Dataset of tomato plants growth observations obtained from multiple sources in a production-like setting

Monique Oliveira, Rafaella Amaro, Henrique Pescarini, Luiz Henrique Rodrigues

https://doi.org/10.1590/SciELOPreprints.7667

Submitted on: 2023-12-08
Posted on: 2023-12-11 (version 1)
(YYYY-MM-DD)

Preprint of manuscript *Dataset of tomato plants growth observations obtained from multiple sources in a production-like setting* This version of the manuscript has not been peer-reviewed.

Article information

Article title

Dataset of tomato plants growth observations obtained from multiple sources in a production-like setting

Authors

Monique Pires Gravina de Oliveira[1a*], Rafaella Pironato Amaro[1b], Henrique Baeninger Pescarini[1c], Luiz Henrique Antunes Rodrigues[1d]

Affiliations

[1]Universidade Estadual de Campinas (UNICAMP). Faculdade de Engenharia Agrícola – FEAGRI. Av. Cândido Rondon, 501 - Barão Geraldo 13083-875 – Campinas – SP, Brazil

[a] https://orcid.org/0000-0001-7167-4473

[b] https://orcid.org/0000-0002-2186-0857

[c] https://orcid.org/0000-0002-9044-8831

[d] https://orcid.org/0000-0002-1756-7367

Corresponding author's email address and Twitter handle

moniquepgoliveira@gmail.com

Keywords

Protected growth, Computer vision, Image segmentation, Greenhouse, Digital agriculture

Abstract

This dataset contains observations of tomato growth in a production-like setting, at research greenhouses. Two plants in each of three growth cycles were continuously monitored and pictures were taken every other day from above and from a side view while a weighting system was used to record changes in weight of plant and water in the substrate. Other plants in the environment were subjected to destructive analysis in general every two weeks to quantify aspects of growth that required destructive measurements, such as dry weight and plant leaf area, and these records are also included in the dataset, including the scans of digitized leaves. Plant samples destined to destructive measurements also had their pictures taken before removal. In total, 618 photos of monitored and removed plants were annotated, and masks of leaf, fruit and mature fruit are also provided. The dataset also includes measurements of photosynthetically active radiation and air temperature recorded inside the greenhouses by two sets of different sensors during the growth cycles. The dataset allows for applications regarding growth monitoring and computer vision tasks.

Specifications table

Subject	Agricultural Sciences: Horticulture; Data science: mathematical modelling; Computer science: Computer Vision and Pattern Recognition
Specific subject area	Tomato growth data obtained for crop modeling and for computer vision applications.
Type of data	Tomato plant RGB images Masks of tomato plants' organs (Image) Digitized tomato leaves RGB images Tomato plant growth data tables Environmental monitoring data tables
How the data were acquired	During three growth cycles in research greenhouses, two tomato plants were continuously monitored, using load cells and fixed cameras. Other tomato plants grown in the same environment were subjected to destructive analysis for dry mass and leaf area characterization, as well as non-destructive characterization, through pictures taken with a smartphone from a top-down and a lateral view. All images from cameras were annotated at the pixel-level using the GIMP software. Environmental data including air temperature, humidity, and photosynthetically active radiation were measured during the growth cycles by two sets of different sensors for each variable.
Data format	Raw Analyzed Filtered

Description of data collection	Three growth cycles in which plant growth was continuously monitored. Growth resembled a production-like environment, with no specific treatment, as the goal was obtaining data representative of growth. While monitoring pictures were obtained every half hour, from each camera, only one in alternate days was annotated and are included in the dataset. Other observations are provided as collected.
Data source location	Institution: School of Agricultural Engineering, University of Campinas City/Region: Campinas, São Paulo Country: Brazil Latitude and longitude: 22º 49' 06” S, 47º 03' 40” W, 635 m altitude).
Data accessibility	Repository name: Tomato growth in production-like setting — Repositório de Dados de Pesquisa da Unicamp. Data ID: 10.25824/redu/EP4NGO Direct URL to the data: https://doi.org/10.25824/redu/EP4NGO
Related research article	

Value of the data

- This dataset includes 618 annotated images of tomato plants. Images were annotated in the pixel level indicating three categories: leaves, and green and mature fruits. These images can be used for computer vision segmentation tasks.
- Tomato growth data from production-like environments, including data from the environmental conditions and destructive and non-destructive observations from plants, allows for characterizing the relationship between growth and environment. Such measurements are often useful for modeling purposes.
- Destructive and non-destructive characterization of the same tomato plant are available, which allows for establishing relationship between direct and indirect measurements.
- Digitized images of leaves are included, which allows for developing algorithms for determining plant leaf area.

Objective

This dataset was obtained within a project aimed at performing data assimilation in a tomato growth model. As new sources of data become available, new approaches for intertwining real-time observations and simulation models must be assessed for extracting the most of them. Since deep learning techniques have not yet been explored for processing the images collected in the research project, either for area quantification or for obtaining observation models, the dataset provides annotated images so that the approaches could be explored. It also includes additional data connected to growth that could be included in growth simulation modeling studies exploring different techniques.

Data description

All data described in the article is included in the online repository [1]. Three main categories for data: destructive and non-destructive growth observations and environmental data (Table 1). Growth data from destructive analysis comprises measurements of leaf area, aboveground fresh and dry mass separated into leaves, green fruits, and mature fruits. Environmental data includes measurements of relative humidity, air temperature, photosynthetically active radiation, and luminosity from inside the greenhouse environment. Non-destructive observations include continuous weight monitoring of two plants and pictures taken from top-down and lateral views of two monitored plants and of plants subjected to destructive analysis. Data set also includes masks for leaves, green fruits, and mature fruits.

Table 1. Description of dataset files.

File group	File description	Number of files
Destructive	Weight (fresh and dry) from tomato plants, separated in leaf, stem, and green and mature fruits for all growth cycles. Cycle 02 includes number of nodes and Cycle 03, number of nodes and plant height. Separated by growth cycle and organized in .xlsx files. Different variables are presented in different tabs.	3 .xlsx files
Leaf area index (LAI) data	Leaf area data consists in .png files of leaf scans. The ID of each plant is indicated in the name of the file, as is the date of the analysis. For Cycle 01, size reference is a black rectangle (length 5.7 cm and area 50.73 cm^2). For Cycle 02 and Cycle 03, size reference is a pink post-it (length 5 cm and area 19 cm^2).	6 .zip files
PAR	Photosynthetically active radiation obtained by quantum sensors Licor LI-190SA with a datalogger Licor LI-1400, using range from 400 to 700 nm. Organized in .txt files.	1 .zip file
Pixel-level plant organs masks	Annotations of leaves and fruits in pictures taken of 1. plants used in destructive analyses and 2. monitored plants. Size references are detailed in a text file inside the zip. Organized in .png files.	1 .zip file
Plants' photos	Photos taken of plants 1. used in destructive analyses and 2. continuously monitored. Size references are described on a txt with the masks' files. Most photos from monitoring cameras are .png files, but some are .jpg. Photos of plants used in destructive analyses are .jpg files.	1 .zip file
Scripts	Scripts used in the processing of raw data.	1 .zip file
Substrate moisture	Substrate moisture measured in mV with EC-05 sensors. Organized in .csv files.	1 .zip file
Temperature, Relative Humidity - Transdutors	Air temperature and relative humidity obtained with SHT75 transducers protected by porous capsules. Data organized in .csv files.	1 .zip file
Temperature, Relative Humidity, Luminosity - Pi	Air temperature and relative humidity obtained with DHT22 sensors and luminosity obtained with BH1750 sensors. All sensors were connected to a Raspberry Pi model B computers. Data organized in .csv files.	1 .zip file
Weight	Weight data for each minute during growth cycles for each of the two monitored plants. Measurements in newton [N]. Data organized in .csv files.	1 .zip file

Experimental design, materials and methods

Growth infrastructure

The experiments were conducted in research greenhouses at the School of Agricultural Engineering of the University of Campinas (22º 49' 06" S, 47º 03' 40" W, 635 m altitude). Four cycles of minitomatoes growth were performed (Table 2).

Table 2. Summary of growth cycles for data gathering.

Growth Cycle	Cultivar	Start date	End date
Cycle 0	Seminis – DRC 564	11/jan/2019	04/apr/2019
Cycle 1	Fercam – Milla	12/jul/2019	28/oct/2019
Cycle 2	Feltrin – Carolina	05/nov/2020	12/feb/2021
Cycle 3	Seminis – DRC 564	16/mar/2021	11/jun/2021

- Cycle 0

This growth cycle was conducted in the same setting of Cycles 1 and 2 described below, but as monitoring was not completely established from the beginning, it was treated as a pilot to test the installation of sensors, the irrigation system and to obtain experience in growing the plants.

- Cycle 1 and Cycle 2

These growth cycles were conducted in a research greenhouse with 6.4 m of width, 10.98 m of length, 3.0 m of height from the floor to the gutter and 4.5 m of total height. The greenhouse has a gable roof covered with low density polyethylene of 150 µm width with light diffuser and anti-UV treatment. The ridge of the greenhouse section was oriented North-South. The section was only bounded to the East, by another greenhouse of the same dimensions. Its cooling system consisted of a pad-fan system, activated by a scheduling device. The South-facing side wall was covered with an evaporative cooling pad and an insect screen. Other walls were covered with the same plastic as the cover. Seedlings provided by commercial units were transplanted to polyethylene pots (8 L) filled with coconut fiber approximately 30 days after seeding. They were distanced 1.5m x 0.5m (1.33 plants m^{-2}). Figure 1 shows the overall disposition of pots.

Figure 1. Disposition of the pots in the first research greenhouse, used for cycles one and two.

- Cycle 3

The experiment was conducted in a research greenhouse with dimensions 6.4 m of width, 18 m of length and 3.0 m of height from the floor to the gutter. It has a gable roof covered with low density polyethylene, light diffuser with 150 μm width and anti-UV treatment. The ridge of the greenhouse section was oriented North-South. The section was not bounded in any directions. All walls were covered by an insect screen. Locally cultivated seedlings were transplanted to polyethylene pots (8 L), filled with coconut fiber, distanced 0.9 m x 0.5 m (2.22 plants m^{-2}) approximately 30 days after seeding. Lines of pots were intercalated with lines of nutrient film hydroponic growth (Nutrient Film Technique – NFT), which were not used in this project. Figure 2 shows the environment. Cycle 3 also included application of neem oil, Bordeaux mixture and lime-sulfur prevent the occurrence of pests and an abamectin-based pesticide (Syngenta's VERTIMEC® 18 EC) after appearance of rust mite.

Figure 2. Disposition of the pots in the second research greenhouse used, for cycle three.

Management practices

All cycles consisted of around 100 days and in all of them plants had reached the highest wire, when the growth was stopped. Management practices (thinning, staking, pruning, pest management and diseases) overall followed the recommendations for hydroponic growth in Alvarenga (2013) [2]. Only one stem was grown per pot. Removal of side shoots happened once to three times a week. Leaves were pruned only when their senescence was dominant. Harvest happened when the whole truss was mature. Irrigation consisted of fertigation through a drip irrigation system and while the nutritive solution mainly followed the recommendations in Pires et al. (2011) [3], concentrations were changed according to plants' responses. In Cycles 0 and 1, irrigation length used fixed time through the cycle, but in Cycle 1 the duration proved insufficient by the end of the cycle. In Cycles 2 and 3, total irrigation time was defined as that which would not allow for deficit by the end of the day, to minimize mass fluctuations from one day to the following caused by variation in irrigations. Cycle 1 showed water deficits throughout growth. Cycle 2 suffered with excessive nitrogen fertilization followed by rust mite while Cycle 3 more closely resembled full irrigation and fertilization.

Data

Environmental data

Sensors used for gathering environmental data were characterized as scientific grade (SG) and low-cost (LC) (Table 3). The scientific grade sensors for temperature and relative humidity corresponded to SHT75 transducers protected by porous capsules which, by their turn, were protected by polyvinylchloride tubes, with downstream fans in the bottom, coated with aluminum foil. The sensors were installed in a hardware platform for wireless sensor networks (Radiuino BE900), with daily backup. For photosynthetically active radiation (PAR), the scientific grade sensors corresponded to quantum sensors Licor LI-190SA with a datalogger Licor LI-1400. Low-cost sensors were connected to Raspberry Pi model B computers.

Table 3. Sensors and frequency of data acquisition used for monitoring the environment.

Variable	Type	Model	Frequency
Air temperature	SG	SHT75	5 minutes
Relative humidity	SG	SHT75	5 minutes
Air temperature	LC	DHT22	5 minutes
Relative humidity	LC	DHT22	5 minutes
Substrate moisture	SG	EC-05	10 minutes
Luminosity	LC	BH1750	5 minutes
Photosynthetically active radiation	SG	LI190SA	15 minutes

Each sensor node was positioned close to one of the monitored plants and each node included two sensors of the same type and for the same variable for redundancy, except for radiation and luminosity, which only include one of each type. As there were differences in the experimental set-ups, they are separately detailed as follows. Sensors were positioned as in Figure 3 for Cycles 0 to 2 and as in Figure 4 for Cycle 3.

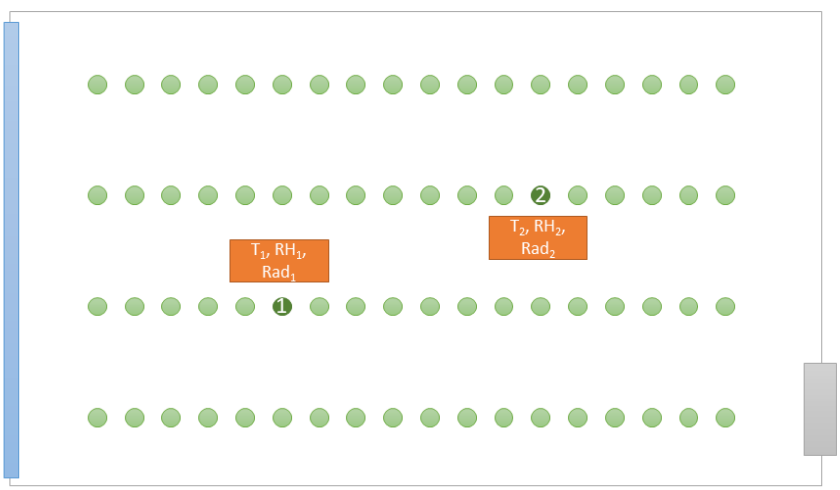

Figure 3. Positions of sensors of temperature (T), relative humidity (RH), and PAR and luminosity (Rad) during growth cycles 0 to 2. Gray rectangle refers to the door and blue rectangle refers to the wet pad. Green circles correspond to the vases. Monitored plants are highlighted in dark green. Distances are at scale.

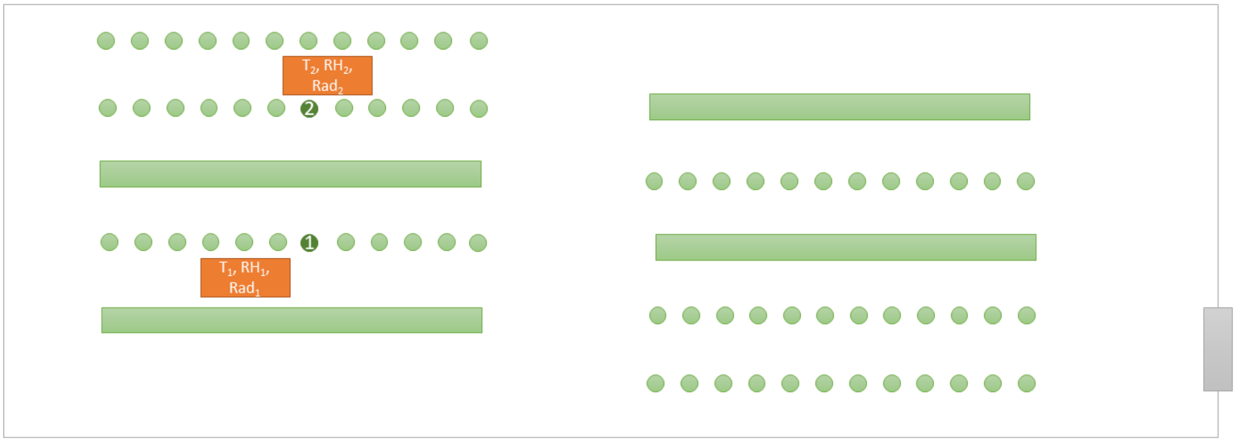

Figure 4. Positions of sensors of temperature (T), relative humidity (RH), and PAR and luminosity (Rad) during growth cycle 3. Gray rectangle refers to the door and green rectangles refer to the NFT grown tomatoes. Green circles correspond to the vases. Monitored plants are highlighted in dark green. Distances are at scale.

Growth data

In each growth cycle, two plants were continuously monitored through the weighting system and by two cameras, positioned above the plant and in an adjacent row. Continuous weight monitoring used force transducers HBM S2M with nominal force of 10 N (0.02 % accuracy) and stored in a data logger PMX WGX002, measure card PX455, and to the pictures taken from side (Figure 5) and top-down (Figure 6) views, with fixed Raspberry Pi Camera Modules v2, connected to Raspberry Pi Zero computers.

Figure 5. Example of image captured by fixed camera in an adjacent row, along with respective annotation Green tomatoes are marked in yellow and leaves are marked in green. Reference is marked in cyan. Left: Photo with overlayed label. Center: Photo as captured. Right: Only label (dark background is merely to increase contrast and is transparent in original image).

Figure 6. Example of image captured by fixed camera above the plant, along with respective annotation Leaves are marked in green. Reference is marked in cyan. Left: Photo with overlayed label. Center: Photo as captured. Right: Only label (dark background is merely to increase contrast and is transparent in original image).

Every one to three weeks, three plants were removed and used for destructive analysis. Keeping guard plants of the destructed plants was not always possible and could affect the results. Before being removed, plants destined for analysis were first photographed from above (Figure 7) and laterally (Figure 8) with a smartphone camera. Plant material was weighted before and after drying for four days or as until constant weight was reached. Leaves, stem and green and mature fruits were separated for weighting. After being weighted, while fresh, leaves were digitized with a scanner (Figure 9). Digitization included a reference of known dimensions.

Figure 7. Example of image captured by a smartphone, from a plant destined to destructive analysis, from the top-down angle, along with respective annotation Leaves are marked in green. Reference is marked in cyan. Left: Photo with overlayed label. Center: Photo as captured. Right: Only label (dark background is merely to increase contrast and is transparent in original image).

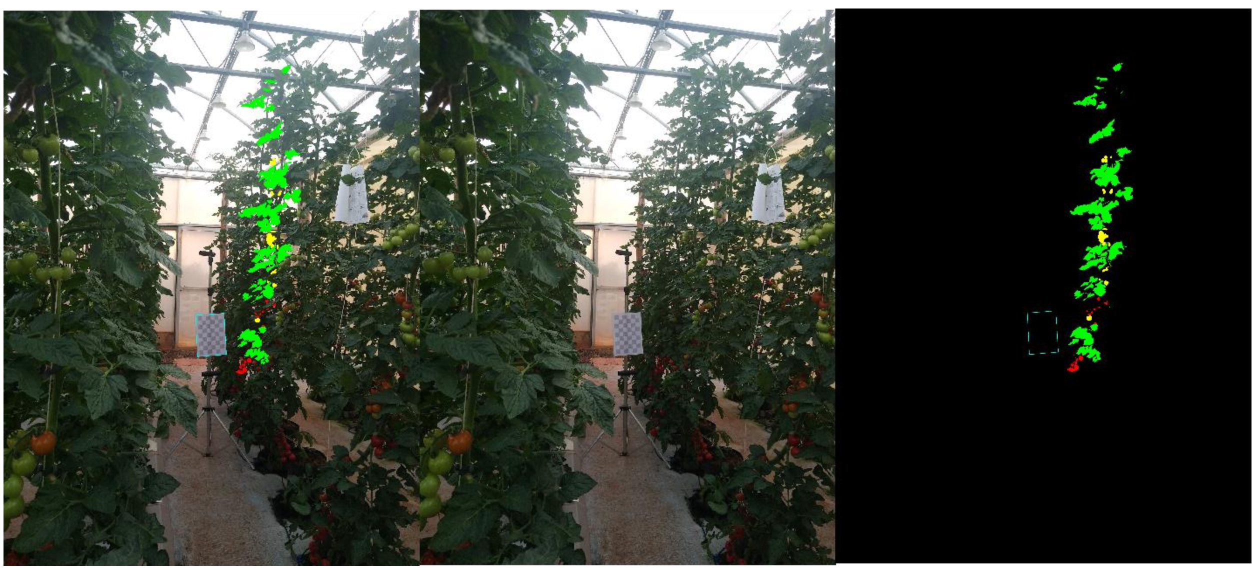

Figure 8. Example of non-destructive observation of lateral leaf and fruit areas of plant destined to destructive analysis, along with respective annotation. Green tomatoes are marked in yellow and leaves are marked in green. The A4 sheet used as reference is marked in cyan. Left: Photo with overlayed label. Center: Photo as captured. Right: Only label (dark background is merely to increase contrast and is transparent in original image).

Figure 9. Example (resized) of leaf scans from Cycle 1 (left) and Cycle 3 (right), including the references of dimensions that were used to calculate the area corresponding to leaf in the picture.

Figures Figure 5 to Figure 8 show examples of annotations along with the original images, but they are provided separately, in different files. Labeling of the plant organs in the images was done manually, using the software GIMP. Only areas in which there was confidence the organ corresponded to the correct plant were marked, which entailed that if there was uncertainty or occlusion, the area was not marked. Leaves, green fruits, and mature fruits were colored differently, in green, yellow and red, respectively. It should be noted that even as a lot of care was taken in the annotation process, often with two persons responsible for each image, it cannot be guaranteed that annotation was perfect, especially for green areas. In total, 618 different pictures of tomato plants were annotated (Table 4).

Table 4. Number of pictures annotated in each growth cycle.

Growth cycle	Monitoring pictures	Calibration pictures
Cycle 01	170	53
Cycle 02	167	44
Cycle 03	148	36

Most images include references of size. For fixed cameras, when the reference was obstructed by the plant, reference values from previous pictures can be used. Table 5 presents the references and their colors. For digitized leaves, the color refers to the reference itself, while for photos, to the color of the annotation. Area was used as a way of checking the conversion.

Table 5. Reference size information.

Source	Growth cycle	Color of annotation or of reference	Type of reference (line/rectangle)	Value of the reference
Scan	1	Black	Rectangle	Length: 5.7 cm; Area: 50.73 cm^2
Scan	2, 3	Pink	Rectangle	Length: 5 cm; Area: 19 cm^2
Photo – fixed camera – above view	1	Magenta	Line	Length: 31 cm
Photo – fixed camera – above view	2, 3	Cyan	Rectangle	Length: 29.7 cm; Area: 623.7 cm^2
Photo – fixed camera – above view	2, 3	Magenta	Rectangle	Length: 3,0 cm; Area: 9 cm^2
Photo – fixed camera – lateral view	1	Magenta	Line	Length: 27 cm
Photo – fixed camera – lateral view	2, 3	Cyan	Rectangle	Length: 29.7 cm; Area: 623.7 cm^2
Photo – smartphone – above view	1	Red	Line	Length: 4 cm
Photo – smartphone – above view	2, 3	Cyan	Rectangle	Length: 29.7 cm; Area: 623.7 cm^2
Photo – smartphone – above view	2, 3	Magenta	Rectangle	Length: 3,0 cm; Area: 9 cm^2
Photo – smartphone – lateral view	1	Magenta	Rectangle	Length: 8.9 cm; Area: 50.73 cm^2
Photo – smartphone – lateral view	2, 3	Cyan	Rectangle	Length: 29.7 cm; Area: 623.7 cm^2

Data is provided as raw as possible and the data repository includes the scripts used to process the observations to obtain time-series without missing data, for environmental observations, and of quantified areas in images, for pictures and masks. Scripts include comments to aid in processing the data accounting for particularities in collection. In case of updates in processing, changes will be available at the Zenodo versioned repository [4].

Ethics statements

The authors confirm: that the manuscript conform to the reporting standards, that raw data is public, that the work is original and all sources have been cited, that the content presented in this manuscript has not been submitted for publication in other journals, that authorship is limited to those who have made a significant contribution to the conception, design, execution, or interpretation of the reported study, that no confidential information is used or disclosed, that fundamental errors, if detected, will be notified to the journal editor or publisher, that there are no competing interests, and that no human, animal or social network data was used.

•

Authors’ contributions

Monique Oliveira: Conceptualization, Methodology, Software, Investigation, Data Curation, Writing - Original Draft; **Rafaella Amaro**: Investigation, Writing - Review & Editing; **Henrique Pescarini**: Investigation; Writing - Review & Editing; **Luiz Rodrigues**: Writing - Review & Editing, Supervision, Funding acquisition

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and by grant #2018/12050-6, São Paulo Research Foundation (FAPESP). We thank Dr. Rafael Ribeiro for the suggestions regarding the methods used and Dr. Antonio Camargo for the support in the last growth cycle.

Declaration of interests

Please **tick** the appropriate statement below and declare any financial interests/personal relationships which may affect your work in the box below.

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Please declare any financial interests/personal relationships which may be considered as potential competing interests here.

References

[1] M.P.G. de Oliveira, R.P. Amaro, H.B. Pescarini, L.H.A. Rodrigues, Tomato growth in production-like setting, (2021). https://doi.org/10.25824/redu/EP4NGO.

[2] M.A.R. Alvarenga, Tomate: produção em campo, casa de vegetação e hidroponia, 2nd ed., Editora Universitária de Lavras, Lavras, 2013.

[3] R.C. de M. Pires, P.R. Furlani, R.V. Ribeiro, D. Bodine Junior, E. Sakai, A.L. Lourencao, A. Torre Neto, Irrigation frequency and substrate volume effects in the growth and yield of tomato plants under greenhouse conditions, Sci. Agric. 68 (2011) 400–405. https://doi.org/10.1590/S0103-90162011000400002.

[4] M. Oliveira, Leveraging high frequency data for improving crop growth estimates, (2023). https://doi.org/10.5281/zenodo.7632419.