Smart Agricultural Technology 1 (2021) 100011

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.elsevier.com/locate/atech

Studies of evolutionary algorithms for the reduced Tomgro model calibration for modelling tomato yields

Liyun Gong[a,*], Miao Yu[a], Shouyong Jiang[a], Vassilis Cutsuridis[a], Stefanos Kollias[a], Simon Pearson[b]

[a] School of Computer Science, University of Lincoln, United Kingdom
[b] Lincoln Institute for Agri-Food Technology, University of Lincoln, United Kingdom

ARTICLE INFO

Keywords:
Reduced Tomgro model
Evolutionary algorithms
Model calibration
Yield prediction

ABSTRACT

The reduced Tomgro model is one of the popular biophysical models, which can reflect the actual growth process and model the yields of tomato-based on environmental parameters in a greenhouse. It is commonly integrated with the greenhouse environmental control system for optimally controlling environmental parameters to maximize the tomato growth/yields under acceptable energy consumption. In this work, we compare three mainstream evolutionary algorithms (genetic algorithm (GA), particle swarm optimization (PSO), and differential evolutionary (DE)) for calibrating the reduced Tomgro model, to model the tomato mature fruit dry matter (DM) weights. Different evolutionary algorithms have been applied to calibrate 14 key parameters of the reduced Tomgro model. And the performance of the calibrated Tomgro models based on different evolutionary algorithms has been evaluated based on three datasets obtained from a real tomato grower, with each dataset containing greenhouse environmental parameters (e.g., carbon dioxide concentration, temperature, photosynthetically active radiation (PAR)) and tomato yield information at a particular greenhouse for one year. Multiple metrics (root mean square errors (RMSEs), relative root mean square errors (r-RSMEs), and mean average errors (MAEs)) between actual DM weights and model-simulated ones for all three datasets, are used to validate the performance of calibrated reduced Tomgro model.

1. Introduction

Nowadays, greenhouses are preferred by many tomato growers. Compared with the field growing, growing tomatoes in a greenhouse can extent the tomato growing season, protect tomatoes against temperature and weather changes as well as provide tomatoes with a safe growing environment [15]. With the development of modern techniques, the environmental parameters (e.g., carbon dioxide concentration, temperature, photosynthetically active radiation (PAR)) in a greenhouse nowadays can be controlled to guarantee tomatoes to grow at the most appropriate environmental conditions [18,23]. The optimal control of greenhouse environments is vitally important for optimizing the cultivation management to guarantee the maximum tomato yield. And to determine the optimal greenhouse environmental parameters, most importantly, it is needed to accurately model the tomato growth/yield based on environmental parameters in a greenhouse

A variety of models have been developed for crop yield modelling, which is divided into two categories: data-driven model and explanatory biophysical model as in [14]. Data-driven ones ([1,2,13]) estimate crop yield by machine learning/deep learning models trained from a large collection of environmental parameters and historical yield data. While the explanatory model describes the relationship between greenhouse environmental factors (e.g., CO2, temperature, etc.) and crop morphological development based on ordinary differential equations (ODEs), for modelling the crop yield. Compared with the data-driven model, a biophysical model is more practical to reflect the actual growth process of crops, which is commonly integrated in the greenhouse environmental control system for optimizing the tomato yields [24].

Different biophysical models have been applied for the greenhouse crop yield modelling. The common-use HORTISIM [4] is a greenhouse crop system simulation model that can be used to predict crop yield in response to climate conditions inside a greenhouse and cultural operations, which is developed for providing effective strategies for greenhouse environmental control and management. Ritchie et al. [19] proposed the CERES model, which simulates a variety of factors (such as above-ground biomass, harvest index) related to crop yield in response to climate, soil, genotypes and management. Van Keulen [12] proposed the SUCROS model, which simulated the dry matter yield in various or-

* Corresponding author.
E-mail addresses: lgong@lincoln.ac.uk (L. Gong), myu@lincoln.ac.uk (M. Yu), sjiang@lincoln.ac.uk (S. Jiang), vcutsuridis@lincoln.ac.uk (V. Cutsuridis), skollias@lincoln.ac.uk (S. Kollias), SPearson@lincoln.ac.uk (S. Pearson).

https://doi.org/10.1016/j.atech.2021.100011
Received 23 April 2021; Received in revised form 24 June 2021; Accepted 10 September 2021

gans of wheat-based on solar radiation. The biophysical model proposed in [17] describe the effects of electrical conductivity, nitrogen, phosphorus, potassium, and light quality on dry matter yield and photosynthesis of greenhouse cucumbers. With respect to the biophysical models focusing on tomato growth modelling, the developed Tomgro model in [10] models the evolution of seven state variable vectors related to the tomato growth: numbers of leaves, numbers of main stem segments, numbers of fruits, dry weights of leaves plus petioles, dry weights of main stem segments, dry weights of fruits, and areas of leaves based on dynamically changing temperature, solar radiation and CO2 concentration inside a greenhouse. A more complex Tomsim biophysical model is proposed in [7], which contains multiple sub-modules developed for modelling different aspects (i.e., photosynthesis, dry matter production truss appearance rate, fruit growth period and dry matter partitioning, etc.) related to tomato growth. A model that describes the effects of greenhouse climate on tomato yield based on a series of ODEs was described and validated in [24]. This yield model was validated for four temperature regimes. Results demonstrated that the tomato yield was simulated accurately for both near-optimal and non-optimal temperature conditions in the Netherlands and southern Spain with varying light and CO2 concentrations. An integrated yield prediction model, which is an integration of Tomgro model [10] and Vanthoor model [24], is applied in [14] to predict the tomato yield in greenhouses based on controllable greenhouse environmental parameters.

Although different models [7,10,14,24] have been proposed for modelling tomato growth/yield, these models are complex with tremendous amounts of state variables and associated model parameters (for example, the Tomgro model in [10] has 69 state variables and that in [7] has 574 state variables), which makes it inconvenient for calibrating related models for accurately modelling tomato yield at a real greenhouse site. To ameliorate this, a reduced Tomgro model is proposed in [9]. Compared with models in [7,10,14,24], the reduced Tomgro model is relatively simple and has much fewer model parameters, which makes it more convenient to be calibrated for real-site tomato growth modelling as shown in [9]. And the reduced Tomgro model has been successfully applied to model the tomato yield on different real greenhouse sites, as reported in [3,20,25].

The genetic algorithm (GA) is usually applied as detailed in [25], to calibrate the reduced Tomgro model to make it applicable to model tomato yield on a particular site. The results obtained on a real tomato growing site has shown the fruit dry matter (DM) weights simulated by the calibrated reduced Tomgro model are close to the actual measurement data as shown in [25]. Although there are many other evolutionary algorithms that have been successfully applied in different engineering applications [21], however, no research of exploiting other optimization strategies for calibrating the reduced Tomgro model is investigated. To fill this gap, in this work we have evaluated different evolutionary algorithms for calibrating the reduced Tomgro model, to model the tomato fruit DM yield at a real greenhouse site. The chosen evolutionary algorithms include genetic algorithm (GA), particle swarm optimization algorithm (PSO) and differential evolutionary (DE) algorithm, which all have been successfully applied for a variety of optimization problems in different domains [8,11,16]. Each algorithm is applied to tune optimal model parameters in the reduced Tomgro model for modelling tomato fruit DM yields. The performance of each evolutionary algorithm for calibrating the reduced Tomgro model is evaluated on three datasets obtained from a tomato grower, with each dataset containing recorded carbon dioxide concentration, temperature, photosynthetically active radiation (PAR) and yield information in a greenhouse during one year's period. From statistical analysis of different error metrics (include root mean square errors (RMSEs), relative root mean square errors (rRMSEs) and mean absolute errors (MAEs)) obtained from the model simulated DM yields and ground-truth ones recorded in every dataset, it is shown that the reduced Tomgro model calibrated by the PSO algorithm achieves the most accurate result for modelling the fruit DM yields.

This paper is organized as below: Section 2 gives the descriptions of the reduced Tomgro model and different evolutionary algorithms: GA, DE and PSO. Detailed experimental studies on the performance of three evolutionary algorithms for calibrating the reduced Tomgro model are provided in Section 3. Dicussions of this research work are provided in Section 4.

2. Methods

2.1. Reduced Tomgro model

The reduced state-variable tomato model [9] is used for the simulation of tomato growth on the basis of three inputs that are measured inside a greenhouse environment: the photosynthetically active radiation in [mmol/m^2/s] or photosynthetically active radiation (PAR) (W/m^2), air temperature [°C] and CO2 concentration [ppm]. The reduced state-variable tomgro model, with a number of mainstem nodes (N), leaf area index (LAI), total plant weight(W), fruit weight(W_F), and mature fruit dry weight (W_M) as state variables, contains the same process equations for photosynthesis, respiration, and development as the comprehensive Tomgro/TomSim model in [7,10], but with new leaf area and dry matter growth relationships being developed. The dynamics of the total dry matter production and distribution in fruit and mature dry weight in the reduced state-variable tomgro model are dependent on photosynthesis and respiration processes, based on temperature, CO2 concentration and PAR radiation.

In specific, main process equations for the reduced state-variable Tomgro model are shown in the following equations:

$$\frac{dN}{dt} = N_m \cdot f_N(T) \tag{1}$$

$$\frac{d(LAI)}{dt} = \rho \cdot \delta \cdot \lambda(T_d) \frac{exp(\beta \cdot (N - N_b))}{1 + exp(\beta \cdot (N - N_b))} \cdot \frac{dN}{dt} \tag{2}$$

$$\frac{dW}{dt} = \frac{dW_F}{dt} + (V_{max} - p_1) \cdot \rho \cdot \frac{dN}{dt} \tag{3}$$

$$\frac{dW_F}{dt} = GR_{net} \cdot \alpha_F \cdot f(T_d) \cdot (1 - exp(-v(N - NFF))) \cdot g(T_{daytime}) \tag{4}$$

$$\frac{dW_M}{dt} = D_F(T_d) \cdot (W_F - W_M) \tag{5}$$

The rate of node development $\frac{dN}{dt}$ was modelled as a the maximum rate of node appearance rate per day N_m multiplied by a function that reduces vegetative development depending on non-optimal temperatures. Leaf area index is updated daily in the reduced model based on (2), which is dependent on the node number N and affected by the daily temperature T_d. W in (3) indicates the total above-ground dry weight. From Eq. (3), we can see that the growth rate of W is calculated as a weighted sum of the fruit growth rate and node growth rate. The partitioning of above-ground growth to fruit each day begins at node position NFF and increases asymptotically to a maximum partitioning using the Eq. (4). While in the reduced Tomgro model, the average development rate of mature fruit DM weight $\frac{dW_M}{dt}$, used to move the fruit from green to mature stages is given in (5), from which we can see that it depends on the daily average temperature T_d, fruit weight W_f and mature fruit weight W_M. More detailed descriptions of relevant equations and parameters can be found in [9].

2.2. Evolutionary algorithms

The calibration of the reduced Tomgro model can be regarded as an optimization problem, that is, finding the optimal set of model parameters to model the tomato yield to be as approximate as the actual one

[25]. In this work, we have compared three mainstream evolutionary algorithms (GA,DE and PSO) for reduced Tomgro model calibrations. Algorithm descriptions are shown in the next sections.

2.2.1. Genetic algorithm

Genetic Algorithm (GA) [6] is an adaptive heuristic search based evolutionary algorithm, which is based on the ideas of natural selection and genetics and commonly used to generate solutions for optimization problems. It simulates the process of natural selection which means those species who can adapt to changes in their environment are able to survive and reproduce and go to the next generation.

To exploit GA to solve an optimization problem requires a genetic representation of the solution domain of the problem as well as a fitness function for evaluating a solution. Once the genetic representation and the fitness function are defined, the GA proceeds to initialize a population of individuals(solutions) and then to improve it to obtain optimal solutions through repetitive application of the following operators:

Selection Operator: This operator gives preference to the individuals with good fitness function scores and allows them to pass to the successive generations.

Crossover Operator: This operator 'mates' between individuals. Two individuals are selected using the selection operator and crossover sites are chosen randomly. Then the genes at these crossover sites are exchanged thus creating a completely new individual (offspring).

Mutation Operator: This operator inserts random genes in offspring to maintain the diversity in the population to avoid premature convergence.

The aforementioned three operators are applied repeatedly until a termination condition has been reached, for example, a fixed number of generations reached or the highest-ranking solution's fitness is reaching or has reached a plateau such that successive iterations no longer produce better results. The procedure of the GA is described as in Algorithm 1. More details of the genetic algorithms can be found in [6].

Algorithm 1 Genetic algorithm procedure.

1) Randomly initialize populations p
2) Determine fitness of population
repeat
  a) Select individuals from population
  b) Crossover and generate new population
  c) Perform mutation on new population
  d) Calculate fitness for new population
until convergence

2.2.2. Differential evolution algorithm

Differential evolution algorithm [22] is a stochastic, population-based optimization algorithm for solving a nonlinear optimization problem. Similar to the generic algorithm, a fitness function and an initial population of individuals(solutions) are both defined for using DE to solve an optimization problem. However, it does not require the genetic representation of the solutions.

In specific, for exploiting the DE algorithm to solve an optimization problem, initially a population of N candidate solutions $[X_1^0, \ldots, X_N^0]$ is defined, where N represents the population size and 0 represents that it is the initial 0-th generation. Mutation, recombination and selection operations are performed to update candidate solutions until termination condition is satisfied. Details of different operations are described as follows:

Mutation: Based on a particular candidate solution X_i^g at the gth generation, randomly select three other individuals $(X_{r1}^g, X_{r2}^g, X_{r3}^g)$ and construct a donor vector V_i^{g+1} by using the following equation:

$$V_i^{g+1} = X_{r1}^g + F(X_{r2}^g - X_{r3}^g) \tag{6}$$

where F is a number between 0 and 1.

Recombination: A trail vector U_i^{g+1} is develop from X_i^g and the donor vector V_i^{g+1} by:

$$U_{i,j}^{g+1} = \begin{cases} V_{i,j}^{g+1} & rand() \le C_p \ \ or \ \ i = I_{rand} \\ X_{i,j}^g & rand() > C_p \ \ or \ \ i \ne I_{rand} \end{cases} \tag{7}$$

where $U_{i,j}^{g+1}$, $V_{i,j}^{g+1}$ and $X_{i,j}^g$ mean the jth element of U_i^{g+1}, V_i^{g+1} and X_i^g. I_{rand} is a integer random number between [1,D], where D is the vector dimension. C_p is the recommendation probability.

Selection: the target vector is compared with the trial vector and the one with the lower fitness function value is select to replace for the next generation, as in the following equation:

$$X_i^{g+1} = \begin{cases} U_i^{g+1} & f(U_i^{g+1}) < f(X_i^g) \\ X_i^g & otherwise \end{cases} \tag{8}$$

where $f(\cdot)$ represents the fitness function.

2.2.3. Particle swarm optimization algorithm

Particle swarm optimization (PSO) [5] is a population-based stochastic optimization algorithm motivated by the intelligent collective behaviour of some animals such as flocks of birds or schools of fish. In specific, initially N solution candidates (also called particles) $[X_1^0, \ldots, X_N^0]$ and their associated velocities $[V_1^0, \ldots, V_N^0]$ are generated. All particles have a fitness value, which can be calculated using a fitness function associated with a particular optimization problem.

Particle values are updated by the following equations, until the optimal solution is found:

$$\begin{aligned} V_i^{g+1} &= V_i^g + C_1 * rand() * (PB_i - X_i^g) + C_2 * rand() * (GB^g - X_i^g) \\ X_i^{g+1} &= X_i^g + V_i^{g+1} \end{aligned} \tag{9}$$

where V_i^g and X_i^g represent the ith velocity and particle in the gth generation. w is the inertia weight and C_1, C_2 represent the learning factors. $rand()$ generates a random number in [0,1]. PB_i and GB^g represent personal best performance for the ith particle and the best performance of the group for the gth generation respectively.

3. Results

The evaluations of different evolutionary algorithms for calibrating the reduced Tomgro model are presented in this Section. Detailed experimental analysis is given out in the following sub-sections.

3.1. Datasets descriptions

Three datasets are collected from a tomato growing site in Newcastle, UK, which contain recorded CO2 concentrations (mmp), temperatures (^oC),photosynthetically active radiation (PAR) (W/m^2) and tomato DM yield information in different greenhouses during different time periods. The details of these datasets are described in the following table:

As an illustration, the daily recorded CO2 concentration, temperature and PAR for all three datasets are shown in Fig. 1. Besides, the descriptive statistics analysis on environmental parameters for different datasets is summarized in Table 2. We can observe that for Dataset 2, the descriptive statistics (min, max median, and mean values) of recorded CO2 concentration values are comparatively lower than those in the other two datasets. While the descriptive statistics of other environmental parameters for these three datasets are almost consistent.

Tomato DM yield (g/m^2) in different greenhouses is measured weekly. Measured weekly DM yields associated with all three datasets are shown in Fig. 2.

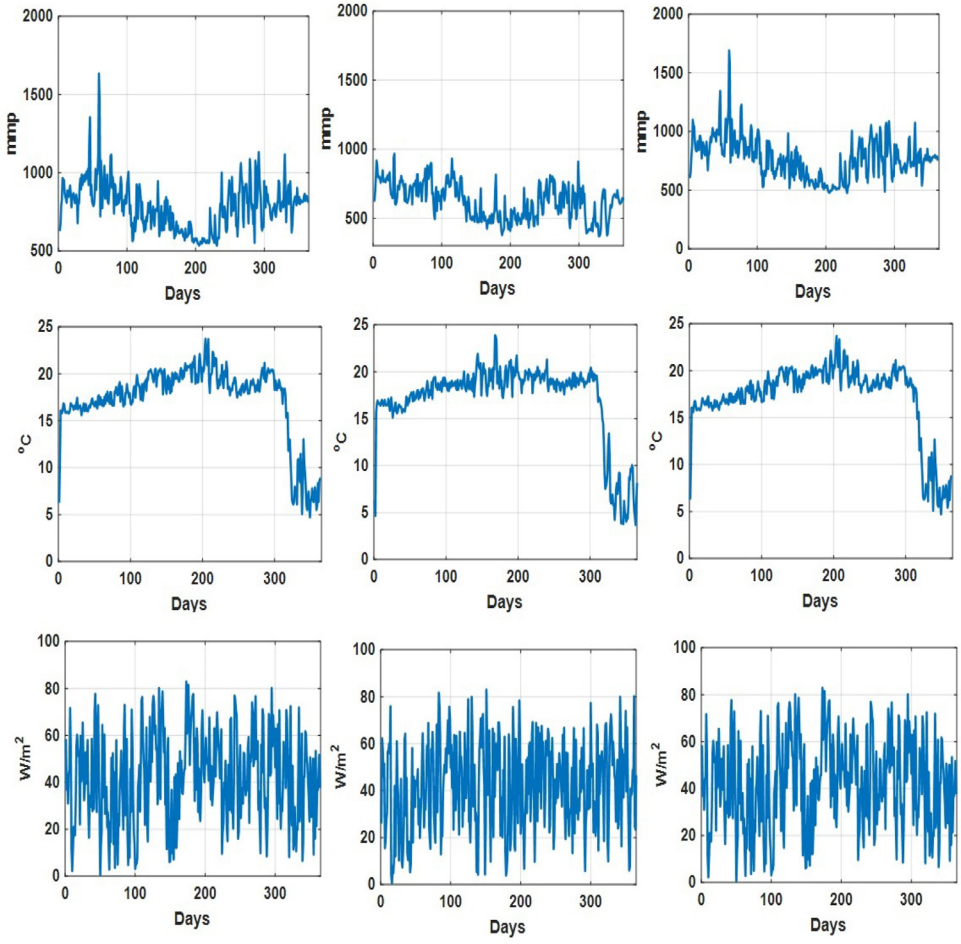

Fig. 1. Daily recorded CO2 concentration (mmp), temperature (^oC) and PAR(W/m^2) associated with dataset 1 (left column), dataset 2 (middle column) and dataset 3 (right column).

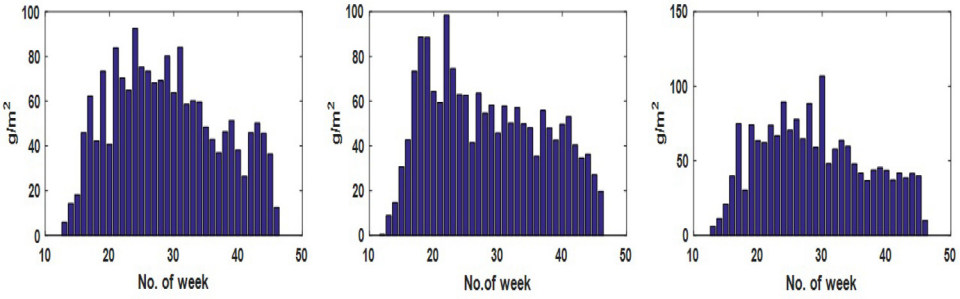

Fig. 2. Recorded weekly tomato DM yields in dataset 1 (left), dataset 2 (middle) and dataset 3 (right).

Table 1
Datasets descriptions.

	Dataset1	Dataset2	Dataset3
Location	Greenhouse 1	Greenhouse 2	Greenhouse 2
Time period	2018	2017	2018
Information included	yield information (g/m^2) CO2 concentration (mmp) temperature (^oC) PAR(W/m^2)		

3.2. Algorithms evaluations

Based on recorded environmental parameters and fruit DM yield information in every dataset, we've exploited different evolutionary algorithms for calibrating the reduced Tomgro model for modelling tomato fruit DM yield. The 14 key parameters of the reduced Tomgro model that needed to be tuned/calibrated are shown in Table 3. The detailed explanations on the physical meanings of different parameters can be found in [9].

Three evolutionary algorithms (GA, PSO, DE) are exploited and compared, for tuning the optimal set of model parameters (the ones in Table 3). For a fair comparison, both the initial populations of model parameters settings and the iteration number for three algorithms are set the same. In specific 100 initial solution candidates of model parameters are chosen randomly among the parameter ranges as in Table 1, for all three algorithms. The iteration number for all three algorithms is set to be 100. Note, we've also tried larger candidate and iteration numbers (such as 150 and 200), but found no obvious performance improvement for reduced Tomgro model calibration. All chosen algorithm parameters

Table 2
Descriptive statistics of greenhouse environmental parameters associated with different datasets.

		Dataset1	Dataset2	Dataset3
CO2 (mmp)	Min	535.97	370.94	478.05
	Max	1634.10	967.40	1691.43
	Median	793.95	629.97	769.79
	Mean	785.95	624.19	770.37
	Standard deviation	152.52	129.58	175.61
Temperature ($°C$)	Min	4.73	3.68	4.72
	Max	23.73	23.89	23.69
	Median	18.30	18.46	18.31
	Mean	17.25	17.01	17.18
	Standard deviation	3.97	4.25	3.94
PAR (W/m^2)	Min	0.59	0.58	0.59
	Max	82.91	83.02	82.91
	Median	42.81	43.41	42.81
	Mean	42.17	42.19	42.17
	Standard deviation	19.37	18.92	19.37

are summarized in Table 4. The root mean square error (RMSE) between the ground truth recorded DM yields in a dataset and ones estimated by the reduced Tomgro model, is taken as the fitness function for all three algorithms and shown below.

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}\|y_i^{real} - \hat{y}_i^{model}\|^2} \tag{10}$$

where y_i^{real} and y_i^{model} represent the ith recorded ground-truth DM yield and the corresponding estimated one by the calibrated reduced Tomgro model respectively.

The evolutions of RMSE fitness function values with respect to iteration numbers, for all three evolutionary algorithms applied to calibrate the reduced Tomgro model based on all three datasets, are shown in Fig. 3. From Fig. 3, we can see that for all scenarios, as the iteration number increases, the minimum fitness value of solution candidates (represented by the blue line in the figure) all decrease and converge to minimum values, which indicates model parameters solutions minimizing the fitness function can be successfully obtained after certain iteration numbers of all the three evolutionary algorithms.

After obtaining model parameter solutions based on evolutionary algorithms, the calibrated reduced Tomgro model is then used to model the fruit DM yield. Firstly, we evaluate the performance of the calibrated model for simulating the accumulated DM yields. Fig. 4 shows the comparisons between the accumulated ground-truth DM yields recorded in all three datasets and simulated ones by the calibrated reduced Tomgro model. We can observe the simulated accumulated DM yields match well with the ground-truth ones as shown in the figure.

Table 4
Chosen values for algorithm parameters.

GA	population size	100
	max. iteration	100
	mutation probability	0.001
DE	population size	100
	max. iteration	100
	F value in (2)	0.001
	crossover probability $C_p in (2)$	0.001
PSO	population size	100
	max. iteration	100
	w in (5)	0.8
	c1,c2 in (5)	0.5,0.5

Table 5
Performance comparisons on accumulated fruit DM yield estimations using reduced Tomgro model calibrated by different evolutionary algorithms.

		Avg.RMSE	Avg.r-RMSE	Avg.MAE
Dataset 1	GA	0.057	0.06	0.046
	DE	0.078	0.08	0.066
	PSO	0.043	0.05	0.032
Dataset 2	GA	0.052	0.06	0.043
	DE	0.041	0.05	0.033
	PSO	0.037	0.03	0.031
Dataset 3	GA	0.082	0.09	0.073
	DE	0.075	0.08	0.069
	PSO	0.068	0.07	0.059

A statistical study is made for quantitatively comparing the performance of the reduced Tomgro models calibrated by different evolutionary algorithms for modelling/estimating the accumulated DM yields. Each evolutionary algorithm is applied for calibrating the reduced Tomgro model multiple times. And for each time, a set of metrics (including RMSE, relative-RMSE and mean absolute error (MAE)) between simulated accumulated DM yields by calibrated models and ground truth recorded ones is calculated. The averages of obtained RMSEs, relative-RMSEs and MAEs are calculated and summarized inTable 5. We can observe that the reduced Tomgro model calibrated by the PSO algorithm can always achieve the minimum values of different metrics for all three datasets, which indicates that calibrated model by the PSO algorithm achieves the most accurate result for modelling the accumulated tomato fruit DM yield.

Moreover, we've also made comparisons on reduced Tomgro models calibrated by different evolutionary algorithms, for modelling the weekly DM weights increases. An illustration of the comparison of the ground-truth weekly DM yields recorded in different datasets and the simulated ones using reduced Tomgro models calibrated by different evolutionary algorithms is shown in Fig. 5. Similar to the accumulative

Table 3
Key parameters for the reduced Tomgro model.

Parameter	Description	Range of estimate	Unit
N_m	Max. rate of nodes	[0.35,0.4]	$node\ d^{-1}$
N_b	Param. in expolinear eq.	[14,16]	$node$
δ	Max. leaf area expansion	[0.05,0.08]	$m^2\ node^{-1}$
β	Param. in eq. expolinear	[0.45,0.55]	$node^{-1}$
$V_{\max}$	Maximum increase in vegetative tissue d.w. growth per node	[8,10]	$g[d.w.]node^{-1}$
τ	CO2 coefficiency	[0.08,0.12]	$\mu mol\ m^2\ s^{-1}$
T_{crit}	Critic temperature	[19,21]	^oC
υ	Transition from vegetative development to fruit development	[0.8,1]	$node^{-1}$
K	Development time from first fruit to fist ripe fruit	[0.8,1]	$node$
m	Light transmission coefficient	[0.01,0.015]	$dimensionless$
N_{FF}	Nodes per plant when first fruit appears	[16,18]	$node$
α_F	Maximum partitioning of new growth to fruit	[0.8,1]	$[fraction]d^{-1}$
E	Growth efficiency	[0.9,1.2]	$g[d.w.]g^{-1}[CH2O]$
D	Conversion coefficient of CO2 to CH2O	[4,6]	$gm^{-2}h^{-1}$

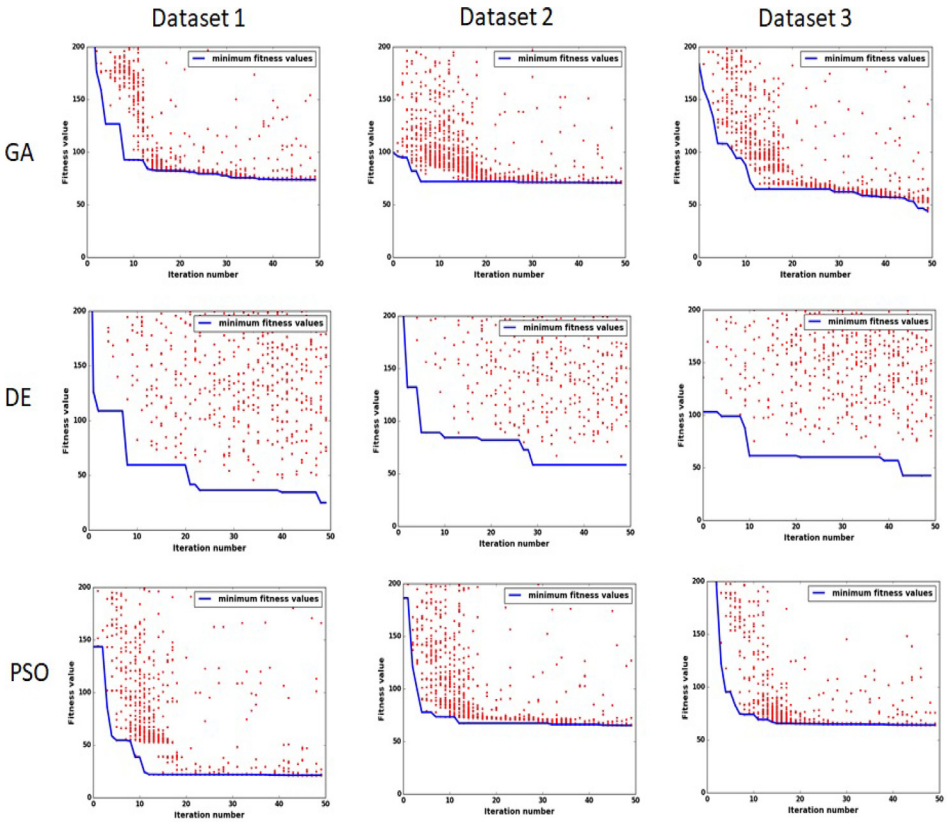

Fig. 3. The evolution of RMSE fitness function values with respect to iteration numbers for different evolutionary algorithms.

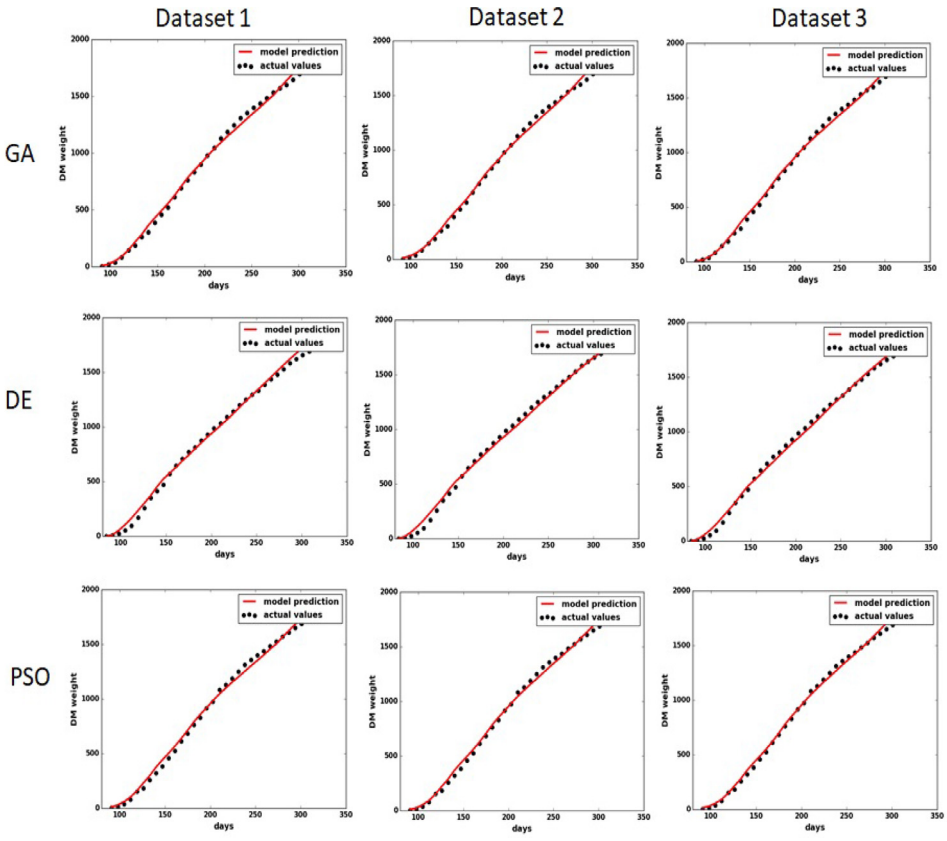

Fig. 4. The comparisons of accumulated recorded DM weights (g/m^2) in all three datasets and the ones simulated by the reduced Tomgro model calibrated by different evolutionary algorithms.

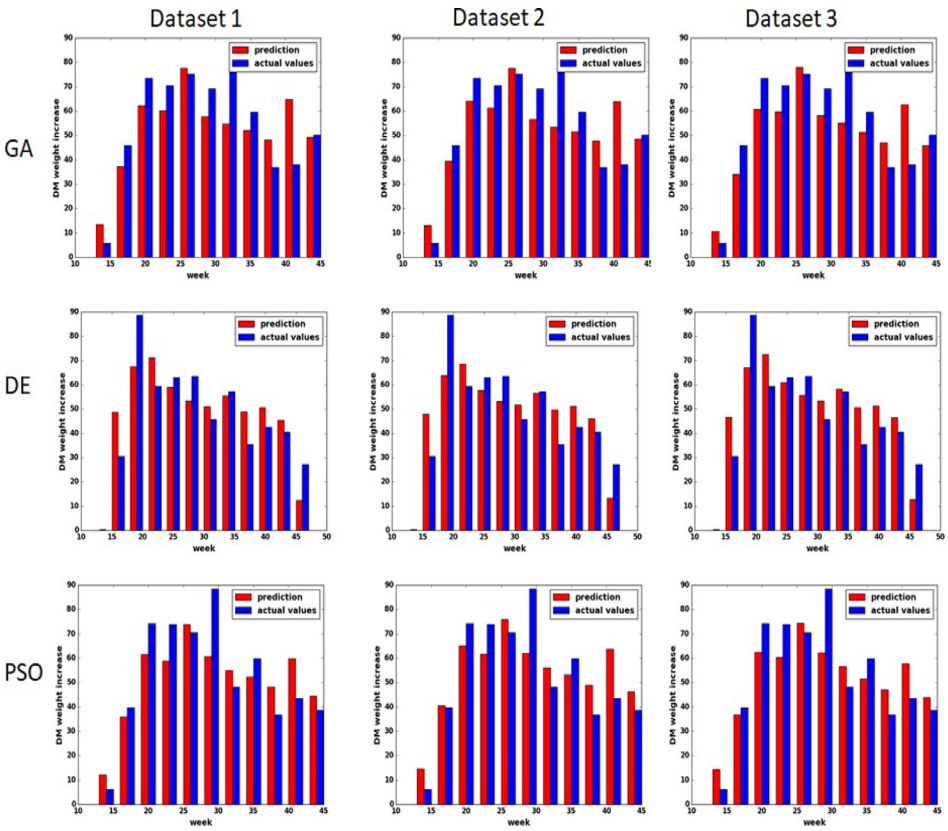

Fig. 5. The comparisons of recorded weekly DM weight growth (g/m^2) in all three datasets and the ones simulated by the reduced Tomgro model calibrated by different evolutionary algorithms.

Table 6
Performance comparisons on weekly fruit DM yield estimations using reduced Tomgro model calibrated by different evolutionary algorithms.

		Avg.RMSE	Avg.r-RMSE	Avg.MAE
	GA	0.015	0.19	0.011
	DE	0.015	0.20	0.011
Dataset 1	PSO	0.014	0.19	0.010
	GA	0.014	0.18	0.011
	DE	0.014	0.18	0.011
Dataset 2	PSO	0.014	0.18	0.010
	GA	0.016	0.21	0.012
	DE	0.017	0.23	0.014
Dataset 3	PSO	0.014	0.17	0.011

DM yield scenario, a quantitative evaluation is also made and the results are summarized in Table 6. From this table, we can see that compared with the accumulative DM yield modelling scenario smaller RMSEs are obtained, however, the r-RMSEs are larger indicating higher relative errors (compared with Table 5). And the model calibrated by the PSO algorithm still achieves the best performance for modelling the weekly DM weight increases, with the smallest RMSEs, r-RMSEs and MAEs for all three datasets.

4. Discussion

In this work, we have evaluated three evolutionary algorithms (GA,PSO and DE) to calibrate the reduced Tomgro model, for modelling the tomato mature fruit DM yield. From experimental analysis on multiple datasets obtained from a tomato grower, it is shown that all three algorithms can successfully calibrate the reduced Tomgro model for tomato yield modelling. Especailly the PSO algorithm achieves the most accurate performance with smallest mean values of RMSEs, r-RMSEs and MAEs, for calibrating the reduced Tomgro mode for simulating both the accumulative and week-by-week fruit DM yields. Although this work provides preliminary results of comparing different evolutionary algorithms for calibrating the reduced Tomgro model, a more objective evaluation of evolutionary algorithms for model calibration is expected to be made on more datasets collected from tomato growers in different locations.

With respect to future works, firstly more advanced biophysical models and optimization algorithms will be investigated. Besides, not only limited to model/estimate fruit DM yields, the modelling of different other factors (such as leaf area index, above-ground biomass) related to the growth of tomato and other crops will be investigated. Besides, both supervised and unsupervised machine learning models will also be investigated for modelling crop growths.

Declaration of Competing Interest

Authors confirm that this manuscript has not been published elsewhere and is not under consideration by another journal. The authors have no conflicts of interests to declare.

Acknowledgment

This research was supported as part of SMARTGREEN, an Interreg project supported by the North Sea Programme of the European Regional Development Fund of the European Union.

References

[1] B. Alhnait, S. Kollias, G. Leontidis, J. Shouyong, B. Schamp, S. Pearson, An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth, Inf. Sci. (560) (2021) 35–50.
[2] Alhnait, B., Pearson, S., Leontidis, G., & Kollias, S. Using deep learning to predict plant growth and yield in greenhouse environments. arXiv:1907.00624
[3] A. Arias, F. Rodriguez, M. Berenguel, E. Heuvelink, Calibration and validation of complex and simplified tomato growth models for control purposes in the southeast of spain, Acta Horticult. (2004).

[4] S. Cohen, H. Gijzen, The implementation of software engineering concepts in the greenhouse crop model hortisim1, Sci. Horticult. (1998) 431–440.
[5] M. Elbes, S. Alzubi, T. Kanan, A. Al-Fuqaha, B. Hawashin, A survey on particle swarm optimization with emphasis on engineering and network applications, Evol. Intell. (2019) 113–129.
[6] C. García-Martínez, F. Rodriguez, M. Lozano, Genetic algorithms, Handb. Heurist. (2018) 431–464.
[7] E. Heuvelink, Evaluation of a dynamic simulation model for tomato crop growth and development, Ann. Bot. 83 (1999) 413–422.
[8] M. Iwana, R. Akmeliawatib, T. Faisala, H. Al-Assadi, Performance comparison of differential evolution and particle swarm optimization in constrained optimization, Procedia Eng. 41 (2012) 1323–1328.
[9] J. J.Jones, A. Kenig, C. Vallejos, Reduced state-variable tomato growth model, Trans. ASAE 42 (1999) 255–265.
[10] J. Jones, E. Dayan, L. Allen, H. VanKeulen, H. Challa, A dynamic tomato growth and yield model (tomgro), Trans. ASAE 34 (1998) 663–672.
[11] V. Kachitvichyanukul, Comparison of three evolutionary algorithms:ga, pso, and de, Ind. Eng. Manag. Syst. 11 (3) (2012) 215–223.
[12] H. Keulen, J. Wolf, L. Fresco, L. Stroosnijder, J. Bouma, Modelling of agricultural production: weather, soils and crops, Agric. Ecosyst. Environ. 30 (1986) 142–143.
[13] S. Khaki, L. Wang, Crop yield prediction using deep neural networks, Front. Plant Sci. 10 (621) (2019).
[14] D. Lin, R. Wei, L. Xu, An integrated yield prediction model for greenhouse tomato, Agronomy 9 (873) (2019) 1–18.
[15] S. Naik, Cultivation of Tomato Production, Processing and Marketing, Agromisa, 2005.
[16] T. Nakane, N. Bold, H. Sun, X. Lu, T. Akashi, C. Zhang, Application of evolutionary and swarm optimization in computer vision: a literature survey, IPSJ Trans. Comput. Vis. Appl. 12 (3) (2020) 1–34.
[17] J. Ni, H. Mao, Dynamic simulation of leaf area and dry matter production of greenhouse cucumber under different electrical conductivity, Trans. Chin. Soc. Agric. Eng. (27) (2011) 105–109.
[18] P. Ponce, A. Molina, P. Cepeda, E. Lugo, B. MacCleery, Greenhouse Design and Control, Publishing House: CRC Press, 2014.
[19] J. Ritchie, D. Nesmith, Temperature and crop development, Sci. Hortic. 74 (1991) 341–342.
[20] R. Shamshiri, D. Ahmad, A. Zakaria, H.M. W. Ismail, M. Yamin, Evaluation of the reduced state-variable tomgro mode using boundary data, in: ASABE Annual International Meeting, 2016.
[21] A. Slowik, H. Kwasnicka, Evolutionary algorithms and their applications to engineering problems, Neural Comput. Appl. 32 (2020) 12363–12379.
[22] P. Suganthan, Differential evolution algorithm: recent advances, in: Proceddings of International Conference on Theory and Practice of Natural Computing, 2012, pp. 30–46.
[23] B. Vanthoor, C. Stanghellini, E.V. Henten, P.D. Visser, A methodology for model-based greenhouse design: part 1, description and validation of a tomato yield model, Biosyst. Eng. (110) (2011a) 363–377.
[24] B. Vanthoor, P. Visser, C. Stanghellini, E. Henten, A methodology for model-based greenhouse design: part 2, description and validation of a tomato yield model, Biosyst. Eng. 110 (4) (2011b) 378–395.
[25] M. Vazquez-Cruz, R. Guaman-Cruz, I. Lopez-Cruz, O. Cornejo-Perez, I. Torres–Pacheco, R. Guevara-Gonzalez, Global sensitivity analysis by means of efast and sobol methods and calibration of reduced state-variable tomgro model using genetic algorithms, Ann. Phys. 100 (2014) 1–12.