* 2nd ed. ed. CABI Pub, Wallingford.
Heuvelink, E., Bertin, N. (1994). Dry-matter partitioning in a tomato crop: Comparison of two simulation models. *Journal of Horticultural Science* 69 (5), 885–903.
Hort Americas (2021). Fresh Weight vs. Dry Weight. https://hortamericas.com/blog/fresh-weight-vs-dry-weight/. Accessed 13 June 2024.
Horton, W., Doxas, I. (1998). A low-dimensional dynamical model for the solar wind driven geotail-ionosphere system. *Journal of Geophysical Research: Space Physics* 103 (A3), 4561–4572.
Huang, W., Ratkowsky, D., Hui, C., Wang, P., Su, J., Shi, P. (2019). Leaf Fresh Weight Versus Dry Weight: Which is Better for Describing the Scaling Relationship between Leaf Biomass and Leaf Area for Broad-Leaved Plants? *Forests* 10 (3), 256.
Iwanaga, T., Usher, W., Herman, J. (2022). Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environmental Systems Modelling* 4, 18155.
Jain, M., Saihjpal, V., Singh, N., Singh, S.B. (2022). An Overview of Variants and Advancements of PSO Algorithm. *Applied Sciences* 12 (17), 8392.
Jones, J.W., Dayan, E., Allen, L.H., van Keulen, H., Challa, H. (1991). A DYNAMIC TOMATO GROWTH AND YIELD MODEL (TOMGRO). *Transactions of the ASAE* 34 (2), 663–672.
Jones, J.W., Kenig, A., Vallejos, C.E. (1999). REDUCED STATE–VARIABLE TOMATO GROWTH MODEL. *Transactions of the ASAE* 42 (1), 255–265.
Katsikogiannis, O.A., Ziar, H., Isabella, O. (2022). Integration of bifacial photovoltaics in agrivoltaic systems: A synergistic design approach. *Applied Energy* 309, 118475.
Kempkes, F., Swinkels, G.J., Hemming, S. (2018). Increase of light transmission of a Venlo-type greenhouse during winter by 10%: a design study. *Acta Horticulturae* (1227), 133–140.
Kennedy, J., Eberhart, R. (1995). Particle swarm optimization. Paper presented at the ICNN'95 - International Conference on Neural Networks, Perth, WA, Australia, 27 Nov.-1 Dec. 1995.
Kujawa et al., 2023. Modeling of bifacial agrivoltaic greenhouses in southern Spain: Poster at EUPVSEC 2023. EUPVSEC, 2023, Lisbon, Portugal.
Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., Monod, H. (2009). Multivariate global sensitivity analysis for dynamic crop models. *Field Crops Research* 113 (3), 312–320.
Lin, D., Wei, R., Xu, L. (2019). An Integrated Yield Prediction Model for Greenhouse Tomato. *Agronomy* 9 (12), 873.

Little, R.J.A., Rubin, D.B. (2020). *Statistical analysis with missing data,* 3rd ed. ed. Wiley, Hoboken, NJ.
López-Cruz, I.L., Rojano-Aguilar, A., Salazar-Moreno, R., Ruiz-García, A., Goddard, J. (2012). A COMPARISON OF LOCAL AND GLOBAL SENSITIVITY ANALYSES FOR GREENHOUSE CROP MODELS. *Acta Horticulturae* (957), 267–273.
López-Díaz, G., Carreño-Ortega, A., Fatnassi, H., Poncet, C., Díaz-Pérez, M. (2020). The Effect of Different Levels of Shading in a Photovoltaic Greenhouse with a North–South Orientation. *Applied Sciences* 10 (3), 882.
Memmert GmbH + Co.KG (n.d.). Universal Oven U: Lab oven for universal use Universal heating oven and drying oven. https://www.memmert.com/products/heating-drying-ovens/universal-oven/#!filters=%7B%7D. Accessed 13 May 2024.
Miranda, L.J.V., Moser, A., Cronin, S.K. (2017). PYSWARMS - a research toolkit for Particle Swarm Optimization in Pythn. https://pyswarms.readthedocs.io/en/latest/index.html. Accessed 17 April 2024.
Montgomery, D.C., Peck, E.A., Vining, G.G. (2021). *Introduction to linear regression analysis,* Sixth edition. ed. John Wiley & Sons, Inc, Hoboken.
Morris, M.D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics* 33 (2), 161–174.
Mulla, D.J. (2003). Mathematical Models of Small Watershed Hydrology and Applications. *Journal of Environmental Quality* 32 (1), 374.
Murphy, K.P. (2012). *Machine learning*. *A probabilistic perspective.* MIT Press, Cambridge, Mass.
Næs, T., Isaksson, T., Fearn, T., Davies, T. (2017). *A user-friendly guide to Multivariate Calibration and Classification.* IM Publications Open.
OTT HydroMet B.V. (2016). SP Lite2 Pyranometer. https://www.kippzonen.com/Product/9/SP-Lite2-Pyranometer. Accessed 7 May 2024.
OTT HydroMet B.V. (2023). CMP10 Pyranometer. https://www.kippzonen.com/Product/276/CMP10-Pyranometer. Accessed 7 May 2024.
Ozer, D.J. (1985). Correlation and the coefficient of determination. *Psychological bulletin* 97 (2), 307.
Ramírez, A., Rodríguez, F., Berenguel, M., Heuvelink, E. (2004). CALIBRATION AND VALIDATION OF COMPLEX AND SIMPLIFIED TOMATO GROWTH MODELS FOR CONTROL PURPOSES IN THE SOUTHEAST OF SPAIN. *Acta Horticulturae* (654), 147–154.
Rubio, A. (2022). El proyecto Lygalán o la búsqueda del sabor en tomate. https://agroautentico.com/2022/03/el-proyecto-lygalan-o-la-busqueda-del-sabor-en-tomate/. Accessed 8 May 2024.
Saltelli, A. (1999). Sensitivity analysis: Could better methods be used? *Journal of Geophysical Research* 104, 3789–3793.

Saltelli, A. (2008). *Global sensitivity analysis*. *The primer.* Wiley, Chichester, West Sussex.
Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M. (2004). Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models.
Saltelli, A., Tarantola, S., Chan, K. (2012). A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics* 41.
Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data,* 1st ed. ed. Chapman & Hall, London.
Smajgl, A., Ward, J., Pluschke, L. (2016). The water–food–energy Nexus – Realising a new paradigm. *Journal of Hydrology* 533, 533–540.
Sobol, I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* 55, 271–280.
Suganthan, P.N. (2012). Differential Evolution Algorithm: Recent Advances. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Dediu, A.-H., Martín-Vide, C., Truthe, B. (Eds.) *Theory and Practice of Natural Computing*, vol. 7505. Springer Nature, Berlin, Heidelberg, pp. 30–46.
Tibshirani, R., Hastie, T., Witten, D., James, G. (2021). *An Introduction to Statistical Learning: With Applications in R.* Springer, New York, NY.
Ureña-Sánchez, R., Callejón-Ferre, Á.J., Pérez-Alonso, J., Carreño-Ortega, Á. (2012). Greenhouse tomato production with electricity generation by roof-mounted flexible solar panels. *Scientia Agricola* 69 (4), 233–239.
Valera, D.L., Belmonte, L.J., Molino-Aiz, F.D., Lopez, A. (2016). *Greenhouse Agriculture in Almeria. A comprehensive techno-economic analysis.* Cajamar Caja Rural.
Vallejos, C.E., Jones, J.W., Williams, F.W. (1997). *High temperature tomato experiments, Ch. II-3*.
Vanthoor, B., Stanghellini, C., van Henten, E.J., Visser, P. de (2011). A methodology for model-based greenhouse design: Part 1, a greenhouse climate model for a broad range of designs and climates. *Biosystems Engineering* 110 (4), 363–377.
Varella, H., Guérif, M., Buis, S. (2010). Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. *Environmental Modelling & Software* 25 (3), 310–319.
Vazquez-Cruz, M.A., Guzman-Cruz, R., Lopez-Cruz, I.L., Cornejo-Perez, O., Torres-Pacheco, I., Guevara-Gonzalez, R.G. (2014). Global sensitivity analysis by means of EFAST and Sobol' methods and calibration of reduced state-variable TOMGRO model using genetic algorithms. *Computers and Electronics in Agriculture* 100, 1–12.
Vindel, J., Valenzuela, R., Navarro, A., Zarzalejo, L., Paz-Gallardo, A., Souto, J., Méndez-Gómez, R., Cartelle, D., Casares, J. (2018). Modeling Photosynthetically

Active Radiation from Satellite-Derived Estimations over Mainland Spain. *Remote Sensing* 10 (6), 849.

Vos, J., Evers, J.B., Buck-Sorlin, G.H., Andrieu, B., Chelle, M., Visser, P.H.B. de (2010). Functional-structural plant modelling: a new versatile tool in crop science. *Journal of experimental botany* 61 (8), 2101–2115.

Willmott, C.J., Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30, 79–82.

Willockx, B., Lavaert, C., Cappelle, J. (2022). Geospatial assessment of elevated agrivoltaics on arable land in Europe to highlight the implications on design, land use and economic level. *Energy Reports* 8, 8736–8751.

Yonryu (2020). TOMGRO with LSTM · GitHub. https://gist.github.com/gyosit/ce2b2fd726f3256d6796ed2bfa8a37e0#comments. Accessed 26 June 2024.

Zhao, C., Liu, B., Xiao, L., Hoogenboom, G., Boote, K.J., Kassie, B.T., Pavan, W., Shelia, V., Kim, K.S., Hernandez-Ochoa, I.M., Wallach, D., Porter, C.H., Stockle, C.O., Zhu, Y., Asseng, S. (2019). A SIMPLE crop model. *European Journal of Agronomy* 104, 97–106.