Journal of Cleaner Production 324 (2021) 129120

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Duckweed growth model for large-scale applications: Optimizing harvesting regime and intrinsic growth rate via machine learning to maximize biomass yields

Ozgul Calicioglu [a], Mert Y. Sengul [b], Pandara Valappil Femeena [a], Rachel A. Brennan [a,*]

[a] Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA, 16802, USA
[b] Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA, 16802, USA

ARTICLE INFO

Handling Editor: Mingzhou Jin

Keywords:
Wastewater upcycling
Duckweed yield
Duckweed harvesting
Stella architect
LASSO regression
Deep learning

ABSTRACT

Duckweed has emerged as a potential feedstock for the environmentally sustainable and economically viable production of biofuels and protein. The aim of this study was to: (1) enhance an existing intrinsic duckweed growth model and use it to develop a general regression model that enables users to easily predict annual duckweed yield for large scale applications; and (2) determine the optimal parameter sets that produce the highest annual duckweed yield at a specific location for known values of daily temperature and photoperiod. Simulations performed using Stella Architect were used to compute annual duckweed yield and generate separate datasets for developing the regression model and optimization model. To improve duckweed yields for large-scale applications which incorporate regular harvesting, a harvesting regime was added to the intrinsic duckweed growth model. Two new parameters (harvest frequency and harvest ratio) and a control (harvest threshold) were used to describe the harvesting regime in the model. The general model was developed by fitting LASSO regression ($R^2 = 0.95$) with four variables: initial mat density, intrinsic growth rate, harvest ratio, and harvest frequency. This model offers a simple method for users to estimate annual duckweed yield in practical applications without the need for dynamic simulation runs. Optimum parameter values to maximize biomass production at a location in southwest Florida, USA, were determined using an optimization framework involving a deep neural network machine learning algorithm. Using an existing daylength model to predict daily photoperiod and inputting local temperature data, machine learning calculated a maximum yield of 70 dry tons per hectare per year for the Florida case study, under the following conditions: initial mat density = 169 g_{dry} m^{-2}; harvest threshold = 76 g_{dry} m^{-2}; nitrogen = 50.1 mg L^{-1}; phosphorus = 7.5 mg L^{-1}; harvest ratio = 0.35; and harvest frequency = 1 day.

1. Introduction

The economic and environmental disadvantages of fossil fuel consumption have increased the search for alternative resources to fulfill the world's growing energy and chemical needs (Jung et al., 2016). At the same time, conventional bioenergy crops have also been posing social, economic, and environmental challenges. Duckweed (*Lemnaceae*), is a technically feasible alternative feedstock due to several advantages, such as its: starch accumulation capacity (Zhao et al., 2015); small size (0.1 cm–1 cm); uniform structure; and low lignin content (1%–3%). Since duckweed floats, its harvesting is also simpler compared to microalgae (Cui and Cheng, 2015).

Duckweeds offer a sustainable pathway to promote a circular bioeconomy by upcycling waste nutrients into valuable bio-based products. Since they are resilient to a broad range of nutrient concentrations, duckweeds can be grown on a variety of wastewater sources (Cheng and Stomp, 2009). Due to their rapid growth rate and ease of harvesting, duckweeds have the potential to be used for large-scale bioenergy production ensuring uninterrupted feedstock supply (Calicioglu et al., 2021). In addition, their high protein content of up to 45% (of the dry weight) makes duckweeds an excellent candidate to be used as a protein supplement for humans and animals such as broiler chickens, laying hens, ducks, pigs, ruminants, fish, and shrimp (reviewed in Roman et al., 2021). Since duckweeds can be vertically farmed and thereby cultivated

* Corresponding author.
E-mail address: rab44@psu.edu (R.A. Brennan).

https://doi.org/10.1016/j.jclepro.2021.129120
Received 9 April 2021; Received in revised form 8 September 2021; Accepted 19 September 2021
Available online 30 September 2021
0959-6526/

within a lower areal footprint than land grown crops, utilizing them as a substitute for other protein feeds could result in lower land use changes and greenhouse gas emissions (Roman et al., 2021). Several studies have demonstrated duckweed as a promising feedstock for the production of bioethanol (Yu et al., 2014), biomethane (Calicioglu and Brennan, 2018), and volatile fatty acids (Calicioglu et al., 2018). Calicioglu et al. (2019) further examined the potential of duckweed as a biorefinery feedstock to produce multiple end products sequentially. Other studies have directly examined the value of duckweed as soil amendment to replace conventional fertilizers (Kreider et al., 2019), tested the quality of extracted duckweed protein as a supplement for animal feed (Roman et al., 2021), and measured the protein yield of duckweed grown under different water quality conditions such as swine wastewater (Mohedano et al., 2012) and wastewater from an ecological treatment system (Roman and Brennan, 2019). Regardless of the end use, framing out a complete biorefinery approach is essential for delivering competitive products to end-user markets. Such a comprehensive approach, in turn, necessitates a robust, reliable, and sustainable feedstock supply.

Duckweed growth is primarily affected by factors such as nutrient concentrations, mat density, temperature, and photoperiod (Lasfar et al., 2007). Several growth models have been proposed for studying duckweed growth, including a popular one by Lasfar et al. (2007) which correlates intrinsic growth rate to temperature and photoperiod, as well as nitrogen (N) and phosphorous (P) concentrations in the growth medium. Other experimental studies have reported optimal values of temperature (Zhao et al., 2014), nutrient concentrations (Al-Nozaily et al., 2000), metal concentrations (Boniardi et al., 1999), and light intensity (Filbin and Hough, 1985) beyond which duckweed growth will be slowed. However, previous models and experimental studies have not focused on continuous large-scale production of duckweed, which can be significantly influenced by the harvesting regime.

Harvest frequency (the time interval at which biomass is harvested), harvest ratio (the ratio of biomass harvested to the total available biomass), and harvest threshold (the biomass quantity above which harvesting takes place) are critical parameters in a harvesting regime. Frédéric et al. (2006) developed a mathematical growth model involving harvest frequency in which they demonstrated that a higher harvesting frequency results in a higher specific growth rate of duckweed. While revealing that duckweed growth can be affected by the amount of duckweed harvested and the interval at which the harvesting is done, that study did not incorporate a complete harvesting regime into the duckweed growth model. A duckweed growth model that incorporates harvesting regime parameters would help increase the duckweed biomass obtained from a large-scale system. Optimization models coupled to duckweed growth models can further offer opportunities to maximize biomass yield by appropriately optimizing the growth and harvest parameters. While optimization models have been used in the past to estimate optimal parameters for algae-based wastewater treatment (Sundui et al., 2021), to the authors' knowledge, the use of optimization models for maximizing duckweed biomass production have not previously been explored. Real-world large-scale applications of duckweed production, such as a biorefinery, would benefit from such optimization frameworks to predict location-specific optimal parameters.

The objectives of this study were to: (1) enhance an existing intrinsic duckweed growth model and use it to develop a general regression model that enables users to easily predict annual duckweed yield for large scale applications involving regular harvesting; and (2) determine the optimal parameter sets that produce the highest annual duckweed yield at a specific location for known values of daily temperature and photoperiod.

2. Methodology

2.1. General duckweed growth and harvest model development

2.1.1. Intrinsic duckweed growth model

Stella Architect (Version 1.9.1) was used to conduct the simulations in this study. Stella Architect is a dynamic modeling software that enables users to create system diagrams and include features to perform extended simulations. Duckweed growth was simulated according to semi-empirical intrinsic growth model developed by Lasfar et al. (2007) which is based on Michaelis-Menten kinetics as shown in Equations (1)–(3).

$$r_i = R\cdot\theta_1^{\left(\left(T-T_{op}\right)/T_{op}\right)^2}\cdot\theta_2^{\left(\left(T-T_{op}\right)/T_{op}\right)}\cdot\theta_3^{\left(\left(E-E_{op}\right)/E_{op}\right)^2}\cdot\theta_4^{\left(\left(E-E_{op}\right)/E_{op}\right)}\cdot\frac{C_P}{C_P+K_P}\cdot\frac{K_{IP}}{K_{IP}+C_P}\cdot\frac{C_N}{C_N+K_N}\cdot\frac{K_{IN}}{K_{IN}+C_N} \tag{1}$$

$$D=\frac{D_L\cdot D_O}{(D_L-D_O)\cdot e^{-r_i\cdot t}+D_O} \tag{2}$$

$$r_s=\frac{1}{t}\cdot\ln\left(\frac{D}{D_O}\right)=\frac{1}{t}\cdot\ln\left(\frac{D_L}{(D_L-D_O)\cdot e^{-r_i\cdot t}+D_O}\right) \tag{3}$$

where K_P, K_{IP}, K_N, and K_{IN} are the saturation and the inhibition constants of P and N, respectively; C_P and C_N are the P and N concentrations (mg L^{-1}), respectively; R is a constant (maximum intrinsic growth rate in day^{-1}); T is the temperature in °C with T_{op} being the optimum temperature; E is the photoperiod (h); r_i and r_s are the intrinsic and specific growth rates (day^{-1}), respectively; D_o is the initial mat density (g_{dry} m^{-2}) of the duckweed; D is the instant mat density (g_{dry} m^{-2}) (i.e., the duckweed biomass per square meter of covered water surface at a specific moment in time); and D_L is the limiting mat density (i.e. the upper limit of the mat density beyond which the duckweed growth is strongly inhibited); t is the duckweed retention time (day); and $\theta_{1\text{-}4}$ are nondimensional constants smaller than one. Values of constants are $K_P = 0.31$, $K_{IP} = 101$, $K_N = 0.95$, $K_{IN} = 604$, $R = 0.62$, $T_{op} = 26$ °C, $\theta_1 = 0.0025$, $\theta_2 = 0.66$, $\theta_3 = 0.0073$, $\theta_4 = 0.65$.

This model considers mat density as a variable, which was particularly important in our study, considering that the harvesting regime changes the mat density frequently. Other parameters used for the description of the general duckweed growth function were N and P concentrations, temperature, initial mat density, limiting mat density, and photoperiod, as well as the proximity of actual values to their optima.

2.1.2. Harvesting module

Two parameters were determined to be critical in estimating the amount of duckweed harvested in the general model: (1) Harvest frequency; and (2) Harvest ratio. Harvest frequency (H_f) is the duration (in days) between two harvesting events, and harvest ratio (H_r) is defined as the ratio of harvested duckweed to the total duckweed available in the pond on the harvest day. To determine the optimum growth and harvesting conditions, a harvesting module with these parameters was incorporated into the intrinsic duckweed growth model (Lasfar et al., 2007).

2.1.3. Incorporation of the harvesting module into the intrinsic growth model

The harvesting module was incorporated to the intrinsic duckweed growth model in Stella Architect (see **Figure S-1**, Supporting Information, for a schematic of the model). The model first estimates the intrinsic growth rate based on the initial values provided, and then calculates the change in duckweed biomass, taking into account the initial and limiting mat densities. The harvest ratio determines the amount of duckweed harvested on each day with a harvesting event (Box D-2, Supporting Information).

2.1.4. Stella simulations for the general duckweed growth and harvest model

The effects of C_N, C_P, T, E, D_o, H_r, and H_f on duckweed yield were examined using Stella simulations. The range of values used for all of these input parameters are equal to the parameter ranges used in the study by Lasfar et al. (2007) and are provided in Table 1.

2.1.5. LASSO regression for the development of the general duckweed growth and harvest model

As can be seen in Table 1, there are seven input parameters in the model, which makes this a high dimensional dataset. To reduce model complexity, four selected parameters – D_o, r_i, H_r, and H_f – were used for fitting the non-linear least absolute shrinkage and selection operator (LASSO) regression model, out of which r_i can be computed according to Eq. (1) using input values of T, E, C_N, and C_P. A fourth order polynomial regression incorporating exponential function and parameter interactions was used to fit the general model. We used LASSO regression because it performs a variable selection via penalization, which shrinks the least influential coefficients of the fitted model to zero and simplifies the final model. In order to explore the dataset in an efficient way, and by taking advantage of the deterministic nature of computer experiments, we used initial design algorithms. Computer simulations, unlike physical experiments, do not have randomness; therefore, such simulation studies can be conducted by smartly designing the input configurations. In this study, we used a Latin Hypercube Design algorithm to generate high dimensional uniformly distributed input configurations by reducing the pairwise correlations and maximizing the distance between the input configurations in high dimensional space (Joseph and Hung, 2008). A total of 30,000 data points were generated to explore the input parameter space. The simulation period for the Stella model was set to 360 days (assuming each month has 30 days) and the annual cumulative yield (D_{annual}, in g_{dry} m^{-2}) obtained with corresponding input values was used as the response variable for regression.

Table 1
Input parameters and their ranges used in the Stella model to compute intrinsic growth rate and cumulative harvest for the general model regression fit.

	Range		Increment	Reference
Input Parameter	Minimum Value	Maximum Value		
Initial Mat Density, D_o (g_{dry} m^{-2})	0.2	200	continuous	Lasfar et al. (2007)
Temperature, T (°C)	5	32	continuous	Lasfar et al. (2007)
Photoperiod, E (h)	2	17	continuous	Lasfar et al. (2007)
N Concentration, C_N (mg L^{-1})	1.1	350	continuous	Lasfar et al. (2007)
P Concentration., C_P (mg L^{-1})	0.12	54	continuous	Lasfar et al. (2007)
Harvest Ratio, H_r	0.05	1	continuous	Introduced in this study
Harvest Frequency, H_f (day)	1	30	1	Introduced in this study

2.2. Florida USA case study

2.2.1. Florida case study model

A location in Florida (FL), USA, was selected for a case study because of the nearly optimal conditions for duckweed growth throughout the year. Another rationale for selecting this location is that it was previously identified by the National Renewable Energy Laboratory (NREL) for testing large-scale algal biomass production (Davis et al., 2016). The coordinates of the city of Fort Myers, FL (26.6406° N, 81.8723° W), were used in the model, due to the potential availability of sufficient wastewater from urban areas to support duckweed growth. For the FL case study, the Stella model described under Section 2.1.4 and parameter ranges listed in Table 1 were used, with the exception that daily inputs of spatially explicit variables - temperature (T) and photoperiod (E) – were used instead of constant values to better simulate biomass yields. The temperature data was retrieved from National Centers for Environmental Information database for Fort Myers, FL (Vose et al., 2014), and ten-year average temperature values (2008–2018) were used for each day in the model. Based on the day length model provided in Forsythe et al. (1995), E was estimated for each calendar day as a function of geographic coordinates at this location (Box D-1 in Supporting Information).

Since T and E vary daily, a new controlling parameter, harvest threshold (H_t), was introduced to avoid depletion of duckweed biomass due to overharvesting. The harvest threshold determines if the harvest will be carried out depending on the available mat density. A mat density higher than the threshold value triggers duckweed harvest for the given day. The H_t parameter introduced for the FL model was considered as a variable in the initial design (with a continuous increment in the range of 0.1–200 g_{dry} m^{-2}) to understand the effect of this parameter on the annual harvest yield.

2.2.2. Stella simulations for the Florida case study

The effects of C_N, C_P, T, E, D_o, H_r, H_f, and H_t on duckweed yield were examined using Stella simulations. For the highest duckweed yields within 10% of the maximum yield, selected scenarios of parameter combinations were analyzed for its practical significance. The dataset consisting of parameter values and Stella simulated cumulative duckweed harvest values were further used in the optimization study as detailed in Section 2.2.3.2.2.3. Optimization of annual duckweed yield for Florida case study.

Optimization was performed for the FL case study using six input parameters (D_o, C_N, C_P, H_r, H_f, and H_t) to maximize the annual harvest yield. We used a machine learning-based high dimensional optimization method that has been previously applied to optimize datasets generated through computer simulations (Sengul et al., 2020; Sengul eet al., 2021). The optimization algorithm (created using Python programming language) starts by fitting a machine learning regression model to the dataset. The dataset was generated by running the Stella model for all parameter combinations generated using the Latin Hypercube Design algorithm. The cumulative annual duckweed harvest yield was used as the response variable for the regression model. We tested three different machine learning methods, namely: (1) linear regression; (2) LASSO regression with polynomial features; and (3) deep neural network regression. The best fit was obtained using the deep neural network model and therefore, only this model is described below.

We used a fully connected feedforward type deep neural network as the machine learning model (Lecun et al., 2015). Feedforward neural network utilizes layers with a unidirectional decision flow in which the weighted sum of inputs from one layer is used to compute the result for the subsequent layer through a non-linear function. Three hidden layers were selected with forty nodes each (Figure S-2, Supporting Information). Nonlinear mapping between layers were incorporated using rectified linear unit (ReLu) activation function defined as $f(x) = \max(x, 0)$ (Glorot et al., 2011). In order to train the deep learning model, the dataset was randomly separated into training and test sets using a 4:1

ratio, and the trained model was tested on the test dataset. The mean absolute error was used as a measure of accuracy between the predicted and true values (i.e. dataset values) during the training of the deep learning model. The accuracy of the prediction capability of our model was measured using the Coefficient of Determination (R^2) between the predicted and true values, with $R^2 = 1$ representing the best fit model.

We used a brute force type of optimization algorithm to find the parameter set that produces the maximum annual harvest yield. The brute force algorithm works as follows. The algorithm assigns randomly generated values to each input parameter within specified ranges. The harvest yield value is then calculated for each randomly generated input parameter set using the fitted deep neural network model. The resulting harvest yield value is compared with the highest previously obtained value. The algorithm iterates until the highest possible harvest yield value is obtained within a given number of iterations with each iteration taking less than 1 ms. In this study, the criterion for stopping the algorithm was to obtain a harvest yield value that was higher than the ones in the dataset. In order to obtain the maximum annual yield value, the algorithm iterated around two million times.

3. Results and discussion

3.1. General duckweed growth and harvest model development

3.1.1. Stella simulations for the general duckweed growth and harvest model

The cumulative annual duckweed yields predicted for varying parameter values using the Stella Model are illustrated in Fig. 1. A wide range of yield values were obtained across the entire range of initial mat density (D_o), nitrogen concentration (C_N), and phosphorus concentration (C_P) values. Biomass yield tends to increase with increasing nitrogen concentration (C_N) but tapers off for concentrations above 100 mg L^{-1}. A similar trend was observed for phosphorus (C_P) above 10 mg L^{-1}. These are in agreement with the intrinsic growth rate versus nutrient concentration curve reported by Lasfar et al. (2007) that follows Michaelis–Menten kinetics with duckweed growth decline occurring at $C_P > 10$ mg L^{-1} and $C_N > 40$ mg L^{-1}. The remaining parameters (harvest ratio (H_r), temperature (T), and photoperiod (E)) showed specific trends in terms of optimum values that produce the highest theoretical duckweed yield. For example, the highest yields in the general model were produced for harvest ratio (H_r) > 0.2, harvest frequency (H_f) between 1 and 5 days, temperature (T) between 20 and 27 °C, and photoperiod (E) between 10 and 15 h. For comparison, Lasfar at al. (2007) reported optimal duckweed growth at values of 26 °C and 13 h for temperature and photoperiod, respectively. Although the highest yields were observed within these optimum parameter ranges, having an optimal parameter value would not always result in highest yields, since a sub-optimal value of one parameter can affect the yield, which explains the very low yield values within these ranges in Fig. 1. As expected, the trend of intrinsic growth rate (r_i) versus annual duckweed yield showed that higher intrinsic growth rate directly relates to higher yield values. However, depending on the harvest parameter values, the yield could be low even at high r_i values.

The tradeoffs in yield values with two parameter combinations (C_N and C_P; H_r and H_f; T and E) were studied using heat maps (Fig. 2). Within the range of N and P concentrations considered, lower values of C_N and C_P correspond to higher duckweed yield values. Moreover, N limitation could be compensated by an increase in P concentration to achieve high

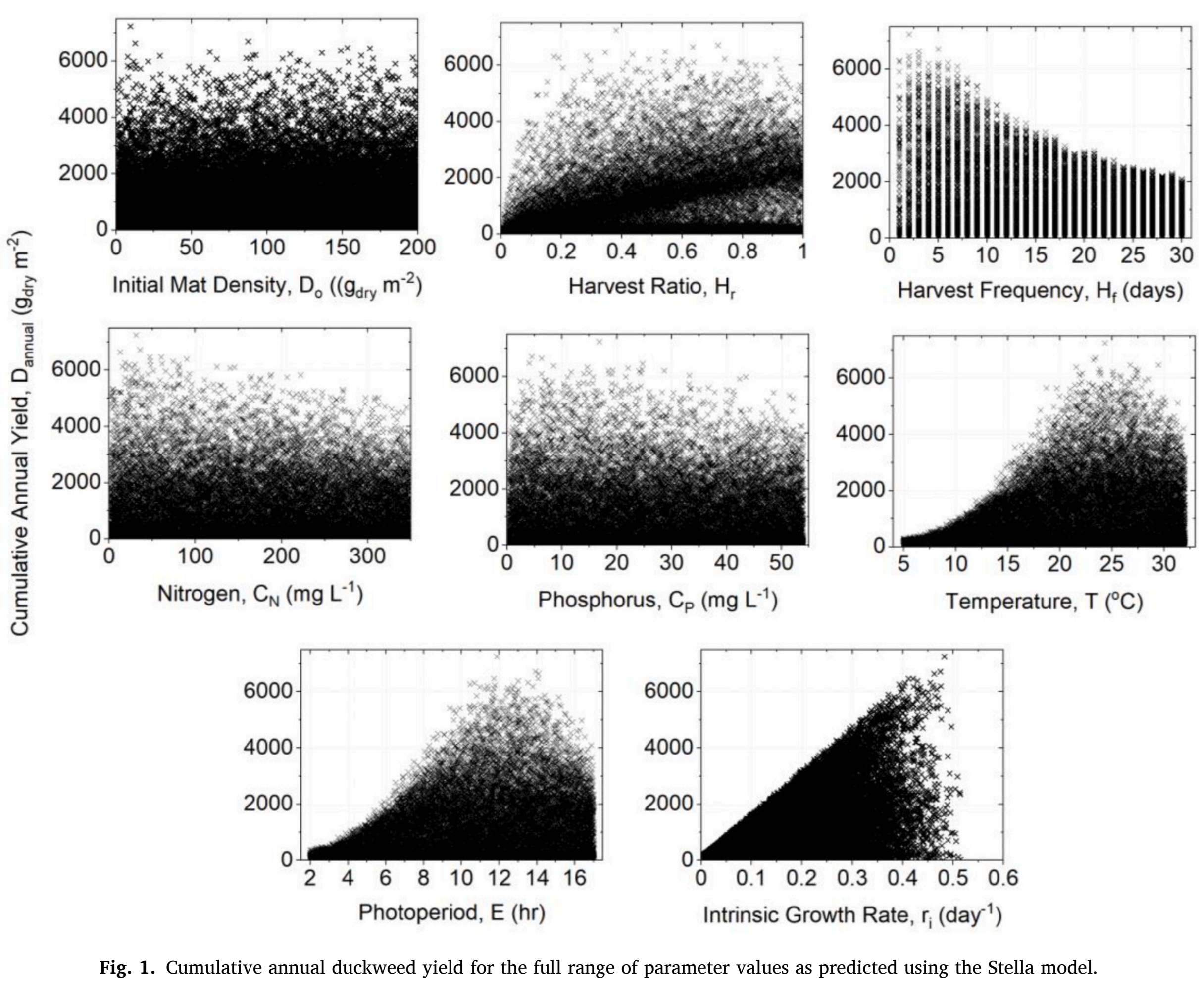

Fig. 1. Cumulative annual duckweed yield for the full range of parameter values as predicted using the Stella model.

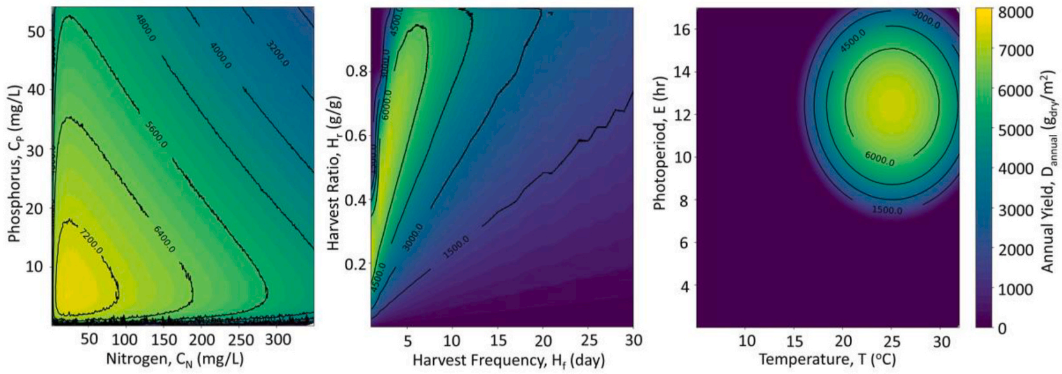

Fig. 2. Maximum annual duckweed yield for the enhanced duckweed growth and harvest model as a function of: A) initial N and P concentrations; B) harvest ratio and harvest frequency; C) photoperiod and temperature.

yields, and vice versa. With respect to harvest parameters, duckweed yields were highest with low harvest frequency (H_f) values, provided the harvest ratio (H_r) > 0.2. In agreement with Fig. 1 which demonstrated optimum values of temperature (T = 20–27 °C) and photoperiod (E = 10–15 h), heat maps revealed the highest yields corresponding to these optimum values.

3.1.2. LASSO regression and development of the general duckweed growth and harvest model

The non-linear LASSO regression model fitted with four parameters (D_o, r_i, H_f, and H_r) and their transformations yielded an analytical equation for the general duckweed growth and harvest model (Eq. (4)). The combination of Eq. (1) and Eq. (4) offers a simple method to estimate long-term duckweed yield under a given set of operating conditions (D_o, C_N, C_P, T, and E) and harvesting regime (H_r and H_f), without the need for simulation software such as Stella.

$$D_{annual} = a_0 + a_1 e^{r_i} + a_2\left(\frac{H_r}{H_f}\right) + a_3(D_o e^{r_i})^2 + a_4(D_o e^{r_i})\left(\frac{H_r}{H_f}\right) + a_5\left(\frac{H_r}{H_f}\right)^2 + a_6(D_o e^{r_i})^3 + a_7(D_o e^{r_i})^2\left(\frac{H_r}{H_f}\right) + a_8(D_o e^{r_i})\left(\frac{H_r}{H_f}\right)^2 + a_9\left(\frac{H_r}{H_f}\right)^3 + a_{10}(D_o e^{r_i})^4 + a_{11}(D_o e^{r_i})^3\left(\frac{H_r}{H_f}\right) + a_{12}(D_o e^{r_i})^2\left(\frac{H_r}{H_f}\right)^2 + a_{13}(D_o e^{r_i})\left(\frac{H_r}{H_f}\right)^3 + a_{14}\left(\frac{H_r}{H_f}\right)^4 \quad (4)$$

where D_{annual} is the cumulative annual duckweed yield (g_{dry} m^{-2}).

The range of parameter values used for model fitting were: Initial mat density (D_o) = 0.4–200 g_{dry} m^{-2}; intrinsic growth rate (r_i) = 0.3–0.51 day^{-1}; harvest ratio (H_r) = 0.05–0.9; and harvest frequency (H_f) = 1–30 days. These r_i ranges were further narrowed down from that used by Lasfar et al. (2007) in order to exclude parameter values that produce very low yield. Eq. (4) is therefore only valid for the parameter ranges which primarily exclude lower growth rates (r_i < 0.3 day^{-1}). Regression coefficients corresponding to each variable in Eq. (4) are provided in Table 2. The equation showed that H_r/H_f and its polynomial terms had the largest effect on model prediction, as seen by its high coefficients. This also indicates the high influence of harvest parameters on duckweed yield when compared to other parameters in the model including initial mat density and intrinsic growth rate. An R^2 value of 0.95 was obtained between D_{annual} values simulated by Stella model and those predicted using the developed regression equation. The regression model fit is illustrated in the Supporting Information (Figure S-3). A slight deviation was observed between the predicted and true values of high duckweed yields indicating model underprediction of these high yield values. This can be attributed to the smaller number of parameter combinations producing high yield and hence a limited amount of high yield data available for model fitting. Although studies in the past have

Table 2
Coefficients corresponding to the fitted general duckweed growth and harvest model.

Parameter	Coefficient	Value
Bias	a_0	0[a]
$D_o e^{r_i}$	a_1	43.79
H_r/H_f	a_2	4658.63
$(D_o e^{r_i})^2$	a_3	0[a]
$D_o e^{r_i}.(H_r/H_f)$	a_4	60.20
$(H_r/H_f)^2$	a_5	−4215.52
$(D_o e^{r_i})^3$	a_6	0[a]
$(D_o e^{r_i})^2.(H_r/H_f)$	a_7	0[a]
$D_o e^{r_i}.(H_r/H_f)^2$	a_8	9.04
$(H_r/H_f)^3$	a_9	948.01
$(D_o e^{r_i})^4$	a_{10}	−1.40
$(D_o e^{r_i})^3.(H_r/H_f)$	a_{11}	−7.82
$(D_o e^{r_i})^2.(H_r/H_f)^2$	a_{12}	−1.49
$D_o e^{r_i}.(H_r/H_f)^3$	a_{13}	−2.89
$(H_r/H_f)^4$	a_{14}	−64.77

[a] Parameters with values of zero (0) indicate that, although considered, they were found through penalized regression to not have a significant effect on the model.

developed regression equations relating duckweed growth to temperature, photoperiod, and metal concentrations (Diritgen and Nel, 1994; Rejmánková, 1973; Zhang et al., 2009), this study is the first time a regression model has been developed to predict long-term duckweed yield. As field-scale data on annual duckweed yield is collected from facilities operating with a defined environment and harvesting regime, the model proposed in this study can further be validated.

3.2. Florida case study

3.2.1. Stella simulations for the Florida case study

The dataset involving D_o, C_N, C_P, H_r, H_f, and H_t was generated through an initial design algorithm, and was used as an input into Stella simulations. Similar to the plots from the general duckweed growth and harvest model, the overall trends in the Stella simulations show that the initial mat density (D_o) does not have significant impact on the end result (Fig. 3). Frédéric et al. (2006) reported that duckweed growth rate decreases with increasing D_o, with the highest growth rate occurring for D_o between 3.4 and 9.6 g_{dry} m^{-2} and lower growth rates observed for D_o between 86 and 128 g_{dry} m^{-2}. The insignificant impact of D_o in this study was expected, as the dynamic nature of the harvesting module reduces the effect of D_o over the time duration of 1 year. The harvest threshold (H_t) also did not have a major impact if it is below 150 g_{dry}

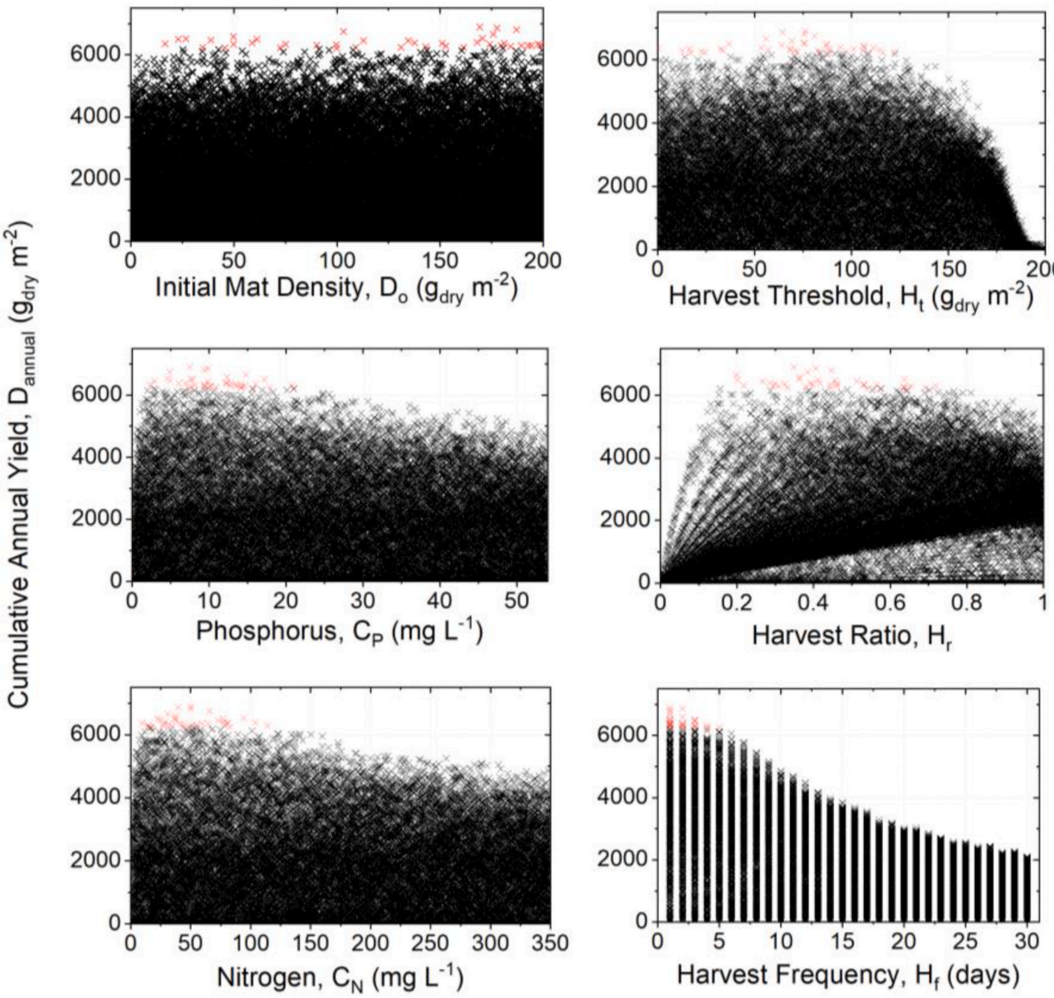

Fig. 3. Cumulative annual duckweed yields obtained for the Florida case study over the range of parameter values considered during optimization. Red highlighted points indicate data within 10% of the maximum yield value.

m^{-2}. Increasing nutrient concentrations were observed to increase biomass yields, but the effect tapered off above $C_N = 100$ mg L^{-1} and $C_P = 10$ mg L^{-1}. The system appears to be more resilient to changing operational conditions and less dependent on the nutrient concentrations within a certain range of harvesting ratio (H_r). In general, the harvest frequency (H_f) inversely affects the yield across the entire parameter range (1–30 days) considered, which means more frequent harvesting results in higher yields.

An analysis within 10% of the maximum yield was also performed to identify various patterns and scenarios among the dataset (see subset of values in Table 3 and full dataset in Tables S–2, Supporting Information). This maximum yield dataset includes parameter values in the following ranges: $D_o = 17$–199 g_{dry} m^{-2}; $C_N = 9.6$–113.9 mg L^{-1}; $C_P = 2.4$–21 mg L^{-1}; $H_t = 0$–123 g_{dry} m^{-2}; $H_r = 0.20$–0.72; and $H_f = 1$–5 days; $D_{annual} = 6209$ to 6898 g_{dry} m^{-2}. Overall, in terms of nutrient concentrations, the majority of the parameters fall within the classification of raw wastewater or primary effluent for domestic wastewater (Metcalf and Eddy, 2003). Lower values of H_f (i.e., more frequent harvesting) with low H_r produced higher annual yields (Fig. 4). The equifinality property of the parameter set (which causes different combinations of parameters to produce similar duckweed yields) offers flexible options to design duckweed growing systems depending on the availability of nutrients, environmental conditions, and operational or harvesting constraints.

Among the entire dataset, the maximum biomass yield obtained was 6898 g_{dry} m^{-2}. This yield was achieved under the following conditions: $D_o = 169$ g_{dry} m^{-2}; $H_t = 76$ g_{dry} m^{-2}; $C_N = 50.1$ mg L^{-1}; $C_P = 7.5$ mg L^{-1}; $H_r = 0.35$; and $H_f = 1$ day. As can be seen from Fig. 4, there is an inverse correlation between C_N and C_P values, which is identical to the trend observed in general duckweed growth and harvest model. Low N concentration can be compensated with higher P concentration to achieve similar yields, and vice versa, provided minimum required C_N and C_P are available in the growing media. H_f and H_r are directly correlated, and operating at a higher frequency yields higher duckweed biomass. Similar to the N–P concentration correlation, H_r can be adjusted to compensate for H_f, as they can be paired at various optima to obtain high duckweed yields.

Apart from maximum yield, other selected scenarios were analyzed using the high yield dataset (Table 3) as discussed below. Most of the high yield values were observed at low H_f (1 day) and high D_o; however even with low D_o, yields could be maximized by increasing the N/P concentrations or H_r.

3.2.1.1. Low nitrogen scenario. The lowest C_N scenario within the high yield dataset (with $C_N = 9.6$ mg L^{-1}) yields an annual duckweed production value of 6368 g_{dry} m^{-2} (Table 3). The other parameters under this scenario include: $D_o = 181$ g_{dry} m^{-2}; $H_t = 0.1$ g_{dry} m^{-2}; $C_P = 12.3$ mg L^{-1}; $H_r = 0.32$; and $H_f = 2$ days. Low threshold value means high residual duckweed concentration after each harvest, which indicates that the threshold parameter was essentially not triggered in this scenario. The nutrient concentrations are similar to effluent from a wastewater treatment plant that facilitates N removal but with no

Table 3
Subset of Stella simulation parameter sets and results, highlighting selected operational scenarios within 10% of the highest duckweed yield.

Annual Cumulative Yield, D_{annual} (g_{dry} m^{-2})	Shift from max D value (%)	Initial Mat Density, D_o (g_{dry} m^{-2})	Nitrogen, C_N (mg L^{-1})	Phosphorus, C_P (mg L^{-1})	Harvest Threshold, H_t (g_{dry} m^{-2})	Harvest Ratio, H_r (g/g)	Harvest Frequency, H_f (days)
6898▲	0.0▲	169▲	50.1▲	7.5▲	76▲	0.35▲	1▲
6529	5.3	171	94.4	4.6	75	0.36	1
6504	5.7	61	77.0	7.7	37	0.20	1
6499	5.8	23	22.2	17.2	87	0.41	1
6453	6.4	135	37.6	2.9	106	0.27	1
6378	7.5	138	103.9	6.1	54	0.42	2
6368■	7.7■	181■	9.6■	12.3■	0.1■	0.32■	2■
6334✱	8.2✱	50✱	64.8✱	2.4✱	89✱	0.53✱	1✱
6319	8.4	100	43.8	8.2	57	0.56	4
6308	8.5	199	26.4	10.3	115	0.29	1
6294	8.8	193	78.0	9.5	13	0.46	3
6290	8.8	190	14.5	14.7	106	0.72	1
6250♦	9.4♦	143♦	14.5♦	9.5♦	98♦	0.69♦	5♦
6242	9.5	35	82.6	14.2	88	0.33	1
6215	9.9	73	44.2	15.7	17	0.54	3

▲Maximum biomass scenario; ■Low nitrogen scenario; ✱Low phosphorus scenario; ♦Low maintenance scenario

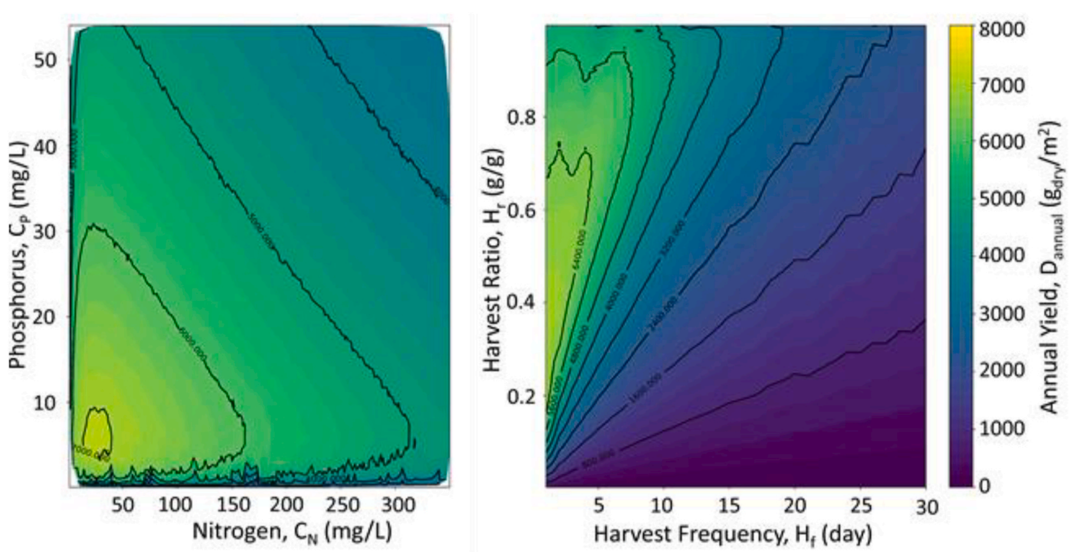

Fig. 4. Maximum annual duckweed yield for the Florida case study as a function of: A) initial N and P concentrations; B) harvest ratio and harvest frequency.

designated P removal units, such as the Modified Ludzack-Ettinger process (Metcalf and Eddy, 2003).

3.2.1.2. Low phosphorus scenario. This scenario was selected for the following parameter set: $D_o = 50$ g_{dry} m^{-2}; $H_t = 89$ g_{dry} m^{-2}; $C_N = 64.8$ mg L^{-1}; $C_P = 2.4$ mg L^{-1}; $H_r = 0.53$; and $H_f = 1$ day (Table 3). This combination of P and N concentrations is typical of raw domestic wastewater in areas where restrictions are imposed on phosphorus detergent use (Sedlak, 1991). In this case, low P concentration can be compensated by high N concentration and frequent harvesting to achieve high biomass yield.

3.2.1.3. Low maintenance scenario. The highest value of H_f (in days) corresponds to the low maintenance scenario in which duckweed is harvested on fewer days hence labor costs are reduced. Under this scenario, the H_f was once in every five days, and H_r was 0.69. The other parameters were set to: $D_o = 143$ g_{dry} m^{-2}; $H_t = 98$ g_{dry} m^{-2}; $C_N = 14.5$ mg L^{-1}; and $C_P = 9.5$ mg L^{-1}. This scenario yields 6250 g_{dry} m^{-2}, which

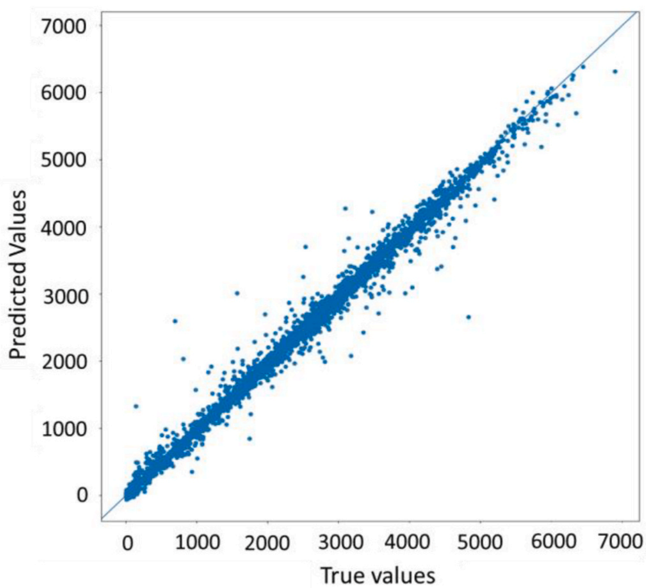

Fig. 5. Optimization model fit ($R^2 = 0.99$) for Florida case study showing predicted and true values of cumulative annual duckweed harvest (g_{dry} m^{-2}).

is 9.6% lower than the maximum (Table 3). However, this is an important scenario to incorporate the labor costs for economical production of duckweed. H_f above 5 days showed decreasing trend in duckweed biomass yield for all combinations of parameter sets.

3.2.2. Optimization of annual duckweed yield for the Florida case study

The machine learning model for FL case study revealed a root mean squared error (RMSE) of 97.63 g_{dry} m^{-2} with testing dataset and 87.17 g_{dry} m^{-2} RMSE with training dataset. An R^2 of 0.99 was obtained for both testing and training datasets (Fig. 5). The results of the brute force optimization (see methods section) predict that a maximum of 7020 g_{dry} m^{-2} (or 70 dry tons ha^{-1}) duckweed can be harvested cumulatively in a year at the location considered under the following growth conditions: $D_o = 85.9$ g_{dry} m^{-2}; $H_t = 98.8$ g_{dry} m^{-2}; $C_N = 17.9$ mg L^{-1}; $C_P = 4.9$ mg L^{-1}; $H_r = 0.2$; and $H_f = 1$ day. This yield value is 1.8% higher than the maximum annual yield obtained through Stella simulations. The optimal N and P concentrations obtained in this case study are within the range of parameter values for optimal duckweed growth reported by Lasfar et al. (2007). In their study, duckweed growth rates have been shown to be practically constant for N between 3-120 mg L^{-1} and P between 1-20 mg L^{-1}, with declining growth rates observed for concentrations below or above these ranges.

The optimal parameter set shows that even with low nutrient concentrations, a high yield of duckweed can be obtained with frequent harvesting. The significance of frequent harvesting in maintaining a high duckweed yield has been emphasized by Said et al. (1979) who reported a 58% increase in duckweed yield with daily harvesting when compared to weekly harvesting for duckweed grown on diluted cattle manure. The higher H_t in the optimized parameter set indicates that it is critical to ensure adequate duckweed biomass accumulation through careful selection of harvesting regime in these production systems. Practical constraints, such as those influenced by the availability of wastewater and labor at a given location, can be incorporated into this optimization framework. By appropriately changing the optimization objectives and/or introducing such constraints, this framework can be used to aid in decision-making related to production and logistics in biorefineries.

3.3. Limitations and future scope

The present study amends an existing duckweed growth model by including a harvesting regime for large-scale applications. With additional model improvements, such as adding a light intensity parameter and incorporating dynamics of nutrient uptake competition between different organisms, duckweed growth can be better represented to improve yield predictions.

The general duckweed growth and harvest model presented in this study calculates the theoretical duckweed yield considering constant parameter values throughout the duration of the simulation period. Daily variation in environmental variables such as temperature and photoperiod was not considered in this model and hence the developed general regression model is only recommended for highly-controlled environments. For large scale applications with less environmental control, a model similar to the Florida case study model would be required to incorporate daily input of temperature and photoperiod. Practical constraints such as wastewater availability (affecting nutrient concentrations in the growth media), and labor availability (influencing harvesting operations), would need to be included to simulate more realistic scenarios.

In terms of future applications, the general duckweed growth and harvest model offers an excellent option for practitioners interested in an easy method to predict duckweed yield. To tailor predictions for a specific geographic location, application of the machine learning-based optimization used in this study would require first running the Stella simulations with location-specific data, and then using the simulation results to fit the machine learning model. With the inclusion of additional model variables such as cloud coverage and duckweed nutrient content, other regression methods and optimization algorithms could be attempted in the future to compare their performance to the methods presented in this study.

Both the general model and site-specific Florida model did not account for harvesting and transportation losses, which may have significant effect on the net duckweed yield in biorefineries. Since harvesting contributes a large share of the total operating cost in biorefineries, a detailed techno-economic analysis with the proposed model would be beneficial in assessing the tradeoffs between maximum biomass yield and production cost.

4. Conclusions

The incorporation of harvest parameters (harvest frequency, harvest ratio, and harvest threshold) play a critical role in improving duckweed yield predictions for large-scale applications with regular harvesting such as a duckweed-based biorefinery. The regression equations developed in this paper, which link simulated duckweed yield to growth and harvest parameters, offer a generalized and simplistic way for users to estimate annual duckweed yields with good accuracy. By integrating a machine learning-based duckweed growth model to an optimization algorithm, we further presented a framework that generated optimal model parameters to maximize duckweed biomass production at a specific geographic location. This framework can be used as a decision-making tool to find parameter combinations that work best to maximize duckweed yield for a given set of location-specific constraints such as labor and wastewater availability. Model refinement utilizing other yield-affecting parameters like cloud coverage and light intensity, and incorporating additional variables like transportation losses, would further enhance the model's performance to better match real-world conditions.

Funding

This work was supported in part by a scholarship from the Fulbright Foreign Student Program for the lead author (O. Calicioglu); a seed grant from the Institutes of Energy and the Environment at The Pennsylvania

State University; and the Northeast Regional Sun Grant Center of the United States Department of Agriculture (USDA).

CRediT authorship contribution statement

Ozgul Calicioglu: Conceptualization, Methodology, Investigation, Data curation, Software, Formal analysis, Validation, Visualization, Writing – original draft. **Mert Y. Sengul:** Conceptualization, Methodology, Investigation, Data curation, Software, Formal analysis, Validation, Visualization. **Pandara Valappil Femeena:** Investigation, Data curation, Software, Formal analysis, Validation, Visualization, Writing – original draft. **Rachel A. Brennan:** Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2021.129120.

References

Al-Nozaily, F., Alaerts, G., Veenstra, S., 2000. Performance of duckweed-covered sewage lagoons-II. Nitrogen and phosphorus balance and plant productivity. Water Res. 34 (10), 2734–2741. https://doi.org/10.1016/S0043-1354(00)00004-X.

Boniardi, N., Rota, R., Nano, G., 1999. Effect of dissolved metals on the organic load removal efficiency of Lemna gibba. Water Res. 33 (2), 530–538. https://doi.org/10.1016/S0043-1354(98)00241-3.

Calicioglu, O., Brennan, R.A., 2018. Sequential ethanol fermentation and anaerobic digestion increases bioenergy yields from duckweed. Bioresour. Technol. 257 (November 2017), 344–348. https://doi.org/10.1016/j.biortech.2018.02.053.

Calicioglu, O., Pandara Valappil, F., Mutel, C., Sills, D.L., Richard, T.L., Brennan, R.A., 2021. Techno-economic analysis and life cycle assessment of an integrated wastewater-derived duckweed pond and biorefinery. ACS Sustain. Chem. Eng. 9 (28), 9395–9408. https://doi.org/10.1021/acssuschemeng.1c02539.

Calicioglu, O., Richard, T.L., Brennan, R.A., 2019. Anaerobic bioprocessing of wastewater-derived duckweed: maximizing product yields in a biorefinery value cascade. Bioresour. Technol. 289 (May), 121716. https://doi.org/10.1016/j.biortech.2019.121716.

Calicioglu, O., Shreve, M.J., Richard, T.L., Brennan, R.A., 2018. Effect of pH and temperature on microbial community structure and carboxylic acid yield during the acidogenic digestion of duckweed. Biotechnol. Biofuels 11 (1), 1–19. https://doi.org/10.1186/s13068-018-1278-6.

Cheng, J.J., Stomp, A.M., 2009. Growing Duckweed to recover nutrients from wastewaters and for production of fuel ethanol and animal feed. Clean 37 (1), 17–26. https://doi.org/10.1002/clen.200800210.

Cui, W., Cheng, J.J., 2015. Growing duckweed for biofuel production: a review. Plant Biol. 17 (s1), 16–23. https://doi.org/10.1111/plb.12216.

Davis, R., Markham, J., Kinchin, C., Grundl, N., Tan, E.C.D., Humbird, D., 2016. Process design and economics for the production of algal biomass: algal biomass production in open pond systems and processing through dewatering for downstream conversion. https://doi.org/10.2172/1239893.

Diritgen, N., Nel, Y.[., 1994. Effects of zinc and copper on growth and metal accumulation in duckweed, lemna minor. Bull. Environ. Contam. Toxicol. 53.

Filbin, G.J., Hough, R.A., 1985. Photosynthesis, photorespiration, and productivity in *Lemna minor* L.1. Limnol. Oceanogr. 30 (2), 322–334. https://doi.org/10.4319/lo.1985.30.2.0322.

Forsythe, W.C., Rykiel, E.J., Stahl, R.S., Wu, H., Schoolfield, R.M., 1995. A model comparison for daylength as a function of latitude and day of year. https://doi.org/10.1016/0304-3800(94)00034-F, 80, 87-95.

Frédéric, M., Samir, L., Louise, M., Abdelkrim, A., 2006. Comprehensive modeling of mat density effect on duckweed (Lemna minor) growth under controlled eutrophication. Water Res. 40 (15), 2901–2910. https://doi.org/10.1016/j.watres.2006.05.026.

Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks. JMLR Workshop and Conference Proceedings. In: http://proceedings.mlr.press/v15/glorot11a.html.

Joseph, V.R., Hung, Y., 2008. Orthogonal-maximin Latin Hypercube designs. In: Statistica Sinica, vol. 18.

Jung, H., Baek, G., Kim, J., Shin, S.G., Lee, C., 2016. Mild-temperature thermochemical pretreatment of green macroalgal biomass: effects on solubilization, methanation, and microbial community structure. Bioresour. Technol. 199, 326–335. https://doi.org/10.1016/j.biortech.2015.08.014.

Kreider, A.N., Fernandez Pulido, C.R., Bruns, M.A., Brennan, R.A., 2019. Duckweed as an agricultural amendment: nitrogen mineralization, leaching, and sorghum uptake. J. Environ. Qual. 48 (2), 469–475. https://doi.org/10.2134/jeq2018.05.0207.

Lasfar, S., Monette, F., Millette, L., Azzouz, A., 2007. Intrinsic growth rate: a new approach to evaluate the effects of temperature, photoperiod and phosphorus-nitrogen concentrations on duckweed growth under controlled eutrophication. Water Res. 41 (11), 2333–2340. https://doi.org/10.1016/j.watres.2007.01.059.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. In: Nature, vol. 521. Nature Publishing Group, pp. 436–444. https://doi.org/10.1038/nature14539, 7553.

Metcalf, E., Eddy, H., 2003. Wastewater engineering: treatment and reuse. In: Tata McGraw-Hill Publishing Company Limited, fourth ed., p. 1819. https://doi.org/10.1016/0309-1708(80)90067-6 New Delhi, India.

Mohedano, R.A., Costa, R.H.R., Tavares, F.A., Belli Filho, P., 2012. High nutrient removal rate from swine wastes and protein biomass production by full-scale duckweed ponds. Bioresour. Technol. 112, 98–104. https://doi.org/10.1016/j.biortech.2012.02.083.

Rejmánková, E., 1973. Seasonal changes in the growth rate of duckweed community (Lemnetum gibbae). Folia Geobot. Phytotaxon. 8 (1), 1–13. https://doi.org/10.1007/BF02854678.

Roman, B., Brennan, R.A., 2019. A beneficial by-product of ecological wastewater treatment: an evaluation of wastewater-grown duckweed as a protein supplement for sustainable agriculture. Ecol. Eng. X 1 (April), 100004. https://doi.org/10.1016/j.ecoena.2019.100004.

Roman, B., Brennan, R.A.A., Lambert, J.D.D., 2021. Duckweed protein supports the growth and organ development of mice : a feeding study comparison to conventional casein protein. J. Food Sci. 86 (3), 1750–3841. https://doi.org/10.1111/1750-3841.15635, 15635.

Said, M.Z.M., Culley, D.D., Standifer, L.C., Epps, E.A., Myers, R.W., Boney, S.A., 1979. Effect of harvest rate, waste loading, and stocking density on the yield of duckweeds. Proceedings of the World Mariculture Society 10 (1–4), 769–780. https://doi.org/10.1111/j.1749-7345.1979.tb00076.x.

Sedlak, R., 1991. In: Sedlak, R. (Ed.), Phosphorus and Nitrogen Removal from Municipal Wastewater- Principles and Practice, Second.

Sengul, M.Y., Song, Y., Nayir, N., Gao, Y., Hung, Y., Dasgupta, T., van Duin, A.C.T., 2020. An initial design-enhanced deep learning-based optimization framework to parameterize multicomponent ReaxFF force fields. In: ChemRxiv. ChemRxiv. https://doi.org/10.26434/chemrxiv.12323909.v2.

Sengul, M.Y., Song, Y., Nayir, N., Gao, Y., Hung, Y., Dasgupta, T., van Duin, A.C.T., 2021. INDEEDopt: a deep learning-based ReaxFF parameterization framework. npj Computational Materials 7 (68), 1–8. https://doi.org/10.1038/s41524-021-00534-4.

Sundui, B., Ramirez Calderon, O.A., Abdeldayem, O.M., Lázaro-Gil, J., Rene, E.R., Sambuu, U., 2021. Applications of machine learning algorithms for biological wastewater treatment: updates and perspectives. Clean Technol. Environ. Policy 23 (1), 127–143. https://doi.org/10.1007/s10098-020-01993-x.

Vose, R.S., Applequist, S., Squires, M., Durre, I., Menne, C.J., Williams, C.N., Fenimore, C., Gleason, K., Arndt, D., 2014. Improved historical temperature and precipitation time series for U.S. climate divisions. Journal of Applied Meteorology and Climatology 53 (5), 1232–1251. https://doi.org/10.1175/JAMC-D-13-0248.1.

Yu, C., Sun, C., Yu, L., Zhu, M., Xu, H., Zhao, J., Ma, Y., Zhou, G., 2014. Comparative analysis of duckweed cultivation with sewage water and SH media for production of fuel ethanol. PloS One 9 (12), 1–15. https://doi.org/10.1371/journal.pone.0115023.

Zhang, X., Zhao, F.J., Huang, Q., Williams, P.N., Sun, G.X., Zhu, Y.G., 2009. Arsenic uptake and speciation in the rootless duckweed Wolffia globosa. New Phytol. 182 (2), 421–428. https://doi.org/10.1111/j.1469-8137.2008.02758.x.

Zhao, Y., Fang, Y., Jin, Y., Huang, J., Bao, S., Fu, T., He, Z., Wang, F., Wang, M., Zhao, H., 2015. Pilot-scale comparison of four duckweed strains from different genera for potential application in nutrient recovery from wastewater and valuable biomass production. Plant Biol. 17 (s1), 82–90. https://doi.org/10.1111/plb.12204.

Zhao, Y., Fang, Y., Jin, Y., Huang, J., Bao, S., He, Z., Wang, F., Zhao, H., 2014. Effects of operation parameters on nutrient removal from wastewater and high-protein biomass production in a duckweed-based (Lemma aequinoctialis) pilot-scale system. Water Sci. Technol. 70 (7), 1195–1204. https://doi.org/10.2166/wst.2014.334.