



HAL
open science

Predicting tomato water consumption in a hydroponic greenhouse: contribution of light interception models

Konstantinos Florakis, Samis Trevezas, Véronique Letort

► To cite this version:

Konstantinos Florakis, Samis Trevezas, Véronique Letort. Predicting tomato water consumption in a hydroponic greenhouse: contribution of light interception models. *Frontiers in Plant Science*, 2023, 14, pp.1264915. 10.3389/fpls.2023.1264915 . hal-04573477

HAL Id: hal-04573477

<https://centralesupelec.hal.science/hal-04573477v1>

Submitted on 13 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Predicting tomato water consumption in a hydroponic greenhouse: contribution of light interception models

Konstantinos Florakis^{1,2*}, Samis Trevezas¹, Veronique Letort²

¹National and Kapodistrian University of Athens, Department of Mathematics, Athens, Greece.

²Université Paris-Saclay, CentraleSupélec, Mathématiques et Informatique pour la Complexité et les Systèmes, Gif-sur-Yvette, France

Correspondence*:

Konstantinos Florakis

konstantinos.florakis@centralesupelec.fr

2 ABSTRACT

3 In recent years, hydroponic greenhouse cultivation has gained increasing popularity: the
4 combination of hydroponics' highly efficient use of resources with a controlled environment and
5 an extended growing season provided by greenhouses allows for optimized, year-round plant
6 growth. In this direction, precise and effective irrigation management is critical for achieving
7 optimal crop yield while ensuring an economical use of water resources. This study explores
8 techniques for explaining and predicting daily water consumption by utilizing only easily readily
9 available meteorological data and the progressively growing records of the water consumption
10 dataset. In situations where the dataset is limited in size, the conventional purely data-based
11 approaches that rely on statistically benchmarking time series models tend to be too uncertain.
12 Therefore, the objective of this study is to explore the potential contribution of crop models'
13 main concepts in constructing more robust models, even when plant measurements are not
14 available. Two strategies were developed for this purpose. The first strategy utilized the Greenlab
15 model, employing reference parameter values from previously published papers and re-estimating,
16 for identifiability reasons, only a limited number of parameters. The second strategy adopted
17 key principles from crop growth models to propose a novel modeling approach, which involved
18 deriving a Stochastic Segmentation of input Energy (SSiE) potentially absorbed by the elementary
19 photosynthetically active parts of the plant. Several model versions were proposed and adjusted
20 using the maximum likelihood method. We present a proof-of-concept of our methodology applied
21 to the ekstasis Tomato, with one recorded time series of daily water uptake. This method provides
22 an estimate of the plant's dynamic pattern of light interception, which can then be applied for
23 prediction of water consumption. The results indicate that the SSiE models could become valuable
24 tools for extracting crop information efficiently from routine greenhouse measurements with further
25 development and testing. This, in turn, could aid in achieving more precise irrigation management.

26 **Keywords:** GreenLab model, Greenhouse, MLE, Stochastic Segmentation of input Energy, Water consumption, Beer-Lambert law, light
27 interception

1 INTRODUCTION

28 Recent years have witnessed considerable research interest in the topic of water uptake by plants, particularly
29 in the context of hydroponic greenhouse cultivation. For tomato plants in particular, numerous studies have
30 stressed the importance of water uptake in comparing different cultivation methodologies. For instance,
31 (Reina-Sánchez et al., 2005) explored the effects of salinity concentration on water uptake, and (Biswas
32 et al., 2016) evaluated the impact of differing drip irrigation techniques. Moreover, investigations have
33 been carried out to assess the influence of nitrogen supply on growth, yield, and Water Use Efficiency
34 (WUE), incorporating biomass measurements either in field conditions (Cheng et al., 2021) or within
35 a hydroponic greenhouse environment (Martínez-Ruiz et al., 2019). In Sigrimis et al. (2001), a distinct
36 approach has been proposed, which suggests a methodology predicated upon measurements taken post-
37 irrigation, accompanied by online re-estimation of the parameters of their model. While the study reported
38 a high degree of accuracy for the proposed methodology, it is crucial to highlight its significant dependence
39 on measurements taken after each individual irrigation event, requiring sophisticated instruments. This
40 continuous monitoring may challenge the average producer, who may lack access to such advanced tools.
41 Furthermore, the solution appears to be largely engineering-based, neglecting the biological representation
42 of the plant's **development**. Adhering to a common practice of daily measurements, the challenge in our
43 setting consists in making the best use of limited information from data that concerns only water uptake and
44 easily accessible meteorological data/parameters. This framework resembles the typical evaluation scheme
45 that an average producer might employ to assess the productivity of their cultivation (Resh, 2022). However,
46 despite the potential benefits for average producers, the proposed simplified experimental protocol may
47 face serious limitations if data are not sufficiently informative due to measurement or even modeling errors
48 caused by oversimplified assumptions. These limitations will be further discussed in the sequel.

49 Plant water uptake is directly related to many greenhouse functions such as electric power usage (fertilizer
50 mixer, climatic regulating facilities, etc.), fertilizer consumption, and yield production and quality (Resh,
51 2022). Predicting water consumption in a given day could help the average producer regulate these costs,
52 prevent excess-deficit irrigation, and increase production. Another aspect of the problem is **water waste**:
53 **In an extensive study spanning 165 countries, the Food and Agriculture Organization of the United**
54 **Nations (FAO) estimated the total requirements and measurements of total withdrawals per country, thus**
55 **documenting a 56% irrigation efficiency only (Food and of the United Nations , FAO).**

56 Water consumption prediction can primarily be accomplished through two distinct methods. The first
57 approach is purely statistical, relying entirely on analyzing data series (Sigrimis et al., 2001). The second
58 approach employs process-based models or Functional-Structural Plant Models (FSPMs) (Sievänen et al.,
59 2014), which, despite requiring detailed and potentially costly longitudinal plant measurements, are
60 invaluable for their ability to convey information about underlying physiological processes and potential
61 interactions. **The GreenLab model (Yan et al., 2004) is such an FSPM, combining both functional and**
62 **structural description of metabolic processes with phytomer-level structures (De Reffye et al., 2021), and**
63 **integrating the effects of water dynamics on plant growth Wang et al. (2012), thus allowing optimization of**
64 **water supply Wu et al. (2005).** However, in many cases of professional practices where access to detailed
65 plant measurements is unavailable, neither of these approaches directly applies. Our objective, therefore,
66 is to find the appropriate level of complexity for a model that remains as mechanistic as possible. The
67 necessity of maintaining a mechanistic orientation stems from a few key reasons: the interpretability
68 of certain parameters allows using a priori biological knowledge and, crucially, our data's limited and
69 uncertain nature. This uncertainty makes a purely statistical approach infeasible, underscoring the need
70 for a method that captures the underlying physiological processes to a suitable extent. **To this end, we**

71 developed a new set of aggregated model versions, which inherit from the GreenLab principles but differ in
72 the representation of the Light Interception Ratio (LIR) of the plant.

73 In GreenLab, as in most process-based models, light interception is classically represented through
74 the Beer–Lambert–Bouguer law, hereafter named as Beer-Lambert (BL), assuming a simple exponential
75 relationship to describe light attenuation within a homogeneous canopy (Monsi & Saeki 1953, cited in
76 English translation in Hirose (2005)). It relies on a simple one-dimensional turbid medium model, which
77 raises several limitations regarding its consistency with experimental data (Ponce de León and Bailey,
78 2019), its stability in relation to other environmental variables (Valladares et al., 2012), and the vertical
79 variation of leaf photosynthetic parameters in the canopy (Sarlikioti et al., 2011), as well as its theoretical
80 foundations (Kostinski, 2002). In particular, this equation assumes that foliage is randomly dispersed, a
81 hypothesis that, depending on the species architecture, can lead to over-estimation of light interception if
82 the foliage is clumped or, on the contrary, under-estimation if the plant plasticity allows optimizing foliage
83 distribution for an enhanced light interception (e.g., for beech tree (Schröter et al., 2012)). In order to
84 overcome these problems, other approaches have been proposed recently, notably that of (Shabanov and
85 Gastellu-Etchegorry, 2018) and Casasanta and Garra (2018). In (Shabanov and Gastellu-Etchegorry, 2018),
86 the authors derive a stochastic formulation of the BL law, which accounts for heterogeneous canopies.
87 Their virtual experiments reveal that the traditional law is not universally applicable across different canopy
88 structures. In Casasanta and Garra (2018), the authors introduce two stochastic approaches to the problem.
89 The first one is based on a fractional Poisson process, resulting in a fractional BL law based on the
90 Mittag-Leffler function, also discussed later in the present study (see Eq. 16). The second is based on
91 weighted Poisson distributions, resulting in a Mittag-Leffler weighted BL law. In line with their work, we
92 also propose some possible generalizations of the BL law. In particular, by modeling appropriately the
93 probability of the event of interception, we derive a class of models for water consumption prediction.

94 Therefore, the main objective of our study was to explore to which extent some mechanistic principles
95 borrowed from physiologically based models could be incorporated into more statistical approaches for
96 predicting the water consumption of plants. To this end, our approach was (i) to analyze the identifiability
97 of some compartmental simplifications of the GreenLab model for tomato plants in the case where the
98 data set consists only of environmental variables, (ii) to derive from the GreenLab principles a new family
99 of models focusing on water consumption as the main state variable and differing by their assumptions
100 regarding the dynamics of the LIR, and (iii) to perform a very preliminary comparison of these models
101 using an experimental dataset of water consumption by tomato plants. Although mostly theoretical at this
102 stage, this work has some conceptual interest in presenting an original stochastic approach to derive a
103 new class of simple models and providing a procedural guideline to further confrontations to experimental
104 results.

2 MATERIAL AND METHODS

105 2.1 Data collection

106 Between May 10 and July 2, 2021, an extensive study was conducted in a hydroponic greenhouse near
107 Therma village, within the Nigrita-Serres region (40.91, 23.55), Greece, to examine the tomato plant's (cv.
108 ecstasis) water consumption patterns. A drip irrigation system was used to ensure precise irrigation for each
109 individual plant. Rockwool was used as a substrate growing medium, a product of basalt mainly composed
110 of Oxide of Calcium (CaO) with small percentages of Iron (Fe) and Aluminum (Al), in keeping with
111 common practices in the region. Plants' density is reported as 5 stems per m^2 (one stem per plant). Indoor

112 measurements were performed using an Efento Logger (Efento, 2020). Additionally, meteorological data
113 were collected using a Davis Vantage Pro 2 (Plus) weather station close to the greenhouse. A comprehensive
114 overview of the measured quantities, including Solar Radiation, Temperature, Humidity, and Air pressure,
115 averaged on a daily level, is presented in Table 1. Notably, the anemometer of the weather station was
116 positioned at a height of 6.5 meters above ground level to guarantee accurate wind speed measurements.

117 Initially, the plants were grown in a location separate from the greenhouse before being relocated to the
118 designated study area. The period spanning from seed planting to the onset of observations was noted to be
119 38 days. Upon arrival, the plants were observed to be in the initial stage of their first inflorescence, i.e., for
120 the majority of plants, the first truss has just been formed. The measurements were thus performed from
121 day 39 to day 92 of the plant growth, from 10/5/2021 to 2/7/2021, for a total of 54 days. Tomato plants are
122 usually grouped into stations, each combining a substrate (slab) with n plants and a pot for water collection.
123 The per average plant volume of water consumed by a station is referred to as *Water Consumption* (W_c),
124 given by

$$W_c(t) = \frac{V_{Irr}(t) - R_{off}(t)}{n}, \quad (1)$$

125 where $V_{Irr}(t)$ is the volume (L) of applied water at time t , and $R_{off}(t)$ the volume (L) of the corresponding
126 collected excess water which accumulates from the application time until the next morning during 12 hours.
127 In our study, $n = 3$, and the Water Consumption (W_c) corresponds to the dependent variable we try to
128 estimate and predict.

129 2.2 Brief presentation of the GreenLab model for tomato

130 The GreenLab model has been extensively explored in the literature, see, e.g., Dong et al. (2008) and
131 Zhang et al. (2009) for its application on tomato. Hereafter, we briefly recall its main principles, and the
132 interested reader is referred to Appendix 1.1 for a more comprehensive description of the model and to
133 Table 6 for the specific parameter values that we used in this work.

134 Plant development is assumed to be deterministically driven by the rules of a parameterized automaton
135 which determines the sequential appearance of phytomers (plant species-specific combinations of organs)
136 and their respective positions. The thermal time elapsing between the appearance of two successive
137 phytomers, assumed to be constant, serves as the discrete (simulation) time step and is referred to as a
138 Cycle of Development (CD). The organogenesis depends solely on thermal time, triggered above a base
139 temperature of $12^\circ C$ (Shamshiri et al., 2018).

140 The structure of a tomato plant can be delineated by four types of primary organs (excluding roots): *blade*
141 (b), *petiole* (p), *internode* (e), and *fruit* (f). Following the simplifications proposed by (Dong et al., 2008)
142 and (Zhang et al., 2009), we considered flowers and fruits as the same organ (i.e. the dynamics of biomass
143 allocation do not distinguish the transition from flower to fruit). Typically, in the cultivation of single-stem,
144 pruned, tomato varieties within greenhouses, seven to eleven phytomers devoid of flowers develop prior
145 to the emergence of the first inflorescence. In the present study, it is assumed that, following the first
146 eight phytomers without flowers, a truss appears at every third phytomer, producing three flowers that bud.
147 This specific assumption is consistent with empirical evidence from the present study and translates to an
148 average of three fruits per truss.

149 The integration of photosynthetic production is calculated using the Beer-Lambert (BL) law (Hirose,
150 2005; Monteith, 1977), which is analogous to the approach employed in most process-based models:

$$Q(t) = E(t) \cdot \text{RUE} \cdot S_p \cdot (1 - \exp\{-k S_L(t)/S_p\}), \quad (2)$$

151 where during $CD(t)$, $Q(t)$ corresponds to the newly synthesized (dried) biomass, $E(t)$ to the **Solar Radiation**,
 152 and **RUE** is the Radiation Use Efficiency (the vegetation efficiency of converting radiative energy into
 153 biochemical energy through photosynthesis). Moreover, S_p represents the projected surface potentially
 154 occupied by a **single** plant, while $S_L(t)$ stands for the plant's photosynthetically active leaf area, calculated
 155 as the sum of the total **photosynthetically active** biomass of the blades multiplied by the specific leaf area
 156 (SLA : coefficient converting a unit of produced biomass to leaf surface). The variable k corresponds to the
 157 extinction coefficient in the Beer-Lambert law, and it is set to 0.8 for the tomato crop, as in Zhang et al.
 158 (2009). For $t = 0$, the initial biomass of the seed is denoted by Q_0 .

159 At each CD, the available biomass is shared between all growing organs of the plants, regardless of their
 160 spatial position and proportionally to their current sink strength, **according to the so-called common pool**
 161 **assumption that was investigated for tomato by** (Heuvelink, 1995). The growth $\Delta q_o(u, t)$ of an organ of
 162 type o and chronological age u (days or CDs), **while the plant is in cycle $t \geq u$, can then be expressed as:**

$$\Delta q_o(u, t) = p_o \cdot f_o \left(\frac{u}{T_o} \right) \cdot \frac{Q(t-1)}{D(t)}, \quad (3)$$

163 where p_o is the relative sink strength of the organ of type o , $f_o(\cdot)$ its sink variation function related to the
 164 organ's biomass demand profile during its expansion and $D(t)$ the total demand in cycle t (see Eq. 23). As
 165 in (Yan et al., 2004), $f_o(\cdot)$ corresponds to a discretized beta law function with shape parameters a_o and
 166 b_o (see Eq. 26). For identifiability reasons, discussed in (Dong et al., 2008), the constraint $a_o + b_o = 5$ is
 167 imposed for tomato plants.

168 2.3 Link with Water Consumption

169 Howell and Musick (1985) demonstrated that transpiration and biomass production are proportional in
 170 their set of environmental conditions that encompasses our experimental conditions (Table 1) (Howell et al.,
 171 1984). In our greenhouse setting, evaporation is assumed to be negligible, so transpiration could, in turn, be
 172 considered proportional to water consumption (Food and Agriculture Organization of the United Nations,
 173 1998), thus rendering the latter linearly related to dry matter production. Disregarding evaporation is not
 174 a particularly far-fetched premise within the framework of hydroponic greenhouses. These greenhouses
 175 are designed to reduce evaporation to a minimum, utilizing substrates wrapped in white sacks that offer a
 176 minimal surface area for water to evaporate from (Resh, 2022).

177 Adding normally distributed homoskedastic errors, we obtain the following initial model:

$$W_c(t) = \mu_0 \cdot Q(t) + \varepsilon_t, \quad \text{where } \varepsilon_t \sim N(0, \sigma^2), \quad (4)$$

178 where μ_0 is a positive proportionality constant and σ^2 is a variance parameter **representing the experimental**
 179 **variability of the measurement process.**

180 As W_c measurements were conducted daily, but the GreenLab model runs on Cycles of development
 181 (CD), we need to map CDs on days. Elapsed days between two successive leaf developments (phyllochron)
 182 can vary from 1.5 (summer) to 3 (autumn) days according to the **genotype**, and the climatic conditions
 183 (Pivetta et al. (2007), Schmidt et al. (2017)). We assume that the phyllochron is stable and equal to 2 days,
 184 as we measured a mean value of 10 $^{\circ}Cd$ with a base temperature of 12 $^{\circ}C$. **To aggregate the two separate**

185 measurements into one CD, a weighted average is utilized with a weight proportional to the fraction of the
186 Solar Radiation of each day.

187 2.4 Identifiability issues and compartmental simplification of the GreenLab model

188 In our realistic setting, where no plant data are available, estimating the parameters of the complete
189 Greenlab model is unfeasible since identifiability problems typically arise: it means that different sets of
190 parameter values generate the same simulated dynamics, for a specified set of output variables. Thus, in
191 our case, plants with different characteristics could have the same dynamics of water consumption.

192 Adopting a general dimensionality reduction strategy for non-identifiability issues—outlined in (Hastie
193 et al., 2009)—we analyzed a simplified version of the model. We trade precision in representing the
194 biological model for enhanced identifiability within the parameter space. In this version, we combined all
195 the biomass of petioles (p), internodes (e), and fruits (f) into a single representative referred to as *body*.

196 Parameters requiring estimation thus comprise:

$$\theta = (a_b, b_b, p_{body}, a_{body}, b_{body}, S_p, \text{RUE}, SLA, \mu_0, \sigma, Q_0) \quad (5)$$

197 We will refer to this specific parametrization as *comp1*.

198 To explore the identifiability of parameters we performed simulation experiments under realistic scenarios.
199 It consists of generating virtual observations from a realistic set of parameter values and investigating
200 which parameters can be accurately estimated: non-identifiability is revealed when the estimated values are
201 different from the ones used for simulating the observations. In that case, the corresponding parameters
202 should be set as a constant and removed from the list of parameters to estimate, in order to reduce the
203 dimension of the parameter vector until reaching an identifiable subset.

204 For the sake of simplicity, we present in Section 3.1 the results from two characteristic cases only, which
205 correspond to the *comp1* model. In the first one, we fix SLA , the specific leaf area, and Q_0 , the initial
206 biomass of the seed, quantities that can typically be measured. Parameters S_p and RUE are also fixed, since
207 we incorporated the μ_0 parameter in the model 4, a simplification justified by the compensation effect
208 between those parameters. In the second case, additionally to the previously mentioned parameters, we fix
209 P_{body} , the sink strength of the “body” compartment. By initializing 5000 randomly selected starting points,
210 we recorded the solutions that maximize the likelihood function of the model, with a tolerance of $< 10^{-3}$
211 to account for numerical approximations. The maximization of the function was performed via a BFGS
212 (Broyden–Fletcher–Goldfarb–Shanno) quasi-Newton algorithm for Bound Constrained Optimization (Byrd
213 et al., 1995).

214 2.5 Two model versions for water consumption time series based on the recurrence 215 equation of GreenLab

216 As shown in Letort et al. (2009), the GreenLab model can be synthesized into a single recurrence equation
217 that, for the sake of simplicity, we chose here to formulate as:

$$Q(t) = E(t) \cdot \text{RUE} \cdot S_p \left(1 - \exp \left\{ -\frac{k \cdot SLA}{S_p} \sum_{n=0}^{t-1} r(n) Q(n) \right\} \right),$$

218 where $r(n)$ represents the proportion of green to the totally produced biomass $Q(n)$, a quantity that can be
 219 calculated as a function of the model parameters. Assuming proportionality (with constant μ_0) between
 220 biomass production and water consumption and no leaf senescence, we obtain a general model form for
 221 water consumption:

$$W_c(t) = \theta_1 \cdot E(t) \cdot \left(1 - \exp \left\{ -\theta_2 \sum_{n=0}^{t-1} r(n) W_c(n) \right\} \right), \quad (6)$$

222 where $\theta_1 = \text{RUE} \cdot S_p \cdot \mu_0$ and $\theta_2 = \frac{k \cdot SLA}{S_p \cdot \mu_0}$ are estimated, while the other parameters which appear implicitly
 223 in the coefficients $r(n)$ are fixed at the values found in Dong et al. (2008) (see Table 6). This model will be
 224 referred to as *GreenLab exp*.

To account for the obviously existing differences between the tomato plants in Dong et al. (2008) and those available in this study, we propose a modified parametric version of the coefficients as follows:

$$r(t) = \frac{t^a}{I(a)}, \quad \text{where} \quad I(a) = \int_0^{t_{max}} t^a dt = \frac{t_{max}^{a+1}}{a+1}$$

225 corresponds to a normalization constant with respect to a , a parameter to estimate, and to the maximum
 226 time of observation t_{max} . This model will be referred to as *exp + rate*.

227 2.6 Stochastic models of light interception to predict water consumption

228 Building upon the prior discussion, we now focus on a novel aspect that broadens the model formulation.
 229 Here, we aim to represent biomass production at time t , as the cumulative byproduct of a composite
 230 stochastic experiment, which consists of many independent individual experiments, each one deciding
 231 whether elementary radiative inputs will be absorbed by the plant or not. We thus derive a family of models,
 232 which will be referred to as “Stochastic Segmentation of input Energy” models (SSiE).

233 2.6.1 Formulation of the water consumption series from a stochastic model of light interception

234 In this section, we discuss the intuition behind a probabilistic interpretation of biomass production, and
 235 we formalize this intuition with tools from theoretical probability. At each time t , a total radiative input
 236 $E(t)$ is channeled into the system per m^2 . We assume that this input is equally quantized into very small
 237 elementary quantities $\{E_i(t)\}_{i=1}^n$ in such a way that either they are completely absorbed by the plant and
 238 converted into biomass by the enlightened parts of the plant or they exit the system without affecting it.
 239 In this case, $E_i(t) = E(t)/n$ where n represents the number of “elementary” units. If no other specific
 240 details are known, one could assume that the individual events of absorption, say $A_i(t)$, are independent
 241 with identical probability of occurrence $p(t)$. With this interpretation and if $\mathbb{1}_{A_i(t)}$ stands for the indicator
 242 function of the corresponding event, each elementary radiative input $E_i(t)$ is associated with a random
 243 variable

$$Q_i(t) = \text{RUE} \cdot S_p \cdot E_i(t) \cdot \mathbb{1}_{A_i(t)}, \quad (7)$$

244 which records its produced biomass, either 0 if the event $A_i(t)$ is not realized or $\text{RUE} \cdot S_p \cdot E_i(t)$ if the
 245 event is realized, and thus it is totally transformed. The total biomass produced by the plant at time t can
 246 thus be expressed as follows:

$$Q^{(n)}(t) = \sum_{i=1}^n Q_i(t) = \text{RUE} \cdot S_p \cdot E(t) \cdot \frac{\sum_{i=1}^n \mathbb{1}_{A_i(t)}}{n}. \quad (8)$$

247 Clearly, the last factor of the above expression corresponds to the sample mean of independent and
 248 identically distributed random variables and in particular Bernoulli random variables with common
 249 probability $p(t)$. Intuitively, one should expect by the strong law of large numbers that the sample mean
 250 value should be very near to their common probability of absorption, that is, $p(t)$. These arguments give an
 251 intuitive interpretation of the fact that the following approximations should be plausible:

$$Q(t) \approx Q^{(n)}(t) \approx \text{RUE} \cdot S_p \cdot E(t) \cdot p(t). \tag{9}$$

252 However, despite the seemingly sound arguments underlying these approximations, a theoretical
 253 justification of their validity is more complex. An obvious theoretical caveat regarding the validity
 254 of these approximations is that we cannot conceptualize a countably infinite sequence of events of
 255 common probability that play the role of the elementary events of biomass absorption, or equivalently the
 256 total radiative input cannot be partitioned into a countably infinite number of positive parts potentially
 257 transformed into biomass. One possibility for justifying the above approximations would be to resort to
 258 an uncountable number of stochastic experiments. This approach involves more mathematical intricacies.
 259 For this reason, and since a rigorous justification of this part is not necessary for the rest of the paper, the
 260 interested reader is referred to Appendix 1.2 for more details.

261 The next step is to appropriately model the probability of absorption $p(t)$, which can classically be done
 262 through a parametric family of continuous distribution functions. For each time t , let $\{Z_u(t)\}_{u=0}^{E(t)}$ represent
 263 the Bernoulli experiments of absorption of the radiative input for all possible u ranging from 0 to $E(t)$. If
 264 we denote by $LIS(t)$ the Light Interception Surface at time t , then, assuming that the maximum available
 265 soil surface is S_p , one could construct a new family of random variables $\{U_u(t)\}_{u=0}^{E(t)}$ uniformly distributed
 266 on $[0, S_p]$ which concretize the above experiments. In particular, the interval $[0, S_p]$ is partitioned into two
 267 subintervals $[0, LIS(t)]$ and $(LIS(t), S_p]$. Then, the absorption events can be written as

$$A_u(t) := \{Z_u(t) = 1\} = \{U_u(t) \leq LIS(t)\}, \quad 0 \leq u \leq E(t). \tag{10}$$

In probability theory, such a family exists; loosely speaking, this reinterpretation of the absorption events corresponds to a collection of idealized experiments where an elementary radiative input enters into the system if it intersects with the green part of the plant. Now, notice that $p(t)$ corresponds exactly to the probability of the event given by (10) which is related to the Light Interception Surface $LIS(t)$ at time t . However, $LIS(t)$ is not directly observable, but only indirectly via the cumulated water consumption prior to time t , denoted by $SW_c(t^-)$ (itself proportional to the cumulated produced biomass). A novelty of this study consists in making a link between $LIS(t)$ and $SW_c(t^-)$ through an increasing (non-decreasing) function $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, that is, $LIS(t) = g(SW_c(t^-))$. By the above argument, Eq. (10) and the fact that $U_u(t) \sim \text{Unif}(0, S_p)$ we get that all the following equalities hold:

$$p(t) = \mathbb{P}(U_u(t) \leq LIS(t)) = \mathbb{P}(U_u(t) \leq g(SW_c(t^-))) = \frac{g(SW_c(t^-))}{S_p} = \frac{LIS(t)}{S_p} =: LIR(t),$$

268 where the last term stands for the Light Interception Ratio. Now, also notice that if $U \sim \text{Unif}(0, S_p)$ is a
 269 copy from the family $\{U_u(t)\}_{u=0}^{E(t)}$ and g is invertible, then the third term above can be rewritten as

$$LIR(t) = \mathbb{P}(g^{-1}(U) \leq SW_c(t^-)) = \mathbb{P}(X \leq SW_c(t^-)) = F_X(SW_c(t^-)), \tag{11}$$

270 where we set $X = g^{-1}(U)$. In fact, since g is assumed to be an increasing function, its inverse exists at
 271 least in a generalised form (generalised inverse) and the above equations still hold. The problem is then to
 272 define the relationship between $LIR(t)$ (or $LIS(t)$) and $SW_c(t^-)$ without having any information on the
 273 plant itself and in the next section we discuss several such possibilities.

274 2.6.2 Different options for the distribution of X

The determination of a mechanistic functional relationship between $LIR(t)$ and $SW_c(t^-)$ is unrealistic. Biologically speaking, the underlying processes are complex and involve, among others, the patterns of biomass allocation to blades and their arrangement in space. An approach to this objective is, however, feasible and a selected number of possible distribution families could be used to compete for their fitting quality and their predictive ability. By introducing additive errors as in Section 2.3, we can derive a model directly applicable to Water Consumption variable

$$W_c(t) \sim N\left(\theta_1 \cdot E(t) \cdot F_X(SW_c(t^-)), \sigma^2\right), \tag{12}$$

275 thereby eliminating the requirement for biomass as intermediary variable. Each model is determined by
 276 specifying F_X in one of the following parametric family of distributions.

Exponential distribution. The exponential distribution is one of the most fundamental suppositions that one can make when faced with an undetermined distribution, since it corresponds to the maximum entropy solution for a given expected value on the positive line (Jaynes, 1957). Besides, in our setting, it leads to a Beer-Lambert-like model. By (11) and the assumption of an exponential model we get:

$$LIR(t; k) = 1 - \exp\left(-k \cdot SW_c(t^-)\right), \quad t \geq 0. \tag{13}$$

277 **Gamma distribution**

The gamma distribution is a generalization of the exponential distribution. This provides a logical progression from our initial assumption of an exponential distribution. By (11) and the assumption of a gamma model, we get:

$$LIR(t; k, a_\gamma) = \int_0^{SW_c(t^-)} \frac{k^{a_\gamma}}{\Gamma(a_\gamma)} s^{a_\gamma-1} e^{-k \cdot s} ds, \quad t \geq 0. \tag{14}$$

278 **Mittag-Leffler distribution**

Mittag-Leffler introduced the function bearing his name in 1903 (Bateman, 1953). Different properties of the distribution generated by the Mittag-Leffler function were explored in Pillai (1990). The concept of generalizing the Beer-Lambert law with the use of the Mittag-Leffler function was proposed by Casasanta and Garra (2018). Following their work, we incorporate this generalization into our analysis, leading to the following LIR term:

$$LIR(t; k, a_{ML}) = 1 - E_{a_{ML}}\left(-\left(k \cdot SW_c(t^-)\right)^{a_{ML}}\right), \quad t \geq 0, \tag{15}$$

where $E_{a_{ML}}$ is the Mittag-Leffler function:

$$E_{a_{ML}}(x) = \sum_{j=0}^{\infty} \frac{x^j}{\Gamma(j \cdot a_{ML} + 1)}, \quad x \in \mathbb{R}, \tag{16}$$

279 with $a_{ML} \in (0, 1]$ the **tail parameter** and $k > 0$ the **rate parameter**. For $a_{ML} = 1$ the above formulation
 280 reduces to the exponential distribution with **rate parameter** k .

281 **Log-normal distribution**

282 The log-normal distribution is commonly employed to model growth rates. Our reasoning for
 283 incorporating this distribution in our analysis stems from the presumption that the elementary events
 284 $(A_i)_{i=1}^n$ are influenced by the incremental growth of smaller plant elements. This growth is contingent
 285 on their size. For the density function, we proceed by adopting the ensuing parametrization:

$$LIR(t; \mu_{\log}, \sigma_{\log}) = \int_0^{SW_c(t^-)} \frac{1}{s \cdot \sigma_{\log} \cdot \sqrt{2\pi}} \exp\left(-\frac{(\log(s) - \mu_{\log})^2}{2\sigma_{\log}^2}\right) ds, \quad t > 0. \quad (17)$$

286 **Pareto distribution**

287 The last distribution we explore is Pareto. Following (Van der Zande et al., 2010) (mainly the results
 288 depicted in Figures 2 and 3), we observe that the percentage of the biomass responsible for most of
 289 the energy interception follows a similar law to the Pareto 80/20 rule (Juran and De Feo, 2010). The
 290 formulation of the distribution function that we adopt is as follows:

$$LIR(t; \theta, \eta) = 1 - \left(\frac{\eta}{SW_c(t^-)}\right)^\theta, \quad SW_c(t^-) > \eta. \quad (18)$$

291 **2.7 Model comparison and prediction performance criteria**

292 To have a challenging baseline model to compete with, we first estimated a linear first-order autoregressive
 293 model with one exogenous variable, namely the **average solar radiation** received at day t , $E(t)$:

$$W_c(t) = b_0 + b_1 W_c(t-1) + b_2 E(t) + \varepsilon_t, \quad (19)$$

294 where the b_i coefficients are estimated via the **ordinary least squares method**.

295 In terms of forecasting, a sequential methodology is employed. From the original dataset, we initially
 296 extract the first 55% days (days 39 to 68) for training and predict the next day's water consumption (day
 297 69). Subsequently, we increase the size of the training set by one additional day at each step, continuing
 298 to predict the following day until we reach the end of the time series. The parameters are re-estimated at
 299 each step of the procedure, using a total of 1000 distinct starting points in our calculations, subsequently
 300 selecting the point with the highest likelihood value as the model's parameter for prediction. However, in
 301 the case of the *mlf* model, the number of initial points was reduced to 20, **to reduce the computational**
 302 **burden**. After the parameter estimation process, a model selection procedure was performed with the two
 303 most classical model selection criteria, namely the corrected Akaike Information Criterion (AICc) (Hurvich
 304 and Tsai, 1989) and the Bayesian Information Criterion (BIC).

305 The setting described above reflects real-world conditions as it emulates the practical scenario where we
 306 have a bunch of observations, and our objective is to forecast Water Consumption for the upcoming day.
 307 **Two different settings were considered for the inputted Solar Radiation (E): (i) assumed to be perfectly**
 308 **known (fixed covariate setting), or (ii) with an additive white noise factor associated with predicting solar**
 309 **radiation, where the standard deviation was set empirically at 20, a value which corresponds to a bound on**
 310 **values typically obtained with current prediction models (Tao et al., 2019).**

The predictive performance of the models was compared with the Root Mean Square Prediction Error (RMSPE):

$$\text{RMSPE}(\hat{y}) = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}, \quad (20)$$

311 where $y = (y_i)$ is the vector of observed values and $\hat{y} = (\hat{y}_i)$ the predicted ones. For the testing set,
 312 according to the previously described protocol we used $m = 24$ observations. The computer programs
 313 were developed in R (version 4.3.1.) and the packages *MittagLeffler* (Gill and Straka, 2018) and *tidyverse*
 314 (Wickham et al., 2019) were used for computations with the Mittag-Leffler distribution and other data
 315 manipulations and visualization respectively.

3 RESULTS

316 3.1 Identifiability analysis of the GreenLab model with compartmental simplification

317 When considering only water consumption data, a certain number of the GreenLab parameters are not
 318 identifiable. This is true even when the simplified and parsimonious *comp1* model is used which has fewer
 319 parameters than the complete one (Section 2.4). The boxplots in Figure 1 allow comparing the case where
 320 P_{body} is estimated (in addition to B_b , B_{body} , μ_0 , and σ) with the case where it is set at its reference value.
 321 Each point represents an estimated parameter value, and specific combinations of these points correspond
 322 to the estimated solutions of the maximization problem. Note that, for scaling purposes, P_{body} has been
 323 normalized by its maximum value. The plots reveal that many distinct solutions yield similar likelihood
 324 values. As can be seen by comparing the ranges of the estimated parameters (see Fig. 1, left and right),
 325 this identifiability issue diminishes as we set more parameters, but never disappear. Even with only four
 326 estimated parameters, we remark compensation effects between B_b and B_{body} , since the resulting estimates
 327 still vary significantly. However, the parameters μ_0 and σ are identifiable, at least locally, around the chosen
 328 reference values, a noteworthy result which enables the elaboration of the stochastic framework discussed
 329 in Section 2.6.

330 3.2 Estimation of the linear and SSiE models' parameters

331 The regression results for the linear autoregressive model 19 with Average Solar Radiation as an
 332 exogenous variable show that all parameters appear to be statistically significant at a 0.01 significance level
 333 (Table 2). The coefficient of determination R^2 and the adjusted one have similar values, of approximately
 334 0.88.

The estimated parameter values of each SSiE model (Table 3) and their relative performances (Table 4) according to the comparison criteria defined in 2.7 highlight a slight superiority of the *lognormal* and *pareto* models in terms of both the Bayesian Information Criterion (BIC) and the corrected Akaike Information Criterion (AICc). A straightforward application of Eq. 6 by estimating the green biomass by an already fitted model (Dong et al., 2008) does not appear to be highly promising, as it still results in higher values in these criteria. Similar behavior is present in the Beer-Lambert-like approach of the exponential distribution. A notable result is the estimation of $a_{ML} \simeq 0.5$ (see Table 3). For $a_{ML} = 0.5$ the Mittag-Leffler function

(16) reduces to (Haubold et al., 2011):

$$E_{1/2}(x) = e^{x^2} \left(1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds \right),$$

where $\frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds$, also known as the Gauss error function, is a quantity which expresses the probability of a typical Gaussian distribution to be found in the interval $[-x, x]$ for $x \geq 0$. In our case this translates to:

$$LIR(t) \simeq 1 - \exp(-k \cdot SW_c(t^-)) \cdot \mathbb{P}(|Z| > \sqrt{k \cdot SW_c(t^-)}), \quad t \geq 0,$$

335 where $Z \sim N(0, 1)$. Another noteworthy finding is related to the *pareto* model and specifically the parameter
 336 η which corresponds to the initial cumulative water consumption $SW_c(t)$ up to the first observation time.
 337 This parameter was estimated at 0.403 (see Table 3) and corresponds approximately to 400 ml over a span
 338 of 38 days.

339 Observing the temporal change of the estimated LIR with the different methods described in Section
 340 2.6.2 reveals that the pairs (*lognormal*, *GreenLab-exp*) and (*gamma*, *lognormal + rate*) exhibit similar
 341 trends (Figure 2). This similarity is even more visible when the LIR is normalized by its maximal value and
 342 displayed with respect to SW_c , as shown in the Supplementary Material (Appendix 4). As the optimization
 343 procedure revealed, there is a compensation effect between θ_1 and the LIR scaling, thus justifying the
 344 normalized representation in Appendix 1.3. However, the *pareto* and *mlf* methodologies demonstrate
 345 distinct trends that can be clearly differentiated from the others. The unique trend of the *pareto* methodology
 346 is also evident in Figure 3, where it manages to track the initial and final trends concurrently during the
 347 observation period, as opposed to the other methods, which are only capable of capturing either the
 348 beginning or the end trend, but not both simultaneously. Another notable result concerns the grouping of
 349 the best-performing models according to the BIC criterion (Fig. 4). The estimated LIR resulting from the
 350 best representative of these models is also shown in Figure 2.

351 3.3 Prediction results

352 The results of the predictive analysis revealed that the *pareto* and the LR models exhibited the best
 353 predictive performance both under known and unknown but predicted Solar Radiation, indicating their
 354 relative superiority within the context of our investigation (Table 5). However, the *lognormal* and the *mlf*
 355 models were slightly inferior and almost equivalent between them in both settings of Solar Radiation,
 356 followed by the *gamma* model. Surprisingly, it is crucial to acknowledge the underperformance of the
 357 *lognormal*, *GreenLab exp* and *lognormal + rate* models, which implement a methodology similar to
 358 the Beer-Lambert law. Compared to other models, these models' inferior performance underlines the
 359 importance of generalising the BL-law for optimising performance.

4 DISCUSSION

360 4.1 Different possible uses of crop models for predicting water consumption without 361 plant information

362 Our primary objective was to investigate methodologies for modeling and predicting Water Consumption
 363 in tomato plants by utilising concepts derived from the crop models but without any information on the
 364 crop. While ambitious, this objective is grounded in uncovering hidden patterns within the crop's behavior
 365 through the model's learning process, particularly patterns of light interception. This approach is inherently

366 interdisciplinary, combining methodologies from data science, statistics, and biology to address a complex
367 biological problem.

368 Initially, we focused on the well-researched GreenLab model (Yan et al. (2004), De Reffye et al. (2021)),
369 but the methodology could be considered generic and applied to other FSPMs (Sievänen et al., 2014) or
370 crop models of tomato (e.g. Marcelis et al. (2008)). Various strategies for using a mechanistic model in
371 such a setting can be considered, each with a unique potential role. The first strategy revolved around
372 employing GreenLab as a completely pre-fitted model, as done in Chew et al. (2014) for the case of
373 *Arabidopsis*. However, the issues here concerned uncertainties and discrepancies among various genotypes.
374 The second strategy would be to estimate all the parameters of GreenLab, but due to the incomplete dataset
375 and the current experimental protocol, this strategy proved infeasible as the model exhibited identifiability
376 problems, as also reported in Letort et al. (2012) for *Coffea* trees when only compartment data are available.
377 Such an estimation procedure would necessitate destructive plant data, at different growth stages, for
378 estimating the sink strengths and their variations Guo et al. (2006). The third strategy involved using
379 GreenLab as a partially pre-fitted model, estimating only a fraction of its parameters. This was done for the
380 *GreenLab exp* model, where only two parameters of the production equation were estimated. This strategy
381 also encompassed the use of GreenLab as a submodel, assuming a similar pattern for the globally allocated
382 biomass fraction to the leaves, as done for the *GreenLab rate*, or by combining some of its basic principles
383 in the proposed SSiE models (e.g. proportionality between biomass production and water consumption is
384 retained as an underlying hypothesis).

385 4.2 Summary of our main findings

386 The proposed models, employing *pareto*, *LR*, *mlf* or *Lognorm*, yielded comparable predictive outcomes
387 (RMSPE 0.2-0.23). In the context of our problem formulation, which involves one measurement of Water
388 Consumption per day and relies solely on climatic data, this RMSPE translates to an error of approximately
389 215ml per day. This level of accuracy can contribute to the sustainability of agricultural practices by
390 optimizing water usage. Importantly, the *pareto* and *mlf* models are feasible for application in a scheme of
391 one measurement per day. However, both of them have disadvantages. The *pareto* model presents some
392 identifiability issues among the μ_0 and θ parameters, which warrants further investigation. On the other
393 hand, the *mlf* is computationally heavy, a disadvantage that can be minimal in a scenario with only one
394 measurement and only one day to predict. Despite these challenges, the models remain viable choices
395 for real-world applications. Even though *lognormal + rate* and *gamma* models do not present equivalent
396 results as the aforementioned, the LIR estimated by these methods, approximately 80%, are similar to the
397 results reported in Wilson et al. (1992) and Ohashi et al. (2022). Measurements at 7 farms showed that
398 in the summer season, the light interception was on average 90%, with values varying between 86% and
399 96%” in Heuvelink et al. (2004), with reported densities of 2.3 and 3.4 stems per m^2 , in contrast to our
400 case, where the reported density is 5 stems per m^2 .

401 This work can be considered as a methodological proposition for determining the LIR profile with only
402 a subset of the variables routinely measured by professional growers in a hydroponic setting, i.e., Water
403 deficit volume, Solar Radiation. Interestingly, the profiles of LIR that we obtain in Figure 2 are consistent
404 with those reported in the literature ((Duursma et al., 2012), Ohashi et al. (2022)). Selecting the models
405 with the best predictive performances seems a reasonable strategy. Nonetheless, this approach warrants
406 further empirical validation. Future research could focus on quantifying the diffusion of light in relation to
407 distinct plant attributes and may include virtual experiments (as in (Duursma et al., 2012)).

408 4.3 Modeling light interception and its relation with plant growth

409 The amount of energy a crop captures, crucial for modeling crop growth and yield, is largely determined
410 by canopy light interception (Higashide and Heuvelink, 2009). There is, however, currently no consensus
411 on how light interception should be modeled: Liu et al. (2021) reviewed the canopy light interception
412 models of 26 crop models of wheat and reported that the uncertainty in simulated wheat growth and final
413 grain yield due to the different light models could be as high as 45%. The light interception modules form
414 a continuum of approaches that range from simple (empirically or theoretically grounded) relationships
415 between some characteristics of the photosynthetically active parts of the plant (usually LAI, e.g., Sarlikioti
416 et al. (2011)) and the way they intercept light, to complex scene illumination algorithm, incorporating a
417 precise 3D geometric representation of the plant (e.g. in Schipper et al. (2023)). Our work contributes to
418 that line of research, proposing a new cost-effective methodology to assess the time course of the LIR
419 through its dependence on the water consumption profile. Our results further address the discussion of the
420 need for generalizations or alternatives to Beer-Lambert law. Shabanov and Gastellu-Etcheberry (2018)
421 and Casasanta and Garra (2018) have already proposed theoretical suggestions in this direction, and we
422 believe our work presents a practical application of these theories.

423 Various variables have been employed in the literature to characterize light interception (e.g. STAR (light
424 interception per leaf area) in Oker-Blom and Smolander (1988), Duursma et al. (2012), FIPAR (Fraction
425 of PAR intercepted by the photosynthetically active radiation elements of canopies) in Liu et al. (2021)).
426 In this context, we utilize the Light Interception Ratio (LIR) that characterizes light interception per soil
427 unit, a term we have intentionally left loosely defined. In our usage, this is primarily because LIR is a more
428 empirically determined global variable rather than one rigorously derived from mechanistic principles.
429 Nevertheless, we anticipate that it may still offer some interpretive value within the scope of our study. Our
430 models cannot disentangle the different factors influencing light interception (leaf density, orientation, etc.)
431 but provide a global representation of light interception at the plant scale, which is easy to obtain using
432 routine measurements, and can assist in simple predictions of Water Consumption.

433 4.4 Limitations of the work

434 Our work presents important limitations that must be acknowledged. First, our modeling approach relies
435 on strong physiological simplifications: e.g. neglecting soil evaporation and respiration of existing organs,
436 constant radiation use efficiency, constant SLA, no influence of external environmental conditions except
437 radiation and applied water volume, proportionality between light intercepted and photosynthesis (a more
438 refined model here would have been to consider Farquhar's photosynthesis model, for instance Farquhar
439 et al. (1980)), proportionality between water consumption and biomass production. Regarding this last
440 assumption, the ratio of biomass to transpiration (Water Use Efficiency (WUE)) is known to vary with
441 weather, genotypes, and practices (Blankenagel et al. (2018), Bhaskara et al. (2022)). Therefore, using a
442 constant value is likely to be valid only in a limited range of environmental conditions that would have to
443 be determined using a more extensive experimental dataset (Lanoue et al., 2017)

444 All these simplifications were required with respect to our objectives and our context of using only
445 routinely recorded variables. They can, however, be considered applicable when describing the average
446 growth of plants in standard conditions, and most of them are also laid in other models (Ma et al.
447 (2022), Winn et al. (2023)).

448 An additional underlying assumption that deserves to be highlighted is that the g function is time-
449 independent. In reality, g aggregates the effects of blade spatial arrangement, which determines the

450 probability of a radiation ray being intercepted, the fraction of biomass allocated to the blades and the
451 senescence of the leaves. This fraction decreases with time, especially due to the progressive appearance of
452 fruits, whose demand competes with that of blades, a phenomenon that our SSiE models do not account for.
453 However, in our case, because the time of observation is at a very later stage than the initial planting, this
454 fraction is, in fact, nearly constant, taking values in the range (0.21-0.24), as simulated using GreenLab
455 (Zhang et al., 2009). This explains why the models *exp* and *GreenLab exp* behave similarly.

456 **In conclusion**, we must acknowledge the limitations of our data, which prevent us from drawing strong
457 conclusions from our results. Measuring and estimating the mean value of water consumption among only
458 three plants could potentially introduce some errors because of the variance within them. **Solar Radiation is**
459 **measured outside of the greenhouse, which introduces the need for simulating an unknown transmission**
460 **coefficient through the greenhouse: such coefficient is accounted for in the constant θ_1 in model 12.**
461 **Lastly, since we do not have access to light distribution measurements in our study, we cannot definitively**
462 **conclude on the validity of our models by comparing our simulation outputs to real measurements, nor**
463 **can we assess the stability of the values of the parameters of our models for different environmental**
464 **conditions.** Nevertheless, we believe that our work can be considered as a proof-of-concept for our
465 proposed methodology and that the SSiE model appears promising for modeling Water Consumption.

466 4.5 Perspectives

467 In light of this, our future research will aim to apply further and investigate the utility of the SSiE model
468 in predicting such quantities. The choice of distribution might be crop-dependent, and we aim to explore
469 this idea in the future **by acquiring data that would enable testing our models' assumptions regarding the**
470 **relationship between water consumption, crop architecture, and the different profiles of light distribution**
471 **within the canopy.**

472 From a methodological point of view, the current formulation is particularly adapted for Bayesian methods,
473 which will allow for an easy way to quantify uncertainty and use the Bayesian predictive distribution for
474 forecasting purposes. An online Bayesian method with sequential Monte-Carlo may be particularly relevant,
475 and MCMC methods could also be applied for more efficient estimation, as in (Logothetis et al., 2022).
476 The comparison of MCMC with sequential Monte-Carlo for MLE was done in Trevezas et al. (2014).

5 CONCLUSION

477 **In this study, we aimed to better understand plant water consumption, a subject of considerable importance**
478 **for greenhouse management. The widely-used GreenLab model was not identifiable in our setting, even**
479 **after compartmental simplifications, but it could be considered in other applications if at least partial**
480 **information on the plant could be collected. Using similar physiological assumptions but in a probabilistic**
481 **framework, we introduced the SSiE model as an alternative, directly applicable to water consumption, thus**
482 **avoiding the need for biomass production as an intermediary variable. Despite the limitations of our data,**
483 **the SSiE model provided some useful preliminary insights, particularly in the area of light interception over**
484 **time. While these findings are still at a mostly theoretical stage, our proof-of-concept on our experimental**
485 **dataset hints at the SSiE model's potential utility for water consumption and light interception analyses.**

486 **The practical implications of these initial findings could be noteworthy and extend toward other crops**
487 **and settings, offering a pathway to more efficient water usage in greenhouses.**

ACKNOWLEDGMENTS

488 The authors wish to thank Kostoudis Greenhouse company for the availability of data and measurement
489 gauges.

1 APPENDIX

490 1.1 Appendix of GreenLab model for tomato

491 The integration of photosynthetic production is calculated using the Beer-Lambert (BL) law (Monteith,
492 1977), which is analogous to the approach employed in most process-based models:

$$Q(t) = E(t) \cdot \text{RUE} \cdot S_p \cdot (1 - \exp\{-k S_L(t)/S_p\}), \quad (21)$$

493 where during $CD(t)$, $Q(t)$ corresponds to the newly synthesized (dried) biomass, $E(t)$ to the system's
494 energetic contribution (in our case, solar radiation), and **RUE** is the Light Use Efficiency (LUE) (the
495 vegetation efficiency of converting radiative energy into biochemical energy through photosynthesis).
496 Moreover, S_p represents the projected surface potentially occupied by a **single** plant, while $S_L(t)$ stands for
497 the plant's photosynthetically active leaf area. This is calculated as the sum of the total **photosynthetically**
498 **active** biomass of the blades multiplied by the specific leaf area (SLA). The variable k corresponds to the
499 extinction coefficient in the Beer-Lambert law, and it is set to 0.8 for the tomato crop, as per Zhang et al.
500 (2009). $Q(0) = Q_0$ is the initial biomass of the seed. In our case, because of the difference in planting and
501 first observation, the initial biomass will also be a parameter under estimation.

502 1.1.1 Dry matter allocation

503 The biomass ascribed to every organ, spread from the common pool, is set proportional to its sink strength
504 (Yan et al. (2004), De Reffye et al. (2021)). In the context of mechanistic models for simulating dry matter
505 partitioning, a common assimilate pool refers to a shared pool of assimilates from which various sink
506 organs of a plant derive their growth resources. This implies that the plant does not segregate into distinct
507 source-sink units, and thus, any resistance encountered during the transport of assimilates from source to
508 sink would not influence the distribution of dry matter (Heuvelink, 1995). Sink strength adjusts during
509 the period of organ expansion, following the same form of sink function for all organs of the same type
510 $o \in \{b, p, e, f\}$ in a cohort. A cohort is a set of organs of the same nature, created at the same CD by the
511 parallel functioning of meristems.

512 If T_o stands for the expansion duration of an organ of type (o) and t stands for its chronological age (days
513 or CDs), then the sink strength is modeled by the function:

$$P_o(t) = p_o \cdot f_o\left(\frac{t}{T_o}\right), \quad 0 \leq t \leq T_o, \quad (22)$$

514 where p_o is its relative sink strength (with respect to the blade's one), $f_o(\cdot)$ is the variation function of the
515 sink related to its development, chosen as a Beta density function.

516 The sum of the sink strength of all organs is the Demand $D(t)$ at a given time t :

$$D(t) = \sum_o \sum_{u=1}^t N_o(t - u + 1)P_o(u), \tag{23}$$

517 where $N_o(t - u + 1)$ is the total number of organs of type o at time t that appeared at time u and are,
 518 therefore, of age $t - u + 1$. The biomass growth of an organ o varies on the value of its sink and the
 519 ratio supply produced to the previous cycle $Q(t - 1)$ (2) by the current demand $D(t)$ (23). The growth of
 520 an organ of type o that manifests during cycle u , while the plant is in a subsequent cycle $t > u$, can be
 521 articulated as:

$$\Delta q_o(u, t) = P_o(t - u + 1) \frac{Q(t - 1)}{D(t)}, \tag{24}$$

522 and the weight of the organ that appeared in cycle u when the plant is at age t is then:

$$q_o(u, t) = \sum_{j=u}^t \Delta q_o(u, j). \tag{25}$$

523 1.1.2 Beta sink function

524 In the initial GreenLab model (Yan et al., 2004), the sink function is defined according to a discretized
 525 beta law function:

$$f_o \left(\frac{t}{T_o} \right) = \frac{1}{M} \left(\frac{t}{T_o} \right)^{a_o-1} \left(1 - \frac{t}{T_o} \right)^{b_o-1}, \quad 0 \leq t \leq T_o, \tag{26}$$

526 where the parameters a_o and b_o , verifying the constraint $a_o, b_o \geq 1$, drive the curve shape and M is a
 527 normalization constant usually modeled as the sum or the max over $1 \leq t \leq T_o$.

528 In this case, the parameters to estimate are:

$$\theta = (a_b, b_b, p_p, a_p, b_p, p_e, a_e, b_e, p_f, a_f, b_f, S_p, \text{RUE}, \text{SLA}, Q_0) \tag{27}$$

529 The absence of p_b arises from the standard practice of establishing $p_b = 1$ as a reference point (Dong
 530 et al. (2008), Zhang et al. (2009)).

531 A dimensionality reduction approach, discussed in (Dong et al., 2008), stabilizes the sum of the two
 532 parameters a_o, b_o . The value of 5 has been specifically chosen for tomato plants, as it has been observed to
 533 produce fine results. Under this assumption, the parameter's space dimension is reduced by 4.

534 1.2 Probabilistic Justification of the SSiE model

535 Let us now try to justify the rationale behind the discussion in Section 2.6. The radiative input $E(t)$ could
 536 be mapped to the interval $[0, E(t)]$ representing an uncountable number of points potentially available for
 537 biomass production. At each point u of the interval, one could attach a Bernoulli random variable, say
 538 $X_u(t)$, deciding whether the point u will enter the system or not. If it enters the system, then it brings
 539 an infinitesimal contribution to biomass production; otherwise, it is rejected and exits the system. One

540 could still keep the independence assumption and assume that there is a common probability $p(t)$ of the
 541 radiative points entering the system, but there is a price to pay. If we assume that the radiative input is a
 542 realization of the stochastic process $\{Z_u(t)\}_{u=0}^{E(t)}$, where the sample (observed) paths would be an interval
 543 of points consisting of 0's and 1's, then it can be proved with tools from probability theory that the resulting
 544 processes are not measurable.

545 To give an interpretation of this nonmeasurability concept, it roughly corresponds to the idea that it
 546 would be impossible to associate the usual notion of length to the set of points that entered the system
 547 and the set of points that exited the system in this ideally conceptualized experiment. Luckily enough,
 548 there is still a solution, and it gives a formal justification for our intuitive approximations. It resides in the
 549 disintegration theorem (Chang and Pollard, 1997), a result of measure and probability theory. In fact, this
 550 theorem gives very powerful tools and a more intuitive approach to the definition of conditional probability
 551 and conditional expectation than the one that is usually presented in standard probability textbooks. A
 552 formal description of this theorem and related conditions for its validity would be out of the scope of this
 553 paper, and we refer to Chang and Pollard (1997). However, we describe the basic ingredients and the result
 554 we need in our context.

555 Instead of selecting points from the interval $[0, E(t)]$, one could think that the same interval is actually
 556 a bundle of Bernoulli experiments, where each one of them is realised when the point u is “activated”.
 557 Formally, one needs a measure space which consists of the set $Y_t := [0, E(t)] \times \{0, 1\}$, an appropriate
 558 measure μ and a function $\pi : Y_t \rightarrow [0, E(t)]$ (usually the projection function) which disintegrates the
 559 measure μ into a family of measures $\{\mu_u\}_{0 \leq u \leq E(t)}$, such that for a measurable A

$$\mu_u(A) = \mu_u\left(A \cap (\{(u, 0), (u, 1)\})\right) \tag{28}$$

560 and induces the measure $\nu = \mu \circ \pi^{-1}$ on $[0, E(t)]$. In our case, the choices are rather simple. Each μ_u is
 561 “living” (has its support) on the fiber $\{u\} \times \{0, 1\}$ and behaves as a Bernoulli measure, while the induced
 562 measure ν should be the Lebesgue measure restricted on $[0, E(t)]$. In this way, the disintegration theorem
 563 justifies the following way of computing the measure of a measurable set A :

$$\mu(A) = \int_0^{E(t)} \mu_u(A) du, \tag{29}$$

564 where $\mu_u(A)$ is given by (28), and the integral should be understood in the Lebesgue sense. We are now
 565 ready to make the correspondence with the computation of the totally produced biomass at time t . Since the
 566 set $B = [0, E(t)] \times \{1\}$ corresponds to the set of all active points, in order to assess the totally absorbed
 567 radiative input, we just have to compute

$$\mu(B) = \int_0^{E(t)} \mu_u(B) du = E(t)p(t), \tag{30}$$

568 since $B \cap \{(u, 0), (u, 1)\} = \{(u, 1)\}$ and $\mu_u(\{(u, 1)\}) = \mathbb{P}(A_u(t)) = p(t)$. Multiplying by RUE S_p to
 569 transform into biomass, we get the expected approximation result given by (9). It is also interesting to
 570 notice that the constant probability $p(t)$ is actually playing the role of a constant flow (with respect to the
 571 incoming radiation) of biomass production. The discussion is continued in paragraph 2.6.1.

Table 1. Summary of all measured variables (units in parenthesis) in column 1. The most basic descriptive statistics (mean, sd, min, max) from a sample of $N = 54$ measurements are given in columns 2-5.

Variables	Mean	St. Dev.	Min	Max
Avg Solar Radiation (W/m^2)	286.944	45.844	114.590	340.310
Max Solar Radiation (W/m^2)	1,021.830	131.702	571	1,329
Avg Air pressure (hPa)	1,014.045	3.601	1,007.120	1,022.720
Avg Temperature ($^{\circ}C$)	22.043	2.642	17.060	28.630
Max Temperature ($^{\circ}C$)	29.306	3.086	23.100	37.800
Min Temperature ($^{\circ}C$)	15.707	3.403	9.100	22.000
Avg Humidity (%)	0.837	0.109	0.618	0.976
Max Humidity (%)	0.984	0.031	0.840	1.000
Min Humidity (%)	0.591	0.161	0.230	0.860
Water Consumption (L)	1.102	0.519	0.090	2.250

Table 2. Summary of the base linear regression model of Water Consumption vs the predictors given in the first column (units in parenthesis). The estimated coefficients (sd in parenthesis) are given in the second column. Asterisks denote the statistical significance according to Student's *t*-test.

	<i>Dependent variable:</i> $W_c(t)$ (L)
Avg Solar Radiation (W/m^2)	0.002 (0.001) ***
$W_c(t - 1)$ (L)	0.893 (0.051) ***
Constant	-0.480 (0.156) ***
Observations	53
R ²	0.884
Adjusted R ²	0.879
Residual Std. Error	0.178 (df = 50)
F Statistic	189.652*** (df = 2; 50)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3. Estimated parameters for the models described in (6) and (12) (see Section 2.6.2). The pair (θ_2, k) is aligned in the same column.

Version	θ_1	σ	k or θ_2	a	μ_{log}	σ_{log}	θ	η
lognorm	0.011	0.165	-	-	3.958	3.273	-	-
pareto	370.112	0.165	-	-	-	-	$3.02 \cdot 10^{-6}$	0.403
mlf	0.01	0.166	0.017	0.501	-	-	-	-
gamma	0.007	0.169	0.01	0.386	-	-	-	-
exp + rate	0.007	0.172	2.037	-0.834	-	-	-	-
GreenLab exp	0.005	0.208	0.559	-	-	-	-	-
exp	0.005	0.211	0.133	-	-	-	-	-

Table 4. Comparison of different distribution choices regarding the formulations in (6) and in (12). The columns refer successively to the method’s name, the estimated log-likelihood value (log_lik_val), the RMSE, the total number of estimated parameters, the BIC and the AICc criteria.

	Version	log_lik_val	RMSE	# param	BIC	AICc
1	lognorm	20.45	0.16	4	-25.02	-31.62
2	pareto	20.39	0.16	4	-24.9	-31.5
3	mlf	19.85	0.17	4	-23.82	-30.42
4	gamma	19.12	0.17	4	-22.36	-28.96
5	exp + rate	18.15	0.17	4	-20.42	-27.02
6	LR	17.7	0.17	3	-19.51	-26.57
7	GreenLab exp	8.14	0.21	3	-4.37	-9.45
8	exp	7.25	0.21	3	-2.59	-7.67

Table 5. Prediction summary among the different suggested methods discussed in (6) and in (12), under: (left) Solar Radiation assumed to be known (right) an additive normal $N(0, 20^2)$ noisy prediction setting. Methods are compared using the RMSPE. The ordering is performed under the setting with noise.

	Version	RMSPE	RMSPE with noise
1	pareto	0.194	0.202
2	LR	0.205	0.212
3	mlf	0.226	0.234
4	lognorm	0.217	0.237
5	gamma	0.234	0.254
6	GreenLab exp	0.282	0.29
7	exp	0.296	0.319
8	exp + rate	0.341	0.364

Table 6. Parameters of the GreenLab model for tomato and their values in our study.

Parameter	Comments	value
p_b	Blade relative sink strength	1
p_p	Petiole relative sink strength	1.09
p_e	Internode relative sink strength	0.93
p_f	Individual fruit relative sink strength	61.3
B_b	Blade sink variation parameter	0.43
B_p	Petiole sink variation parameter	0.45
B_e	Internode sink variation parameter	0.38
B_f	Fruit sink variation parameter	0.36
S_p	Projection surface (cm^2)	5047
k	Beer Lamber coefficient	0.8
RUE	Radiation Use Efficiency	0.05
T_b	maximum expansion time of blade (CDs)	10
T_e	maximum expansion time of internode (CDs)	8
T_p	maximum expansion time of petiole (CDs)	10
T_f	maximum expansion time of fruit (CDs)	15
phyllocron	elapsed time between two leaves emergences (days)	2

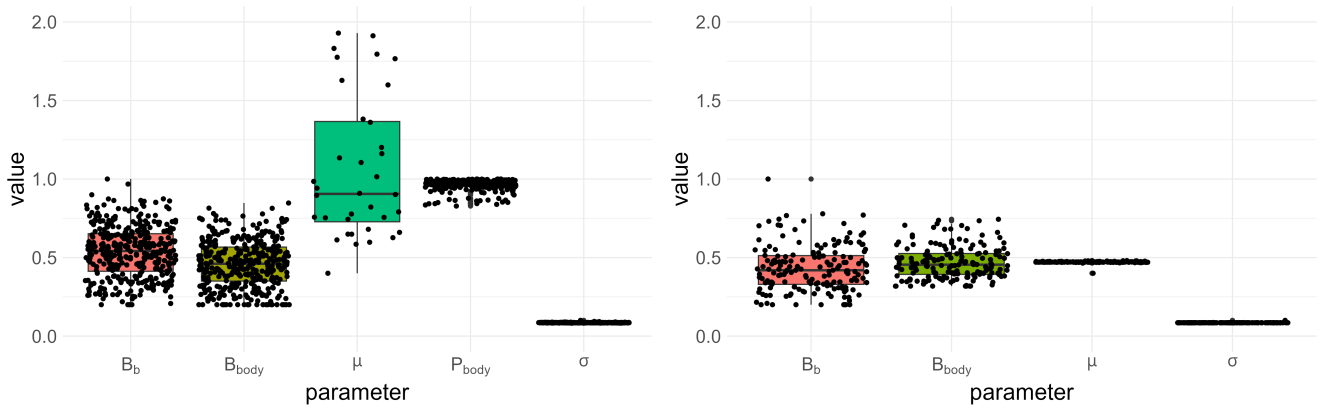


Figure 1. Boxplots of estimated values with similar likelihood for the two cases presented in Section 3.1.. Each dot represents an estimated value. The sink strength of the body compartment (P_{body}) is normalized by its maximum for scaling reasons. (Left) Stabilized parameters: SLA , S_p , RUE , Q_0 . (Right) Stabilized parameters: SLA , S_p , RUE , Q_0 , P_{body} .

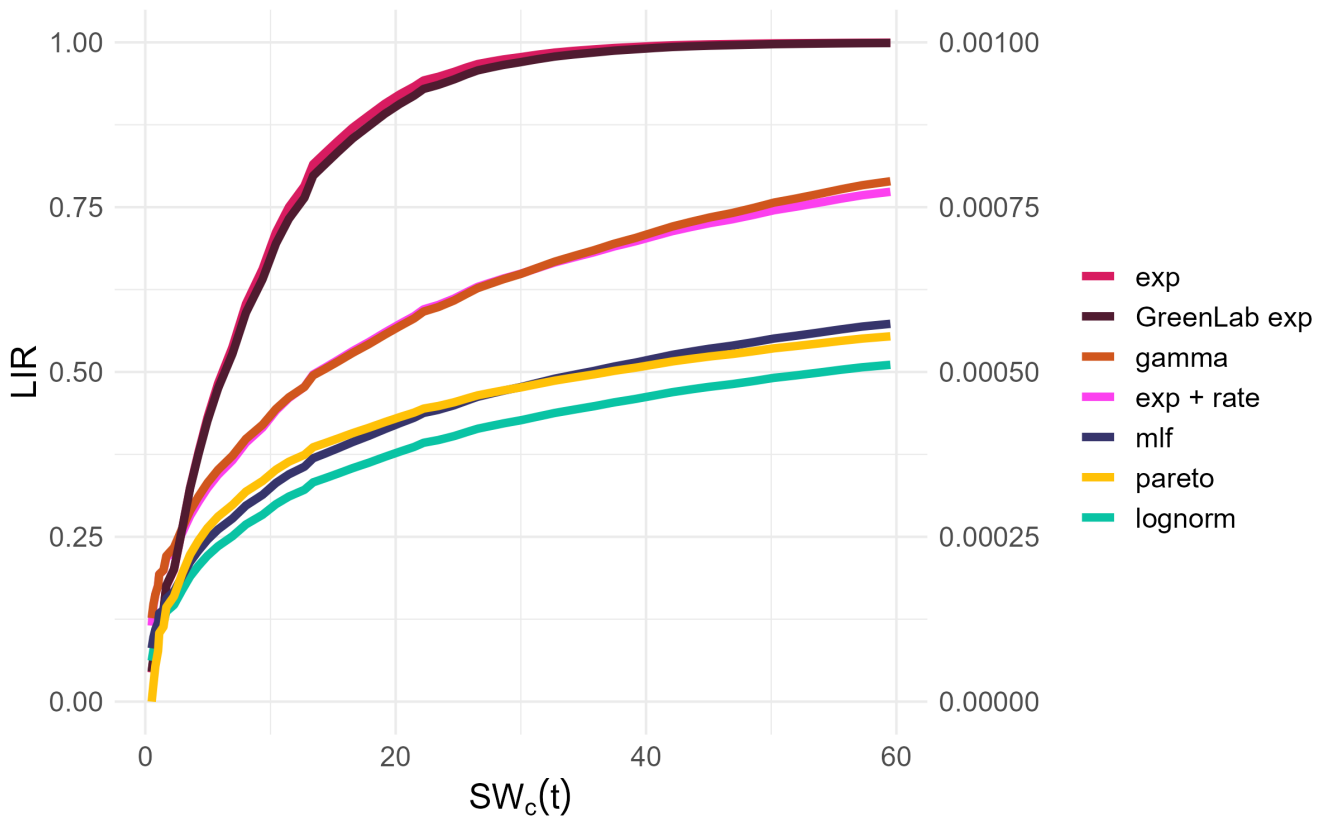


Figure 2. Estimated LIR from the competing models (2.6.2) as a function of the accumulated water usage. The right axis was included for the values of the Pareto distribution. The *Lognorm* and *GreenLab exp* overlap, as well as the *Gamma* and the *LR*.

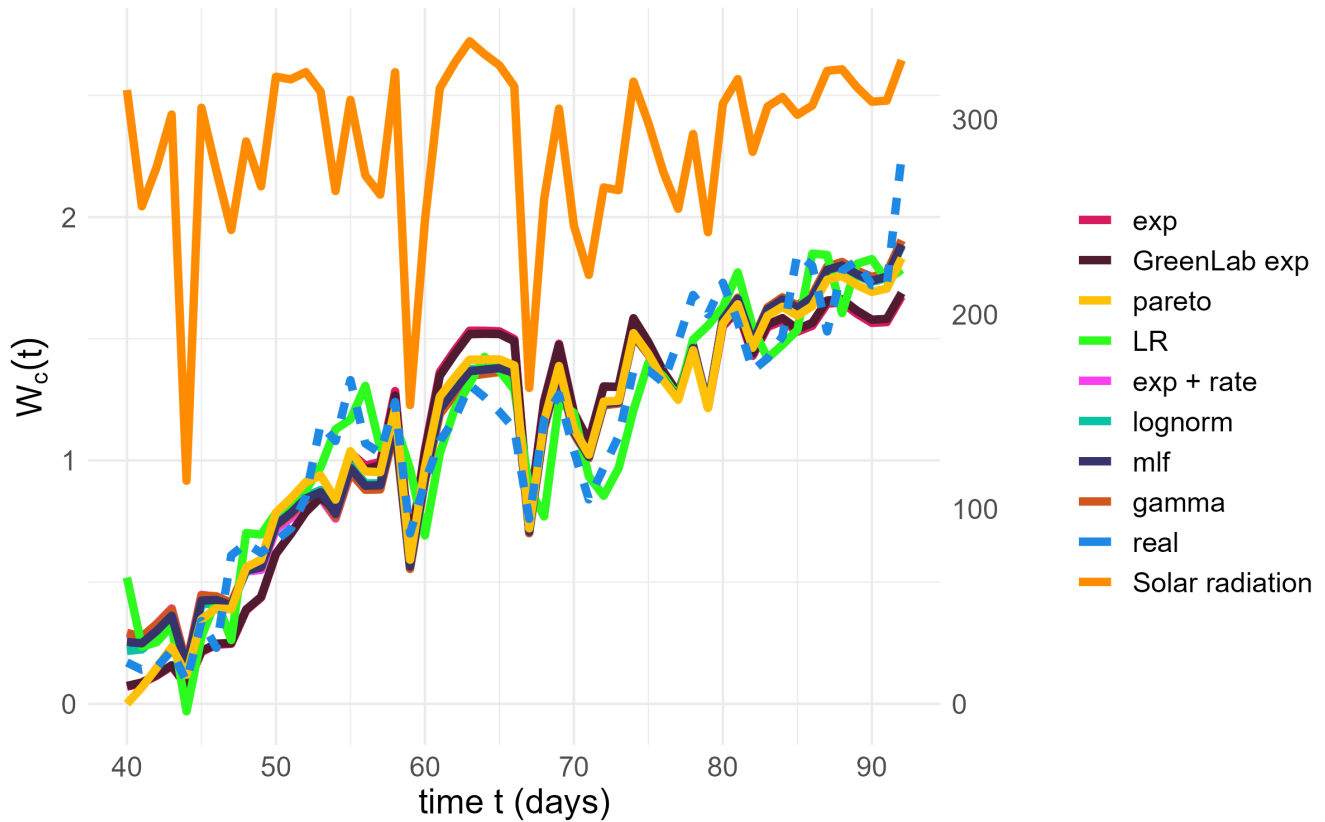


Figure 3. Final fit of the models (solid lines) on the real data (dashed line). Time (days), represented on the x-axis, runs over the days of observation, with $t = 1$ being the day the seed was planted. The left y-axis represents the Water Consumption at time t , in liters. The right y-axis represents values of Avg Solar Radiation (W/m^2). The evolution of Solar Radiation is plotted at the top of the graph, with a dark orange color.

572 1.3 Supplementary figure: normalized LIR w.r.t. cumulated water uptake

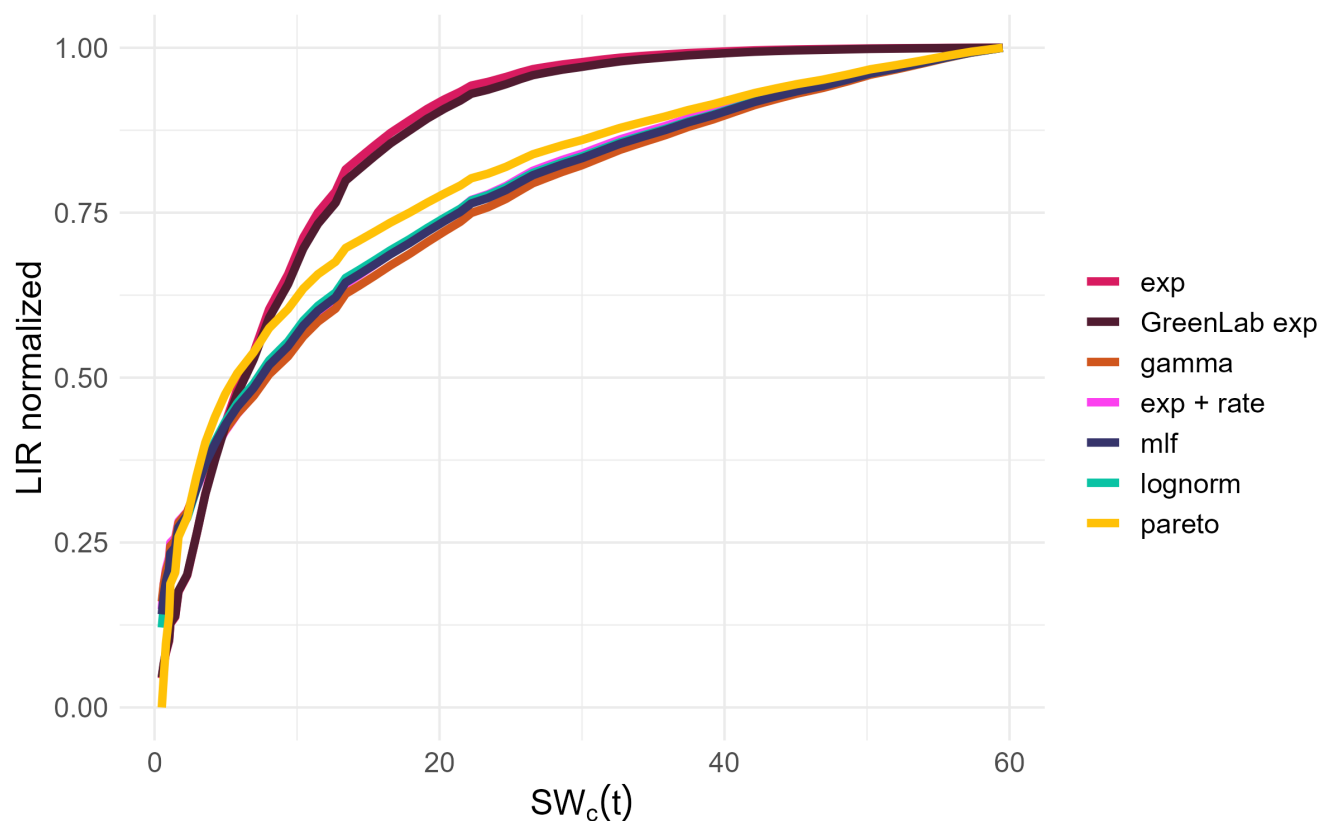


Figure 4. Estimated LIR from the competing models (2.6.2) as a function of the accumulated water usage.

REFERENCES

- 573 Bateman, H. (1953). *Higher transcendental functions [volumes i-iii]*, vol. 1 (McGRAW-HILL book
574 company)
- 575 Bhaskara, G. B., Lasky, J. R., Razzaque, S., Zhang, L., Haque, T., Bonnette, J. E., et al. (2022). Natural
576 variation identifies new effectors of water-use efficiency in arabidopsis. *Proceedings of the National
577 Academy of Sciences* 119, e2205305119
- 578 Biswas, S., Akanda, A., Rahman, M., and Hossain, M. (2016). Effect of drip irrigation and mulching on
579 yield, water-use efficiency and economics of tomato. *Plant, Soil and Environment* 61, 97–102
- 580 Blankenagel, S., Yang, Z., Avramova, V., Schön, C.-C., and Grill, E. (2018). Generating plants with
581 improved water use efficiency. *Agronomy* 8, 194
- 582 Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained
583 optimization. *SIAM Journal on scientific computing* 16, 1190–1208
- 584 Casasanta, G. and Garra, R. (2018). Towards a generalized beer-lambert law. *Fractal and Fractional* 2, 8
- 585 Chang, J. T. and Pollard, D. (1997). Conditioning as disintegration. *Statistica Neerlandica* 51, 287–317
- 586 Cheng, M., Wang, H., Fan, J., Xiang, Y., Tang, Z., Pei, S., et al. (2021). Effects of nitrogen supply on
587 tomato yield, water use efficiency and fruit quality: A global meta-analysis. *Scientia Horticulturae* 290,
588 110553

- 589 Chew, Y. H., Wenden, B., Flis, A., Mengin, V., Taylor, J., Davey, C. L., et al. (2014). Multiscale digital
590 arabidopsis predicts individual organ and whole-organism growth. *Proceedings of the National Academy
591 of Sciences* 111, E4127–E4136
- 592 De Reffye, P., Hu, B., Kang, M., Letort, V., and Jaeger, M. (2021). Two decades of research with the
593 greenlab model in agronomy. *Annals of Botany* 127, 281–295
- 594 Dong, Q., Louarn, G., Wang, Y., Barczi, J.-F., and De Reffye, P. (2008). Does the structure–function model
595 greenlab deal with crop phenotypic plasticity induced by plant spacing? a case study on tomato. *Annals
596 of botany* 101, 1195–1206
- 597 Duursma, R. A., Falster, D. S., Valladares, F., Sterck, F. J., Pearcy, R. W., Lusk, C. H., et al. (2012).
598 Light interception efficiency explained by two simple variables: a test using a diversity of small-to
599 medium-sized woody plants. *New Phytologist* 193, 397–408
- 600 [Dataset] Efento (2020). Wireless temperature logger - bluetooth low energy. Accessed: 18-07-2023
- 601 Farquhar, G. D., von Caemmerer, S. v., and Berry, J. A. (1980). A biochemical model of photosynthetic co
602 2 assimilation in leaves of c 3 species. *planta* 149, 78–90
- 603 [Dataset] Food and of the United Nations (FAO), A. O. (2012). Irrigation water requirement. Accessed:
604 2023-09-07
- 605 Food and Agriculture Organization of the United Nations (1998). *Crop evapotranspiration - Guidelines
606 for computing crop water requirements* (Food and Agriculture Organization of the United Nations).
607 Accessed: 2023-07-12
- 608 Gill, G. and Straka, P. (2018). MittagLeffler: Using the mittag–leffler distributions in r. 2018. URL
609 <https://strakaps.github.io/MittagLeffleR>
- 610 Guo, Y., Ma, Y., Zhan, Z., Li, B., Dingkuhn, M., Luquet, D., et al. (2006). Parameter optimization and
611 field validation of the functional–structural model greenlab for maize. *Annals of botany* 97, 217–230
- 612 Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning:
613 data mining, inference, and prediction*, vol. 2 (Springer)
- 614 Haubold, H. J., Mathai, A. M., Saxena, R. K., et al. (2011). Mittag-leffler functions and their applications.
615 *Journal of applied mathematics* 2011
- 616 Heuvelink, E. (1995). Dry matter partitioning in a tomato plant: one common assimilate pool? *Journal of
617 Experimental Botany* 46, 1025–1033
- 618 Heuvelink, E., Bakker, M., Elings, A., Kaarsemaker, R., and Marcelis, L. (2004). Effect of leaf area
619 on tomato yield. In *International Conference on Sustainable Greenhouse Systems-Greensys2004* 691.
620 43–50
- 621 Higashide, T. and Heuvelink, E. (2009). Physiological and morphological changes over the past 50 years
622 in yield components in tomato. *Journal of the American Society for Horticultural Science* 134, 460–465
- 623 Hirose, T. (2005). Development of the monsi–saeki theory on canopy structure and function. *Annals of
624 botany* 95, 483–494
- 625 Howell, T., Davis, K., McCormick, R., Yamada, H., Walhood, V. T., and Meek, D. (1984). Water use
626 efficiency of narrow row cotton. *Irrigation Science* 5, 195–214
- 627 Howell, T. and Musick, J. (1985). Relationship of dry matter production of field crops to water consumption.
628 In *Les besoins en eau des cultures, conférence internationale, Paris, Versailles, 11-14 septembre 1984*.
629 247–269
- 630 Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples.
631 *Biometrika* 76, 297–307
- 632 Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical review* 106, 620

- 633 Juran, J. M. and De Feo, J. A. (2010). *Juran's quality handbook: the complete guide to performance*
634 *excellence* (McGraw-Hill Education)
- 635 Kostinski, A. B. (2002). On the extinction of radiation by a homogeneous but spatially correlated random
636 medium: reply to comment. *JOSA A* 19, 2521–2525
- 637 Lanoue, J., Leonardos, E. D., Ma, X., and Grodzinski, B. (2017). The effect of spectral quality on
638 daily patterns of gas exchange, biomass gain, and water-use-efficiency in tomatoes and lisianthus: An
639 assessment of whole plant measurements. *Frontiers in plant science* 8, 1076
- 640 Letort, V., Cournède, P.-H., and De Reffye, P. (2009). Impact of topology on plant functioning: a theoretical
641 analysis based on the greenlab model equations. In *2009 Third International Symposium on Plant*
642 *Growth Modeling, Simulation, Visualization and Applications* (IEEE), 341–348
- 643 Letort, V., Sabatier, S., Akaffou, S., Hamon, S., Hamon, P., and De Reffye, P. (2012). Interspecific
644 variability of biomass production of young coffea: no influence of branch pruning. experimental evidence
645 and theoretical analysis. In *2012 IEEE 4th International Symposium on Plant Growth Modeling,*
646 *Simulation, Visualization and Applications* (IEEE), 224–227
- 647 Liu, S., Baret, F., Abichou, M., Manceau, L., Andrieu, B., Weiss, M., et al. (2021). Importance of the
648 description of light interception in crop growth models. *Plant Physiology* 186, 977–997
- 649 Logothetis, D., Malefaki, S., Trevezas, S., and Cournède, P.-H. (2022). Bayesian estimation for the
650 greenlab plant growth model with deterministic organogenesis. *Journal of Agricultural, Biological and*
651 *Environmental Statistics* 27, 63–87
- 652 Ma, C., Liu, M., Ding, F., Li, C., Cui, Y., Chen, W., et al. (2022). Wheat growth monitoring and yield
653 estimation based on remote sensing data assimilation into the safy crop growth model. *Scientific Reports*
654 12, 5473
- 655 Marcelis, L., Elings, A., De Visser, P., and Heuvelink, E. (2008). Simulating growth and development of
656 tomato crop. In *International Symposium on Tomato in the Tropics 821*. 101–110
- 657 Martínez-Ruiz, A., López-Cruz, I. L., Ruiz-García, A., Pineda-Pineda, J., and Prado-Hernández, J. V.
658 (2019). Hortsyst: A dynamic model to predict growth, nitrogen uptake, and transpiration of greenhouse
659 tomatoes. *Chilean journal of agricultural research* 79, 89–102
- 660 Monteith, J. L. (1977). Climate and the efficiency of crop production in britain. *Philosophical Transactions*
661 *of the Royal Society of London. B, Biological Sciences* 281, 277–294
- 662 Ohashi, Y., Murai, M., Ishigami, Y., and Goto, E. (2022). Light-intercepting characteristics and growth of
663 tomatoes cultivated in a greenhouse using a movable bench system. *Horticulturae* 8, 60
- 664 Oker-Blom, P. and Smolander, H. (1988). The ratio of shoot silhouette area to total needle area in scots
665 pine. *Forest science* 34, 894–906
- 666 Pillai, R. N. (1990). On mittag-leffler functions and related distributions. *Annals of the Institute of*
667 *statistical Mathematics* 42, 157–161
- 668 Pivetta, C. R., Tazzo, I. F., Maass, G. F., Streck, N. A., and Heldwein, A. B. (2007). Leaf emergence and
669 expansion in three tomato (*lycopersicon esculentum* mill.) genotypes. *Ciência Rural* 37, 1274–1280
- 670 Ponce de León, M. A. and Bailey, B. N. (2019). Evaluating the use of beer's law for estimating light
671 interception in canopy architectures with varying heterogeneity and anisotropy. *Ecological Modelling*
672 406, 133–143. doi:<https://doi.org/10.1016/j.ecolmodel.2019.04.010>
- 673 Reina-Sánchez, A., Romero-Aranda, R., and Cuartero, J. (2005). Plant water uptake and water use
674 efficiency of greenhouse tomato cultivars irrigated with saline water. *Agricultural water management* 78,
675 54–66
- 676 Resh, H. M. (2022). *Hydroponic food production: a definitive guidebook for the advanced home gardener*
677 *and the commercial hydroponic grower* (CRC press)

- 678 Sarlikioti, V., De Visser, P., and Marcelis, L. (2011). Exploring the spatial distribution of light interception
679 and photosynthesis of canopies by means of a functional–structural plant model. *Annals of Botany* 107,
680 875–883
- 681 Schipper, R., van der Meer, M., De Visser, P., Heuvelink, E., and Marcelis, L. (2023). Consequences of
682 intra-canopy and top led lighting for uniformity of light distribution in a tomato crop. *Frontiers in plant*
683 *science* 14, 1012529
- 684 Schmidt, D., Zamban, D. T., Prochnow, D., Caron, B. O., Souza, V. Q., Paula, G. M., et al. (2017).
685 Phenological characterization, phyllochron and thermal requirement of italian tomato in two cropping
686 seasons. *Horticultura Brasileira* 35, 89–96
- 687 Schröter, M., Härdtle, W., and von Oheimb, G. (2012). Crown plasticity and neighborhood interactions of
688 european beech (*fagus sylvatica* l.) in an old-growth fores. *European Journal of Forest Research* 131,
689 787–798. doi:<https://doi.org/10.1007/s10342-011-0552-y>
- 690 Shabanov, N. and Gastellu-Etchegorry, J.-P. (2018). The stochastic beer–lambert–bouguer law for
691 discontinuous vegetation canopies. *Journal of Quantitative Spectroscopy and Radiative Transfer*
692 214, 18–32. doi:<https://doi.org/10.1016/j.jqsrt.2018.04.021>
- 693 Shamshiri, R. R., Jones, J. W., Thorp, K. R., Ahmad, D., Man, H. C., and Taheri, S. (2018). Review of
694 optimum temperature, humidity, and vapour pressure deficit for microclimate evaluation and control in
695 greenhouse cultivation of tomato: a review. *International agrophysics* 32, 287–302
- 696 Sievänen, R., Godin, C., DeJong, T. M., and Nikinmaa, E. (2014). Functional–structural plant models: a
697 growing paradigm for plant studies. *Annals of botany* 114, 599–603
- 698 Sigrimis, N., Arvanitis, K., Pasgianos, G., and Ferentinos, K. (2001). Hydroponics water management
699 using adaptive scheduling with an on-line optimiser. *Computers and Electronics in Agriculture* 31,
700 31–46
- 701 Tao, H., Ebtehaj, I., Bonakdari, H., Heddam, S., Voyant, C., Al-Ansari, N., et al. (2019). Designing a
702 new data intelligence model for global solar radiation prediction: application of multivariate modeling
703 scheme. *Energies* 12, 1365
- 704 Trevezas, S., Malefaki, S., and Cournède, P.-H. (2014). Parameter estimation via stochastic variants of the
705 ecm algorithm with applications to plant growth modeling. *Computational Statistics & Data Analysis*
706 78, 82–99
- 707 Valladares, F., Garcia-Plazaola, J., Morales, F., and Niinemets, U. (2012). Photosynthetic responses to
708 radiations. In *Terrestrial Photosynthesis in a Changing Environment: A Molecular, Physiological and*
709 *Ecological Approach*, eds. J. Flexas, F. Loreto, and H. Medrano (New York: Cambridge University
710 Press). 239–258
- 711 Van der Zande, D., Stuckens, J., Verstraeten, W. W., Muys, B., and Coppin, P. (2010). Assessment of light
712 environment variability in broadleaved forest canopies using terrestrial laser scanning. *Remote Sensing*
713 2, 1564–1574
- 714 Wang, F., Letort, V., Lu, Q., Bai, X., Guo, Y., De Reffye, P., et al. (2012). A functional and structural
715 mongolian scots pine (*pinus sylvestris* var. *mongolica*) model integrating architecture, biomass and
716 effects of precipitation. *PLOS ONE* 7
- 717 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to
718 the tidyverse. *Journal of open source software* 4, 1686
- 719 Wilson, J. W., Hand, D., and Hannah, M. (1992). Light interception and photosynthetic efficiency in some
720 glasshouse crops. *Journal of Experimental Botany* 43, 363–373
- 721 Winn, C. A., Archontoulis, S., and Edwards, J. (2023). Calibration of a crop growth model in apsim for 15
722 publicly available corn hybrids in north america. *Crop Science* 63, 511–534

- 723 Wu, L., De Reffye, P., Hu, B.-G., Le Dimet, F.-X., and Cournède, P.-H. (2005). A water supply optimization
724 problem for plant growth based on greenlab model. *Revue Africaine de la Recherche en Informatique et*
725 *Mathématiques Appliquées* 3, 197–207
- 726 Yan, H.-P., Kang, M. Z., De Reffye, P., and Dingkuhn, M. (2004). A dynamic, architectural plant model
727 simulating resource-dependent growth. *Annals of botany* 93, 591–602
- 728 Zhang, B., Kang, M., Letort, V., Wang, X., and De Reffye, P. (2009). Comparison between empirical or
729 functional sinks of organs-application on tomato plant. In *2009 Third International Symposium on Plant*
730 *Growth Modeling, Simulation, Visualization and Applications* (IEEE), 191–197